Commun. Math. Phys. 289, 1087–1098 (2009)
Digital Object Identifier (DOI) 10.1007/s00220-009-0825-1

Communications in
Mathematical
Physics

Universal Coding for Classical-Quantum Channel

Masahito Hayashi

Graduate School of Information Sciences, Tohoku University,
Sendai, 980-8579, Japan.
E-mail: hayashi@math.is.tohoku.ac.jp

Received: 18 July 2008 / Accepted: 12 February 2009
Published online: 13 May 2009 –

Abstract: We construct a universal code for a stationary and memoryless classical-quantum channel as a quantum version of the universal coding by Csiszár and Körner. Our code is constructed utilizing a combination of irreducible representations, a decoder introduced through the quantum information spectrum, and the packing lemma.

1. Introduction

How to transmit information via a noisy communication channel is one of the most important problems for current information network systems. The first big step in this direction was Shannon's channel coding theorem [1], in which he proved that there exists a code enabling reliable communication whose transmission rate is the capacity of the channel, i.e., the maximum of mutual information between the input and output systems. In his formulation, Shannon treated the channel as a stochastic matrix.

In the present paper, we consider the ultimate transmission rate for sending classical messages, when the communication channel is given as a pair of a fixed optical fiber and a fixed modulator. In this case, the input system is described by a set $\mathcal{X}$ of classical alphabets, and the output system is described by a quantum system. Therefore, the channel is given as a map from a classical alphabet to a quantum state (i.e., a density matrix), which is called a classical-quantum channel. In contrast, a stochastic matrix is called a classical channel. When all output density matrices commute with each other, the original coding theorem of Shannon can be trivially extended to the quantum case.

However, in the general case, there is a serious non-commutative difficulty for its quantum extension. Although it is not so difficult to extend the mutual information to this non-commutative quantum framework, there have been several obstacles to the establishment of the channel coding theorem, even for the classical-quantum channel. The crucial obstacle was first resolved by Holevo [2] and Schumacher-Westmoreland [3]. They showed that there exists a reliable code realizing transmission of the maximum value of quantum mutual information. In contrast, in 1970s studies by Holevo [4,5], it

was shown that there does not exist a reliable code overcoming the maximum value of quantum mutual information. The combination of the additivity[1] of the maximum of the mutual information and these results yields the capacity theorem for the classical-quantum channel. That is, it implies that this maximum value is equal to the maximum reliable transmission rate, which is called the capacity. After their achievement, Ogawa and Nagaoka [9] and Hayashi and Nagaoka [10] systematically constructed other codes which realized capacity transmission using the information spectrum method. However, since these existing codes depended on the form of the channel, they were not robust with respect to disagreements between the sender's and receiver's coordinate systems.

In the classical system, Csiszár and Körner [11] constructed a universal channel coding, whose construction does not depend on the channel and depends only on the mutual information and the 'type' of the input system, i.e., the empirical distribution of code words. (The notion of type will be explained in Sect. 3.) Here, we should remark that a universal channel code can universally realize not the capacity but the mutual information because the constructed code is based on an (empirical) distribution on the input classical system whereas universal data compression can universally realize the minimum compression rate for both variable-length settings [12,13] and fixed-length settings [11]. In order to extend Csiszár-Körner's universal coding to the quantum case, we have to overcome the non-commutative obstacle.

Concerning the quantum system, Jozsa et al. [14] constructed a universal fixed-length source coding, which depended only on the compression rate and realized the minimum compression rate. Hayashi [15] discussed the exponentially decreasing rate of the decoding error. Further, Hayashi and Matsumoto [16] constructed a universal variable-length source coding for the quantum system. Hence, we can expect to establish a quantum version of universal channel coding. For example, even if the receiver cannot synchronize his coordinate system with the sender's coordinate system, universal coding guarantees reliable communication.

In the present paper, we construct a universal coding for a classical-quantum channel, which enables transmission of the quantum mutual information and which depends only on the coding rate and the 'type' of the input system. Unfortunately, the capacity cannot be attained universally because its construction depends on the distribution of the input system. In the proposed construction, the following three factors play essential roles for resolving the non-commutative obstacle. One is the decoder given by the proof of the information spectrum method. In the information spectrum method, the decoder is constructed by the square root measurement of the projectors given by the quantum analogue of the likelihood ratio between the signal state and the mixture state [10,17].

The second factor is the irreducible decomposition of the dual representation of the special unitary group and the permutation group, which is known as Schur-duality. The method of irreducible decomposition provides the universal protocols in the quantum setting [14,16,18–22]. However, even in the classical case, the universal channel coding requires the conditional type as well as the type [11]. In the present paper, we introduce a quantum analogue of the conditional type, which is the most essential part of the present paper.

[1] Holevo [5] mentioned this type of additivity, whose proof is written in Fujiwara and Nagaoka [6] (Lemma 3) and Holevo [7]. While Fujiwara and Nagaoka [6] treats the case when the input set $\mathcal{X}$ is given as the set of density matrices of the input quantum system, their proof is valid even when the input set is given as an arbitrary finite set. This is because the key point is essentially shown by the chain rule of classical mutual information. Holevo [7] shows this kind of additivity in a more general setting, in which, he regards this kind of channel as a special case of a channel with a quantum input system.

The third factor is the packing lemma, which yields a suitable combination of the signal states independent of the form of the channel in the classical case [11]. This method plays the same role in the present paper.

An independent work, Bjelakovic and Boche [28], treats a code for a classical-quantum channel that universally realizes transmission of quantum mutual information. However, the result [28] is different from the present paper with respect to the following points. Firstly, the present paper explicitly gives the pair of the encoder and the decoder that universally attains transmission of maximal mutual information. Secondly, the present paper provides an upper bound (26) for the average error probability whose decreasing speed is exponential, whereas the paper [28] does not give such an upper bound. Thirdly, the present paper makes use of Schur-duality, which can be regarded as a kind of quantum extension of the method of type by Csiszár and Körner. This fact suggests that the proposed method can be applied to another topic in Csiszár and Körner.

Further, our construction of encoder does not depend on the dimension of the output system. Only the decoder depends on the dimension of the output system. Note that Csiszár and Körner's construction and Bjelakovic et al's construction depend on the output system. The present paper employs Packing lemma in the construction of encoder as well as Csiszár and Körner. However, the present paper uses this lemma in a way different from Csiszár and Körner. Hence, even if the obtained result is restricted to the classical case, it contains a new result in this point.

The remainder of the present paper is organized as follows. In Sect. 2, we explain the notation used herein and the main result including the existence of a universal coding for a classical-quantum channel. In this section, we present the exponential decreasing rate of the error probability of our universal code. In Sect. 3, the notation for group representation theory is presented and a quantum analogue of conditional type is introduced. In Sect. 4, we provide a code that works well universally. In Sect. 5, the exponentially decreasing rate mentioned in Sect. 2 is proven by using the property given in Sect. 3.

2. Main Result

For the classical-quantum channel (see Fig. 1), we focus on the set of input alphabets $\mathcal{X} := \{1, \ldots, k\}$ and the representation space $\mathcal{H}$ of the output system, whose dimension is d. Then, a classical-quantum channel is given as a map from $\mathcal{X}$ to the set of density matrices on $\mathcal{H}$ of the form $i \mapsto W(i)$. The n-fold discrete memoryless extension is given as the map from $\mathcal{X}^n$ to the set of density matrices on the n^{th} tensor product system $\mathcal{H}^{\otimes n}$. That is, this extension maps the input sequence $\boldsymbol{i} = (i_1, \ldots, i_n)$ to the state $W_n(\boldsymbol{i}_n) := W(i_1) \otimes \cdots \otimes W(i_n)$. Sending the message $\{1, \ldots, M_n\}$ requires an encoder and a decoder. The encoder is given as a map φ_n from the set of messages $\{1, \ldots, M_n\}$ to the set of alphabets $\mathcal{X}^n$, and the decoder is given by a POVM $Y^n = \{Y_i^n\}_{i=1}^{M_n}$. Thus,

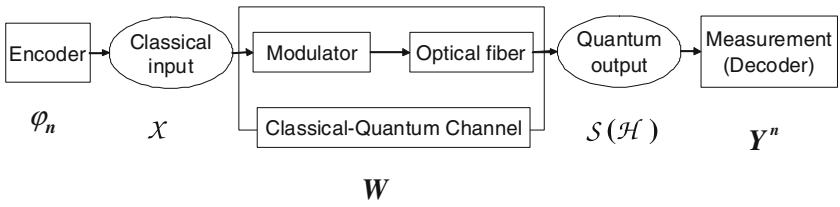

Fig. 1. Figure of classical-quantum channel

the triplet $\Phi_n := (M_n, \varphi_n, Y^n)$ is called a code. Its performance is evaluated by the size $|\Phi_n| := M_n$ and the average error probability, given by

$$\varepsilon[\Phi_n, W] := \frac{1}{M_n} \sum_{i=1}^{M_n} \mathrm{Tr}\, W_n(\varphi_n(i))(I - Y_i^n).$$

The following theorem is known as the classical-quantum channel coding theorem. The optimal reliable transmission rate is equal to the capacity

$$\max_{\boldsymbol{p}} I(\boldsymbol{p}, W),$$

where the mutual information $I(\boldsymbol{p}, W)$ is defined for $\boldsymbol{p} = \{p_i\}_{i=1}^k$ on the set of input alphabets $\mathcal{X} := \{1, \ldots, k\}$ as

$$I(\boldsymbol{p}, W) := \sum_{i=1}^k p_i \,\mathrm{Tr}\, W(i)(\log W(i) - \log W_{\boldsymbol{p}}),$$
$$W_{\boldsymbol{p}} := \sum_{i=1}^k p_i W(i).$$

As stated in the following main theorem, there exists a reliable code that depends only on the coding rate R and the distribution $\boldsymbol{p}$ on the input system when the coding rate R is smaller than the mutual information $I(\boldsymbol{p}, W)$. Note that this theorem does not imply the universal achievement of the capacity $\max_{\boldsymbol{p}} I(\boldsymbol{p}, W)$ because our construction depends on the input distribution $\boldsymbol{p}$.

Theorem 1. *For any distribution $\boldsymbol{p} = \{p_i\}_{i=1}^k$ on the set of input alphabets $\mathcal{X} := \{1, \ldots, k\}$ and any real number R, there is a sequence of codes $\{\Phi_n\}_{n=1}^\infty$ such that*

$$\lim_{n\to\infty} \frac{-1}{n} \log \varepsilon[\Phi_n, W] \ge \max_{0\le t\le 1} \frac{\phi_{W,\boldsymbol{p}}(t) - tR}{1+t},$$
$$\lim_{n\to\infty} \frac{1}{n} \log |\Phi_n| = R$$

for any classical-quantum channel W, where $\phi_{W,\boldsymbol{p}}(t)$ is given by

$$\phi_{W,\boldsymbol{p}}(t) := -(1-t) \log \mathrm{Tr} \left(\sum_{i=1}^k p_i W(i)^{1-t} \right)^{\frac{1}{1-t}}.$$

Note that the code $\{\Phi_n\}_{n=1}^\infty$ does not depend on the channel W, and depends only on the distribution $\boldsymbol{p}$ and the coding rate R.

The derivative of $\phi_{W,\boldsymbol{p}}(t)$ is given as

$$\phi'_{W,\boldsymbol{p}}(0) = I(\boldsymbol{p}, W).$$

When the transmission rate R is smaller than the mutual information $I(\boldsymbol{p}, W)$,

$$\max_{0\le t\le 1} \frac{\phi_{W,\boldsymbol{p}}(t) - tR}{1+t} > 0$$

because there exists a parameter $t \in (0, 1)$ such that $\phi_{W,\boldsymbol{p}}(t) - tR > 0$. That is, the average error probability $\varepsilon[\Phi_n, W]$ goes to zero.

This fact implies that without knowledge of the channel W we can mathematically construct a reliable code based only on the input distribution $\boldsymbol{p}$ when the coding rate R is smaller than the mutual information $I(\boldsymbol{p}, W)$. Therefore, in order to construct a code attaining the capacity $\max_{\boldsymbol{p}} I(\boldsymbol{p}, W)$, we need to know only the value of $\operatorname{argmax}_{\boldsymbol{p}} I(\boldsymbol{p}, W)$. We do not require complete knowledge of the classical-quantum channel $i \mapsto W(i)$. For example, we may not be able to identify the coordinates of the output system, i.e., we cannot identify only the unitary U in the classical-quantum channel $i \mapsto UW(i)U^\dagger$; however we are able to identify the maximizing input distribution $\operatorname{argmax}_{\boldsymbol{p}} I(\boldsymbol{p}, W)$. In this case, we can construct a code that realizes the capacity $\max_{\boldsymbol{p}} I(\boldsymbol{p}, W)$. Furthermore, the proposed code is robust with respect to small disturbances, in other words, our evaluation (26) of the average error probability guarantees reliable communication under the proposed code even if the true channel is a little different from the estimated channel.

3. Group Representation Theory

In this section, we focus on the dual representation of the n-fold tensor product space by the special unitary group $\mathrm{SU}(d)$ and the n^{th} symmetric group S_n.[2] For this purpose, we focus on the Young diagram and the 'type'. The former is a key concept in group representation theory and the latter is the corresponding notion in information theory [11]. When the vector of integers $\boldsymbol{n} = (n_1, n_2, \cdots, n_d)$ satisfies the condition $n_1 \geq n_2 \geq \cdots \geq n_d \geq 0$ and $\sum_{i=1}^d n_i = n$, the vector $\boldsymbol{n}$ is called a Young diagram (frame) of size n and depth d; the set of such vectors is denoted as Y_n^d. When the vector of integers $\boldsymbol{n}$ satisfies the condition $n_i \geq 0$ and $\sum_{i=1}^d n_i = n$, the vector $\boldsymbol{p} = \frac{\boldsymbol{n}}{n}$ is called a 'type' of size n; the set of these vectors is denoted as T_n^d. Further, for $\boldsymbol{p} \in T_n^d$, a subset of $\mathcal{X}^n$ is defined by:

$$T_{\boldsymbol{p}} := \{\boldsymbol{x} \in \mathcal{X}^n | \text{The empirical distribution of } \boldsymbol{x} \text{ is equal to } \boldsymbol{p}\}.$$

The cardinalities of these sets are constrained as follows:

$$|Y_n^d| \leq |T_n^d| \leq (n+1)^{d-1}, \tag{1}$$

$$(n+1)^{-d} e^{nH(\boldsymbol{p})} \leq |T_{\boldsymbol{p}}|, \tag{2}$$

where $H(\boldsymbol{p}) := -\sum_{i=1}^d p_i \log p_i$ [11]. Using the Young diagram, the irreducible decomposition of the above representation can be characterized as follows:

$$\mathcal{H}^{\otimes n} = \bigoplus_{\boldsymbol{n} \in Y_n^d} \mathcal{U}_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}, \tag{3}$$

where $\mathcal{U}_{\boldsymbol{n}}$ is the irreducible representation space of $\mathrm{SU}(d)$ characterized by $\boldsymbol{n}$, and $\mathcal{V}_{\boldsymbol{n}}$ is the irreducible representation space of n^{th} symmetric group S_n characterized by $\boldsymbol{n}$. Here, the representation of the n^{th} symmetric group S_n is denoted as $V : s \in S_n \mapsto V_s$. Hence, Eq. (3) gives the irreducible decomposition of the representation of the group

[2] Christandl [23] contains a good survey of representation theory for quantum information.

$\mathrm{SU}(d) \times S_n$, which is called Schur-duality. For $\boldsymbol{n} \in Y_n^d$, the dimension of $\mathcal{U}_{\boldsymbol{n}}$ is evaluated by

$$\dim \mathcal{U}_{\boldsymbol{n}} \le n^{\frac{d(d-1)}{2}}. \tag{4}$$

Then, denoting the projection to the subspace $\mathcal{U}_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}$ as $I_{\boldsymbol{n}}$, we define the following:

$$\rho_{\boldsymbol{n}} := \frac{1}{\dim \mathcal{U}_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}} I_{\boldsymbol{n}}, \tag{5}$$

$$\rho_{U,n} := \sum_{\boldsymbol{n} \in Y_n^d} \frac{1}{|Y_n^d|} \rho_{\boldsymbol{n}}. \tag{6}$$

Any state ρ and any Young diagram $\boldsymbol{n} \in Y_n^d$ satisfy the following:

$$\dim \mathcal{U}_{\boldsymbol{n}} \rho_{\boldsymbol{n}} \ge I_{\boldsymbol{n}} \rho^{\otimes n} I_{\boldsymbol{n}}.$$

Thus, (1), (4), and (6) yield the inequality

$$n^{\frac{d(d-1)}{2}} |Y_n^d| \rho_{U,n} \ge \rho^{\otimes n}. \tag{7}$$

Next, we focus on two systems $\mathcal{X}$ and $\mathcal{Y} = \{1, \ldots, l\}$. When the distribution of $\mathcal{X}$ is given by a probability distribution $\boldsymbol{p} = (p_1, \ldots, p_d)$ on $\{1, \ldots, d\}$, and the conditional distribution on $\mathcal{Y}$ with the condition on $\mathcal{X}$ is given by $\boldsymbol{V}$, we denote the joint distribution on $\mathcal{X} \times \mathcal{Y}$ by $\boldsymbol{pV}$ and the distribution on $\mathcal{Y}$ by $\boldsymbol{p} \cdot \boldsymbol{V}$. When the empirical distribution of $\boldsymbol{x} \in \mathcal{X}^n$ is $(\frac{n_1}{n}, \ldots, \frac{n_d}{n})$, the sequence of types $\boldsymbol{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d) \in T_{n_1}^l \times \cdots \times T_{n_d}^l$ is called a conditional type for $\boldsymbol{x}$ [11]. We denote the set of conditional types for $\boldsymbol{x}$ by $V(\boldsymbol{x}, \mathcal{Y})$. For any conditional type V for $\boldsymbol{x}$, we define the subset of $\mathcal{Y}^n$:

$$T_{\boldsymbol{V}}(\boldsymbol{x}) := \left\{ \boldsymbol{y} \in \mathcal{Y}^n \,\middle|\, \begin{array}{l} \text{The empirical distribution of} \\ ((x_1, y_1), \ldots, (x_n, y_n)) \text{ is equal to } \boldsymbol{pV}. \end{array} \right\},$$

where $\boldsymbol{p}$ is the empirical distribution of $\boldsymbol{x}$.

We define the state $\rho_{\boldsymbol{x}}$ for $\boldsymbol{x} \in \mathcal{X}^n$. For this purpose, we consider a special element $\boldsymbol{x}' = (\underbrace{1, \ldots, 1}_{m_1}, \underbrace{2, \ldots, 2}_{m_2}, \ldots, \underbrace{k, \ldots, k}_{m_k})$. The state $\rho_{\boldsymbol{x}'}$ is defined as $\rho_{\boldsymbol{x}'} := \rho_{U,m_1} \otimes \rho_{U,m_2} \otimes \cdots \otimes \rho_{U,m_k}$. For a general element $\boldsymbol{x} \in \mathcal{X}^n$, we choose a permutation $s \in S_n$ such that $\boldsymbol{x} = s\boldsymbol{x}'$. Then, we define a state $\rho_{\boldsymbol{x}}$ by $\rho_{\boldsymbol{x}} := U_s \rho_{\boldsymbol{x}'} U_s^\dagger$, where U_s is the unitary representation of S_n. This state plays a similar role to the conditional type in the classical case. Using the inequality (7), we have

$$n^{\frac{kd(d-1)}{2}} |Y_n^d|^k \rho_{\boldsymbol{x}} \ge W_n(\boldsymbol{x}). \tag{8}$$

As is shown here, the density matrix $\rho_{\boldsymbol{x}'} := \rho_{U,m_1} \otimes \rho_{U,m_2} \otimes \cdots \otimes \rho_{U,m_k}$ commutes with $\rho_{U,n}$. For simplicity, we show commutativity between $\rho_{U,m_1} \otimes \rho_{U,m_2}$ and ρ_{U,m_1+m_2}, first. In order to prove this fact, it is sufficient to show the existence of a resolution of the identity by the projections $\{E_i\}_i$ such that

$$\exists \{a_i\}, \quad \rho_{U,m_1} \otimes \rho_{U,m_2} = \sum_i a_i E_i, \tag{9}$$

$$\exists \{b_i\}, \quad \rho_{U,m_1+m_2} = \sum_i b_i E_i. \tag{10}$$

When the resolution $\{E_i^{\mathbf{1}}\}$ of the identity is given as projections to the irreducible spaces of the representation of the group $\mathrm{SU}(d) \times S_{m_1+m_2}$, the resolution $\{E_i^{\mathbf{1}}\}$ satisfies the condition (10) because of the construction of ρ_{U,m_1+m_2}. Similarly, the resolution $\{E_i^2\}$ of the identity is given as projections to the irreducible spaces of the representation of the group $\mathrm{SU}(d) \times \mathrm{SU}(d) \times S_{m_1} \times S_{m_2} = (\mathrm{SU}(d) \times S_{m_1}) \times (\mathrm{SU}(d) \times S_{m_2})$, and the resolution $\{E_i^2\}$ satisfies the condition (9). The group $\mathrm{SU}(d) \times S_{m_1} \times S_{m_2}$ is a subgroup of $\mathrm{SU}(d) \times S_{m_1+m_2}$, and it is also a subgroup of $\mathrm{SU}(d) \times \mathrm{SU}(d) \times S_{m_1} \times S_{m_2}$ via the correspondence $(g, s_1, s_2) \mapsto (g, g, s_1, s_2)$. Now, the resolution $\{E_j^3\}_{j\in J}$ of the identity is given as projections to the irreducible spaces of the representation of the group $\mathrm{SU}(d) \times S_{m_1} \times S_{m_2}$. For any $E_i^1 \in \{E_i^1\}$, there is a subset J_i of J such that $E_i^1 = \sum_{j\in J_i} E_j^3$. The same fact holds for $\{E_i^2\}$. Therefore, the resolution $\{E_j^3\}_{j\in J}$ satisfies (10) and (9). Thus, the density matrix $\rho_{U,m_1} \otimes \rho_{U,m_2}$ commutes with ρ_{U,m_1+m_2}. Applying the same discussion to the group $\mathrm{SU}(d) \times S_{m_1} \times S_{m_2} \times \cdots \times S_{m_k}$, we can show that $\rho_{\boldsymbol{x}'} := \rho_{U,m_1} \otimes \rho_{U,m_2} \otimes \cdots \otimes \rho_{U,m_k}$ commutes with $\rho_{U,n}$. This property is essential for the construction of the proposed decoder.

4. Construction of the Code

According to Csiszár and Körner [11], the proposed code is constructed as follows. The main point of this section is to establish that Csiszár-Körner's Packing lemma provides a code whose performance is essentially equivalent to the average performance of random coding in the sense of (12). In the following discussion, we treat the conditional type in the case when the system $\mathcal{Y}$ coincides with the other system $\mathcal{X}$.

Lemma 1. *For a positive number $\delta > 0$, a type $\boldsymbol{p} \in T_n^d$, and a real positive number $R < H(\boldsymbol{p})$, there exist $M_n := e^{n(R-\delta)}$ distinct elements $\mathcal{M}_n := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M_n}\} \subset T_{\boldsymbol{p}}$ such that their empirical distributions are $\boldsymbol{p}$ and*

$$|T_{\boldsymbol{V}}(\boldsymbol{x}) \cap (\mathcal{M}_n \setminus \{\boldsymbol{x}\})| \le |T_{\boldsymbol{V}}(\boldsymbol{x})| e^{-n(H(\boldsymbol{p})-R)}$$

for $\boldsymbol{x} \in \mathcal{M}_n \subset T_{\boldsymbol{p}}$ and $\boldsymbol{V} \in V(\boldsymbol{x}, \mathcal{X})$.

This lemma can be shown by substituting the identical map into $\hat{V}$ in Lemma 5.1 in Csiszár and Körner [11], which is known as the Packing lemma. Since Csiszár and Körner proved Lemma 5.1 using the random coding method, we can replace δ by $\frac{1}{\sqrt{n}}$. That is, there exist $M_n := e^{nR-\sqrt{n}}$ distinct elements $\mathcal{M}_n := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M_n}\} \subset T_{\boldsymbol{p}}$ such that their empirical distributions are $\boldsymbol{p}$ and

$$|T_{\boldsymbol{V}}(\boldsymbol{x}) \cap (\mathcal{M}_n \setminus \{\boldsymbol{x}\})| \le |T_{\boldsymbol{V}}(\boldsymbol{x})| e^{-n(H(\boldsymbol{p})-R)} \tag{11}$$

for $\boldsymbol{x} \in \mathcal{M}_n \subset T_{\boldsymbol{p}}$ and $\boldsymbol{V} \in V(\boldsymbol{x}, \mathcal{X})$. Note that this encoder $\mathcal{M}_n$ does not depend on the output system because the employed Packing lemma treats the conditional types from the input system to the input system. Now, we transform the property (11) to a more useful form.

Using the encoder $\mathcal{M}_n$, we can define the distribution $P_{\mathcal{M}_n}$ as

$$p_{\mathcal{M}_n}(\boldsymbol{x}) = \begin{cases} \frac{1}{|\mathcal{M}_n|} & \boldsymbol{x} \in \mathcal{M}_n \\ 0 & \boldsymbol{x} \notin \mathcal{M}_n. \end{cases}$$

For any $\boldsymbol{x} \in \mathcal{X}^n$, we define an invariant subgroup $S_{\boldsymbol{x}} \subset S_n$:

$$S_{\boldsymbol{x}} := \{s \in S_n | s(\boldsymbol{x}) = \boldsymbol{x}\}.$$

Since $\boldsymbol{x}' \in T_{\boldsymbol{p}}$ implies that

$$\boldsymbol{p}^n(\boldsymbol{x}') = e^{-nH(\boldsymbol{p})},$$

any element $\boldsymbol{x}' \in T_{\boldsymbol{V}}(\boldsymbol{x}) \cap \mathcal{M}_n \subset T_{\boldsymbol{p}}$ satisfies

$$\begin{aligned}\sum_{s \in S_{\boldsymbol{x}}} \frac{1}{|S_{\boldsymbol{x}}|} p_{\mathcal{M}_n} \circ s(\boldsymbol{x}') &= \frac{|T_{\boldsymbol{V}}(\boldsymbol{x}) \cap \mathcal{M}_n|}{|T_{\boldsymbol{V}}(\boldsymbol{x})|} \cdot \frac{1}{|\mathcal{M}_n|} = \frac{|T_{\boldsymbol{V}}(\boldsymbol{x}) \cap (\mathcal{M}_n \setminus \{\boldsymbol{x}\})|}{|T_{\boldsymbol{V}}(\boldsymbol{x})||\mathcal{M}_n|} \\ &\le e^{-nH(\boldsymbol{p})} e^{\sqrt{n}} = \boldsymbol{p}^n(\boldsymbol{x}') e^{\sqrt{n}} \end{aligned} \tag{12}$$

when the conditional type $\boldsymbol{V}$ is not identical. Relation (12) holds for any $\boldsymbol{x}' (\neq \boldsymbol{x}) \in \mathcal{M}_n$ because there exists a conditional type $\boldsymbol{V}$ such that $\boldsymbol{x}' \in T_{\boldsymbol{V}}(\boldsymbol{x})$ and $\boldsymbol{V}$ is not identical.

Next, for any $\boldsymbol{x} \in \mathcal{X}^n$ and any real number C_n, we define the projection

$$P(\boldsymbol{x}) := \{\rho_{\boldsymbol{x}} - C_n \rho_{U,n} \ge 0\},$$

where $\{X \ge 0\}$ presents the projection $\sum_{i:x_i \ge 0} E_i$ for a Hermitian matrix X with the diagonalization $X = \sum_i x_i E_i$. Remember that the density matrix $\rho_{\boldsymbol{x}}$ commutes with the other density matrix $\rho_{U,n}$. Using the projection $P(\boldsymbol{x})$, we define the decoder:

$$Y_{\boldsymbol{x}'} := \sqrt{\sum_{\boldsymbol{x} \in \mathcal{M}_n} P(\boldsymbol{x})}^{-1} P(\boldsymbol{x}') \sqrt{\sum_{\boldsymbol{x} \in \mathcal{M}_n} P(\boldsymbol{x})}^{-1}.$$

In the following, the above-constructed code $(e^{nR-\sqrt{n}}, \mathcal{M}_n, \{Y_{\boldsymbol{x}}\}_{\boldsymbol{x} \in \mathcal{M}_n})$ is denoted by $\Phi_{U,n}(\boldsymbol{p}, R)$.

5. Exponential Evaluation

Hayashi and Nagaoka [10] showed that

$$I - Y_{\boldsymbol{x}'} \le 2(I - P(\boldsymbol{x}')) + 4 \sum_{\boldsymbol{x} (\neq \boldsymbol{x}') \in \mathcal{M}_n} P(\boldsymbol{x}).$$

Then, the average error probability of $\Phi_{U,n}(\boldsymbol{p}, R)$ is evaluated by

$$\begin{aligned}&\frac{1}{|\mathcal{M}_n|} \sum_{\boldsymbol{x}' \in \mathcal{M}_n} \mathrm{Tr}\, W_n(\boldsymbol{x}')(I - Y_{\boldsymbol{x}'}) \\ &\le \frac{2}{|\mathcal{M}_n|} \sum_{\boldsymbol{x}' \in \mathcal{M}_n} \mathrm{Tr}\, W_n(\boldsymbol{x}')(I - P(\boldsymbol{x}')) + \frac{4}{|\mathcal{M}_n|} \sum_{\boldsymbol{x}' \in \mathcal{M}_n} \mathrm{Tr}\, W_n(\boldsymbol{x}') \sum_{\boldsymbol{x} (\neq \boldsymbol{x}') \in \mathcal{M}_n} P(\boldsymbol{x}) \\ &= \frac{2}{|\mathcal{M}_n|} \sum_{\boldsymbol{x} \in \mathcal{M}_n} \mathrm{Tr}\, W_n(\boldsymbol{x})(I - P(\boldsymbol{x})) \\ &\quad + 4\,\mathrm{Tr} \left[\sum_{\boldsymbol{x} \in \mathcal{M}_n} P(\boldsymbol{x}) \left(\frac{1}{|\mathcal{M}_n|} \sum_{\boldsymbol{x}' (\neq \boldsymbol{x}) \in \mathcal{M}_n} W_n(\boldsymbol{x}') \right) \right]. \end{aligned} \tag{13}$$

Since the density matrix $\rho_{\boldsymbol{x}}$ commutes with the density matrix $\rho_{U,n}$, we have

$$(I - P(\boldsymbol{x})) = \{\rho_{\boldsymbol{x}} - C_n \rho_{U,n} < 0\} \le \rho_{\boldsymbol{x}}^{-t} C_n^t \rho_{U,n}^t \tag{14}$$

for $0 \le t \le 1$. Since the density matrix $\rho_{\boldsymbol{x}}$ commutes with the density matrix $W_n(\boldsymbol{x})$, $W_n(\boldsymbol{x})\rho_{\boldsymbol{x}}^{-t}$ is a Hermite matrix and (8) implies that

$$W_n(\boldsymbol{x})\rho_{\boldsymbol{x}}^{-t} \le n^{\frac{ktd(d-1)}{2}} |Y_n^d|^{kt} W_n(\boldsymbol{x})^{1-t}. \tag{15}$$

Using (14) and (15), we have

$$\begin{aligned} &\operatorname{Tr} W_n(\boldsymbol{x})(I - P(\boldsymbol{x})) \le \operatorname{Tr} W_n(\boldsymbol{x})\rho_{\boldsymbol{x}}^{-t}\rho_{U,n}^t C_n^t \\ &\le n^{\frac{ktd(d-1)}{2}} |Y_n^d|^{kt} C_n^t \operatorname{Tr} W_n(\boldsymbol{x})^{1-t}\rho_{U,n}^t. \end{aligned} \tag{16}$$

Since the quantity $\operatorname{Tr} W_n(\boldsymbol{x})(I - P(\boldsymbol{x}))$ is invariant with respect to the action of the permutation and the relation (2) implies that

$$\boldsymbol{p}^n(\boldsymbol{x}) = e^{-nH(\boldsymbol{p})} \ge \frac{(n+1)^{-d}}{|T_{\boldsymbol{p}}|} \tag{17}$$

for $\boldsymbol{x} \in T_{\boldsymbol{p}}$, we obtain

$$\begin{aligned} &\operatorname{Tr} W_n(\boldsymbol{x})(I - P(\boldsymbol{x})) = \frac{1}{|T_{\boldsymbol{p}}|} \sum_{\boldsymbol{x}' \in T_{\boldsymbol{p}}} \operatorname{Tr} W_n(\boldsymbol{x}')(I - P(\boldsymbol{x}')) \\ &\le (n+1)^d \sum_{\boldsymbol{x}' \in \mathcal{X}^n} \boldsymbol{p}^n(\boldsymbol{x}') \operatorname{Tr} W_n(\boldsymbol{x}')(I - P(\boldsymbol{x}')) \qquad (18) \\ &\le (n+1)^d n^{\frac{ktd(d-1)}{2}} |Y_n^d|^{kt} C_n^t \operatorname{Tr}\Big(\sum_{\boldsymbol{x}' \in \mathcal{X}^n} \boldsymbol{p}^n(\boldsymbol{x}') W_n(\boldsymbol{x}')^{1-t}\Big)\rho_{U,n}^t \qquad (19) \\ &\le (n+1)^{d+\frac{ktd(d-1)}{2}} |Y_n^d|^{kt} C_n^t \max_{\sigma} \operatorname{Tr}\left[\sum_{x \in \mathcal{X}} \boldsymbol{p}(x) W_n(x)^{1-t}\right]^{\otimes n} \sigma^t \\ &\le (n+1)^{d+\frac{ktd(d-1)}{2}} |Y_n^d|^{kt} C_n^t \left(\operatorname{Tr}\left(\left[\sum_{x \in \mathcal{X}} \boldsymbol{p}(x) W_n(x)^{1-t}\right]^{\otimes n}\right)^{\frac{1}{1-t}}\right)^{1-t} \qquad (20) \\ &= (n+1)^{d+\frac{ktd(d-1)}{2}} |Y_n^d|^{kt} C_n^t \left(\operatorname{Tr}\left(\sum_{x \in \mathcal{X}} \boldsymbol{p}(x) W_n(x)^{1-t}\right)^{\frac{1}{1-t}}\right)^{n(1-t)} \\ &= (n+1)^{d+\frac{ktd(d-1)}{2}} |Y_n^d|^{kt} C_n^t e^{-n\phi_{W,\boldsymbol{p}}(t)}, \qquad (21) \end{aligned}$$

where (18) and (19) follow from (17) and (16), respectively. The inequality (20) can be checked in the following way: When X is a positive semi-definite matrix, σ is a density matrix, $0 \le t \le 1$, $p = 1/(1-t)$, and $q = 1/t$, the Hölder inequality implies that

$$\operatorname{Tr} X\sigma^t \le \operatorname{Tr} |X\sigma^t| \le (\operatorname{Tr} X^p)^{\frac{1}{p}} (\operatorname{Tr} \sigma^{tq})^{\frac{1}{q}} = (\operatorname{Tr} X^{\frac{1}{1-t}})^{1-t}$$

because $\frac{1}{p} + \frac{1}{q} = 1$. This inequality yields (20).

Next, we evaluate the second term of (13) using the invariant property of $S_{\boldsymbol{x}}$:

$$\begin{aligned}
&\mathrm{Tr}\left[P(\boldsymbol{x})\left(\frac{1}{|\mathcal{M}_n|}\sum_{\boldsymbol{x}'(\neq\boldsymbol{x})\in\mathcal{M}_n} W_n(\boldsymbol{x}')\right)\right] \\
=& \mathrm{Tr}\left[P(\boldsymbol{x})\sum_{\boldsymbol{x}'(\neq\boldsymbol{x})\in\mathcal{M}_n} p_{\mathcal{M}_n}(\boldsymbol{x}')W_n(\boldsymbol{x}')\right] \\
=& \mathrm{Tr}\left[P(\boldsymbol{x})\sum_{s\in S_{\boldsymbol{x}}}\frac{1}{|S_{\boldsymbol{x}}|}\sum_{\boldsymbol{x}'(\neq\boldsymbol{x})\in\mathcal{M}_n} p_{\mathcal{M}_n}(\boldsymbol{x}')V_s W_n(\boldsymbol{x}')V_s^*\right] \\
=& \mathrm{Tr}\left[P(\boldsymbol{x})\sum_{\boldsymbol{x}'(\neq\boldsymbol{x})\in\mathcal{M}_n}\sum_{s\in S_{\boldsymbol{x}}}\frac{1}{|S_{\boldsymbol{x}}|} p_{\mathcal{M}_n}\circ s^{-1}(\boldsymbol{x}')W_n(\boldsymbol{x}')\right] \\
\le& \mathrm{Tr}\left[P(\boldsymbol{x})\sum_{\boldsymbol{x}'(\neq\boldsymbol{x})\in\mathcal{M}_n} \boldsymbol{p}^n(\boldsymbol{x}')e^{\sqrt{n}}W_n(\boldsymbol{x}')\right] \qquad (22)\\
=& e^{\sqrt{n}}\,\mathrm{Tr}\left[P(\boldsymbol{x})W_{\boldsymbol{p}}^{\otimes n}\right] \\
\le& e^{\sqrt{n}}\,\mathrm{Tr}\left[P(\boldsymbol{x})n^{\frac{d(d-1)}{2}}|Y_n^d|\rho_{U,n}\right] \qquad (23)\\
\le& e^{\sqrt{n}}\,\mathrm{Tr}\left[P(\boldsymbol{x})n^{\frac{d(d-1)}{2}}|Y_n^d|C_n^{-1}\rho_{\boldsymbol{x}}\right] \qquad (24)\\
\le& e^{\sqrt{n}}\,\mathrm{Tr}\left[n^{\frac{d(d-1)}{2}}|Y_n^d|C_n^{-1}\rho_{\boldsymbol{x}}\right] = e^{\sqrt{n}}n^{\frac{d(d-1)}{2}}|Y_n^d|C_n^{-1}, \qquad (25)
\end{aligned}$$

where (22), (23), and (24) follow from (12), (7), and the inequality $P(\boldsymbol{x})(\rho_{U,n} - C_n^{-1}\rho_{\boldsymbol{x}}) \le 0$.

For any $t \in (0,1)$ and $R > 0$, we choose $|\mathcal{M}_n| := e^{nR-\sqrt{n}}$, $C_n := e^{n(R+r(t))}$, and $r(t) := \frac{\phi_{W,\boldsymbol{p}}(t)-tR}{1+t}$. Since $r(t) = \phi_{W,\boldsymbol{p}}(t) - t(R + r(t))$, from (13), (21) and (25), the average error probability can be evaluated as

$$\begin{aligned}
&\varepsilon(\Phi_{U,n}(\boldsymbol{p}, R), W) \\
&\quad\le 2(n+1)^{d+\frac{ktd(d-1)}{2}}|Y_n^d|^{kt}e^{-n(\phi_{W,\boldsymbol{p}}(t)-t(R+r(t)))} + 4n^{\frac{d(d-1)}{2}}|Y_n^d|e^{-nr(t)}. \qquad (26)
\end{aligned}$$

Then, its exponentially decreasing rate is characterized by

$$\lim_{n\to\infty}\frac{-1}{n}\log\varepsilon(\Phi_{U,n}(\boldsymbol{p}, R), W) \ge \min\{\phi_{W,\boldsymbol{p}}(t) - t(R + r(t)), r(t)\} = \frac{\phi_{W,\boldsymbol{p}}(t) - tR}{1+t}.$$

That is, when we choose $t_0 := \mathrm{argmax}_{t\in(0,1)}\frac{\phi_{W,\boldsymbol{p}}(t)-tR}{1+t}$, $|\mathcal{M}_n| := e^{nR-\sqrt{n}}$, and $C_n := e^{n(R+r(t_0))}$, we obtain

$$\lim_{n\to\infty}\frac{-1}{n}\log\varepsilon(\Phi_{U,n}(\boldsymbol{p}, R), W) \ge \max_{t\in(0,1)}\frac{\phi_{W,\boldsymbol{p}}(t) - tR}{1+t}$$

for any channel W. Therefore, we obtain Theorem 1.

6. Discussion

We have constructed a universal code realizing the transmission of quantum mutual information by combining the information spectrum method with group representation theory and using the Packing lemma. The code that we have developed works well because any tensor product state $\rho^{\otimes n}$ is close to the state $\rho_{U,n}$. Indeed, Krattenthaler and Slater [24] demonstrated the existence of a state σ_n such that $\frac{1}{n} D(\rho^{\otimes n} \| \sigma_n) \to 0$ for any state ρ in the qubit system as a quantum analogue of Clarke and Barron's result [25]. Its d-dimensional extension is discussed in another paper [26].

Further, Hayashi [27] derived another exponentially decreasing rate of error probability for a classical-quantum channel, which is $\max_{t:0\le t\le 1} -(\log \sum_i p_i \operatorname{Tr}[W(i)^{1-t} W_{\boldsymbol{p}}^t]) - tR$. Since

$$e^{-\frac{\phi_{W,\boldsymbol{p}}(t)-t(R+r(t))}{1+t}} = e^{-(\phi_{W,\boldsymbol{p}}(t)-t(R+r(t)))} = e^{t(R+r(t))} \max_{\sigma} \operatorname{Tr}\Big(\sum_i p_i W(i)^{1-t}\Big)\sigma^t$$

$$\ge e^{tR} \operatorname{Tr}\Big(\sum_i p_i W(i)^{1-t}\Big)\Big(\sum_i p_i W(i)\Big)^t = e^{-(-(\log \sum_i p_i \operatorname{Tr}[W(i)^{1-t} W_{\boldsymbol{p}}^t]) - tR)},$$

we obtain

$$\max_{t:0\le t\le 1} -(\log \sum_i p_i \operatorname{Tr}[W(i)^{1-t} W_{\boldsymbol{p}}^t]) - tR \ge \max_{t:0\le t\le 1} \frac{\phi_{W,\boldsymbol{p}}(t) - tR}{1+t}.$$

That is, the obtained exponentially decreasing rate is smaller than that of Hayashi [27]. However, according to Csiszár and Körner [11], the exponentially decreasing rate of the universal coding is the same as the optimal exponentially decreasing rate in the classical case when the rate is close to the capacity. Hence, if a more sophisticated analysis were to be applied, a better exponentially decreasing rate could be expected. Such an analysis is left as a future problem.

The proposed encoder does not depend on the output system. Such a construction is realized by employing the Packing lemma in a way different from that of Csiszár and Körner. In the present paper, the Packing lemma treats the conditional types from the input system to the input system. We hope that such a style of application of the Packing lemma yields another new result on information theory in the future.

Acknowledgement. This research was partially supported by a Grant-in-Aid for Scientific Research on Priority Area 'Deepening and Expansion of Statistical Mechanical Informatics (DEX-SMI)', No. 18079014 and a MEXT Grant-in-Aid for Young Scientists (A) No. 20686026. The author thanks the referees and the editor for helpful comments concerning this manuscript. He also acknowledges Professor Hiroshi Nagaoka for an interesting discussion.

References

1. Shannon, C.E.: A mathematical theory of communication. Bell System Technical Journal **27**, 623–656 (1948)
2. Holevo, A.S.: The capacity of the quantum channel with general signal states. IEEE Trans. Inform. Theory **44**, 269–273 (1998)

3. Schumacher, B., Westmoreland, M.D.: Sending classical information via noisy quantum channels. Phys. Rev. A **56**, 131–138 (1997)
4. Holevo, A.S.: Bounds for the quantity of information transmitted by a quantum communication channel. Probl. Inform. Transm. **9**, 177–183 (1973)
5. Holevo, A.S.: On the capacity of quantum communication channel. Probl. Inform. Transm. **15**(4), 247–253 (1979)
6. Fujiwara, A., Nagaoka, H.: *Capacity of a memoryless quantum communication channel*. Mathematical Engineering Technical Reports, METR 94-22, the University of Tokyo, http://www.keisu.t.u-tokyo.ac.jp/research/techrep/1994.html, 1994
7. Holevo, A.S.: *Quantum Coding Theorems*. Russ. Math. Surv. **53**:6, 1295–1331 (1999); *Coding Theorems for quantum channels*, http://arXiv.org/abs/quant-ph/9809023, 1998
8. Hayashi, M.: Quantum Information: An Introduction. Berlin: Springer, 2006
9. Ogawa, T., Nagaoka, H.: Making good codes for classical-quantum channel coding via quantum hypothesis testing. IEEE Trans. Inform. Theory **53**, 2261–2266 (2007)
10. Hayashi, M., Nagaoka, H.: General formulas for capacity of classical-quantum channels. IEEE Trans. Infor. Theory **49**, 1753–1768 (2003)
11. Csiszár, I., Körner, J.: *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London-New York: Academic Press, 1981
12. Lynch, T.J.: Sequence time coding for data compression. Proc. IEEE **54**, 1490–1491 (1966)
13. Davisson, L.D.: Comments on Sequence time coding for data compression. Proc. IEEE **54**:2010, 1966
14. Jozsa, R., Horodecki, M., Horodecki, P., Horodecki, R.: Universal quantum information compression. Phys. Rev. Lett. **81**, 1714 (1998)
15. Hayashi, M.: Exponents of quantum fixed-length pure state source coding. Phys. Rev. A **66**, 032321 (2002)
16. Hayashi, M., Matsumoto, K.: Quantum universal variable-length source coding. Phys. Rev. A **66**, 022311 (2002)
17. Verdú, S., Han, T.S.: A general formula for channel capacity. IEEE Trans. Inform. Theory **40**, 1147–1157 (1994)
18. Hayashi, M.: Asymptotics of quantum relative entropy from a representation theoretical viewpoint. J. Phys. A: Math. and Gen. **34**, 3413–3419 (2001)
19. Keyl, M., Werner, R.F.: Estimating the spectrum of a density operator. Phys. Rev. A **64**, 052311 (2001)
20. Hayashi, M.: Optimal sequence of POVMs in the sense of Stein's lemma in quantum hypothesis. J. Phys. A: Math. and Gen. **35**, 10759–10773 (2002)
21. Bjelaković, I., Deuschel, J.-D., Kruger, T., Seiler, R., Siegmund-Schultze, R., Szkoła, A.: A quantum version of sanov's theorem. Commun. Math. Phys. **260**, 659–671 (2005)
22. Matsumoto, K., Hayashi, M.: Universal distortion-free entanglement concentration. Phys. Rev. A **75**, 062338 (2007)
23. Christandl, M.: *The structure of bipartite quantum states - insights from group theory and cryptography*. PhD thesis, February, University of Cambridge, http://arxiv.org/abs/quant-ph/0604183, 2006
24. Krattenthaler, C., Slater, P.: Asymptotic redundancies for universal quantum coding. IEEE Trans. Inform. Theory **46**, 801–819 (2000)
25. Clarke, B.S., Barron, A.R.: Information-theoretic asymptotics of Bayes methods. IEEE Trans. Inform. Theory **36**, 453–471 (1990)
26. Hayashi, M.: *Universal approximation of multi-copy states and universal quantum lossless data compression*. http://arxiv.org/abs/:0806.1091v2[quant-ph], 2008
27. Hayashi, M.: Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding. Phys. Rev. A **76**, 062301 (2007)
28. Bjelakovic, I., Boche, H.: *Classical capacities of averaged and compound quantum channels*. http://arxiv.org/abs/0710.3027v2[quant-ph], 2009

Communicated by M.B. Ruskai