# Quantum universal variable-length source coding

Masahito Hayashi[1,*] and Keiji Matsumoto[2,†]
[1]*Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan*
[2]*Quantum Computation and Information Project, ERATO, JST, 5-28-3, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
(Received 1 February 2002; revised manuscript received 17 April 2002; published 26 August 2002)

We construct an optimal quantum universal variable-length code that achieves the admissible minimum rate, i.e., our code is used for any probability distribution of quantum states. Its probability of exceeding the admissible minimum rate exponentially goes to 0. Our code is optimal in the sense of its exponent. In addition, its average error asymptotically tends to 0.

DOI: 10.1103/PhysRevA.66.022311 PACS number(s): 03.67.Hk

## I. INTRODUCTION

As was proven by Schumacher [1] and Jozsa and Schumacher [2], we can compress the unknown source state into the length $nH(\overline{\rho}_p)$ with a sufficiently small error when the source state on $n$ quantum systems obeys the $n$-i.i.d. (independently and identically distributed) distribution of the known probability $p$, where $\overline{\rho}_p := \Sigma_\rho p(\rho)\rho$ and $H(\rho)$ is the von Neumann entropy $-\mathrm{Tr}\rho\log\rho$. Jozsa and Schumacher's protocol depends on the mixture state $\overline{\rho}_p$. Concerning the quantum source coding, there are two settings: blind coding, in which the input is the unknown quantum state, and visible coding, in which the input is classical information which determines the quantum state that we want to send, i.e., the encoder knows the input quantum state. In this paper, we treat only blind coding. In our setting, we allow mixed states as input states.

In blind coding, Koashi and Imoto [3] proved that even if we allow mixed states as input states without trivial redundancies, the minimum admissible length is $nH(\overline{\rho}_p)$. Depending only on the coding length $nR$, Jozsa *et al.* [4] constructed a code which is independent of the distribution which the input obeys. In their protocol, if and only if the minimum admissible length of the distribution $p$ is less than $nR$, we can decode with a sufficiently small error. This kind of code is called a quantum universal fixed-length source code.

In the classical system, depending on the input state, the encoder can determine the coding length. Such a code is called a variable-length code. Using this type of code, we can compress any information without error. When we suitably choose a variable-length code for the probability distribution $p$ of the input, the coding length is less than $nH(p)$, except for a small enough probability. In particular, Lynch [5] and Davisson [6] proposed a variable-length code with no error, in which the coding length is less than $nH(p)$ except for a small enough probability under the distribution $p$. Such a code is called a universal variable-length source code. Today, their code can be regarded as the following two-stage code: at the first step, we send the empirical distribution (i.e., the type) which indicates a subset of data, and in the second step, we send information which indicates every sequence belonging to the subset [7].

This paper deals with quantum data compression in which the encoder determines the coding length, according to the input state. In order to make this decision, he must measure the input quantum system. After this measurement, depending on the data, the encoder compresses the final state of this measurement and sends its data and the compressed state. This type of code is called a *quantum variable-length source code*. However, in general, the encoder knows only that the input state is written as a separable state $\rho_{x_1}\otimes\rho_{x_2}\otimes\cdots\otimes\rho_{x_n}$. Therefore, it is impossible to determine the coding length without destruction of the input state.

In particular, independently of the probability distribution $p$, we construct the code satisfying the following conditions: the average error concerning Bures' distance tends to 0. The probability that the coding length is greater than $nH(\overline{\rho}_p)$ tends to 0. Such a code is called a *quantum universal variable-length source code*. In our construction, similarly to Keyl and Werner [8], an essential role is played by the representation theory of the special unitary group and the symmetric group on the tensored space. In our code, the encoder performs a quantum measurement closely related to irreducible decomposition of the two groups, and its resulting data can be approximately regarded as a quantum analogue of type. Thus, our code can be regarded as a quantum analogue of Lynch-Davisson code [5,6]. Of course, if we can estimate the entropy $H(\overline{\rho}_p)$, we can compress the coding rate to the admissible rate $H(\overline{\rho}_p)$ with a probability close to 1. However, when we perform a naive measurement for the estimation of $H(\overline{\rho}_p)$, the input state is destroyed. Therefore, in our code, it is the main problem to treat the trade-off between the estimation of $H(\overline{\rho}_p)$ and the nondemolition of the input state.

One might consider that the universal variable code can be easily realized as follows. First, use the $n\epsilon$ (where $\epsilon$ is small) states for the estimation of $H(\overline{\rho}_p)$. Second, apply Jozsa *et al.* protocol [4] by setting $R = H(\overline{\rho}_p) + \epsilon$, and apply to $n(1-\epsilon)$ states. If we consider individual error (24), this code successfully compresses the source. However, in our

*Electronic address: masahito@brain.riken.go.jp
†Electronic address: keiji@qci.jst.go.jp

 

paper, like Jozsa *et al.* [4], we consider the total Bures' distance (1) between the input state and the output state. In this criterion, ''naive estimate and compress'' strategy destroys the input state a lot. The details will be discussed in Sec. VI. [Note also that our criterion (1) is different from Krattenthanler and Slater's criterion [9] and Schumacher and Westmoreland's criterion [10].]

In this paper, we discuss the universality for the probability family $\mathcal{P}$ consisting of predicted probabilities on $\mathcal{S}(\mathcal{H})$. For any probability family $\mathcal{P}$ on $\mathcal{S}(\mathcal{H})$, we define universality of a quantum variable-length source code and evaluate the exponent of the probability that the coding length is greater than the minimum admissible length, which is called the overflow probability. However, unfortunately, in our approach, it is difficult to construct a quantum universal variable-length source code whose error exponentially tends to 0 in the blind setting. In the visible coding case, it is possible to construct such a code. This topic will be discussed in another paper.

We summarize quantum fixed-length source coding in Sec. II. After this summary, we state our mathematical setting and the main results in Sec. III. Our proofs and our construction of code are given in Secs. V and IV. Moreover, as is demonstrated in Sec. VI, in the two-dimensional case, a naive code destroys the state and is not used as a quantum universal variable-length source code.

## II. SUMMARY OF QUANTUM FIXED-LENGTH SOURCE CODING

Let $\mathcal{H}$ be a finite-dimensional Hilbert space that represents the physical system of interest and let $\mathcal{S}(\mathcal{H})$ be the set of density operators on $\mathcal{H}$. Consider a source of quantum state which produces the state $\vec{\rho}_n := \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n$ with probability the i.i.d. distribution $p^n$ of the probability $p$ on $\mathcal{S}(\mathcal{H})$. In *fixed-length source coding*, a sequence of states $\vec{\rho}_n$ is compressed to the state in a smaller Hilbert space $\mathcal{H}_n \subset \mathcal{H}^{\otimes n}$, whose dimension is $e^{nR}$. Here, the encoder and the decoder is a trace-preserving completely positive (TP-CP) map $E^n$ and $D^n$, respectively. The average of the total error is given by

$$\epsilon_{n,p}(E^n, D^n) := \sum_{\vec{\rho}_n \in \mathcal{S}(\mathcal{H}^{\otimes n})} p^n(\vec{\rho}_n) b^2(\vec{\rho}_n, D^n \circ E^n(\vec{\rho}_n)), \quad (1)$$

where Bures' distance is defined as

$$b(\rho, \sigma) := \sqrt{1 - \mathrm{Tr} |\sqrt{\rho}\sqrt{\sigma}|}.$$

Note that the support of $p$ does not necessarily consist of pure states. In this setting, we focus the infimum of the rate with which the average error goes to zero. The infimum is called the minimum admissible rate $R_p$ of $p$, and is defined by

$$R_p := \inf \left\{ \limsup \frac{1}{n} \log \dim \mathcal{H}_n \middle| \exists \{(\mathcal{H}_n, E^n, D^n)\}, \quad \epsilon_{n,p}(E^n, D^n) \to 0 \right\}.$$

The number $nR_p$ is the called minimum admissible length. When the source has no trivial redundancy in the sense following, it is calculated as

$$R_p = H(\overline{\rho}_p) := -\mathrm{Tr} \overline{\rho}_p \log \overline{\rho}_p,$$

where $\overline{\rho}_p := \Sigma_{\rho \in \mathcal{S}(\mathcal{H})} p(\rho) \rho$. The direct part was proven by Schumacher [1], and Jozsa and Schumacher [2] and the converse part was proven by Barnum *et al.* [14] in the pure state case. In the mixed case, Koashi and Imoto [3] discussed this problem as follows. Indeed, if the source has trivial redundancies, we can compress up to more than the rate $H(\overline{\rho}_p)$. We consider the source to have trivial redundancy if the support $\mathcal{S}(p)$ of $p$ satisfies the following. The Hilbert space $\mathcal{H}$ is decomposed as Eq. (2) satisfying the conditions (i) and (ii):

$$\mathcal{H} = \bigoplus_l \mathcal{H}_{J,l} \otimes \mathcal{H}_{K,l}. \quad (2)$$

(i) Any element $\rho \in \mathcal{S}(p)$ is commutative with $P_l$, where $P_l$ denotes the projection to the subspace $\mathcal{H}_{J,l} \otimes \mathcal{H}_{K,l}$.

(ii) The state $\mathrm{Tr}_{\mathcal{H}_{J,l}} P_l \rho P_l / \mathrm{Tr} P_l \rho$ is independent of $\rho \in \mathcal{S}(p)$.

Precisely, we should state that the conditions (i) and (ii) hold almost everywhere for $p$. In this case, without loss of information, we can transform $\rho$ to $\Sigma_l \mathrm{Tr}_{\mathcal{H}_{K,l}} P_l \rho P_l$. When the encoder sends the state $\Sigma_l \mathrm{Tr}_{\mathcal{H}_{K,l}} P_l \rho P_l$ instead of $\rho$, the decoder can recover the state $\rho$ from the state $\Sigma_l \mathrm{Tr}_{\mathcal{H}_{K,l}} P_l \rho P_l$. This fact implies that we can compress up to the rate $H[\Sigma_\rho p(\rho) \Sigma_l \mathrm{Tr}_{\mathcal{H}_{K,l}} P_l \rho P_l]$, i.e., $R_p \leqslant H[\Sigma_\rho p(\rho) \Sigma_l \mathrm{Tr}_{\mathcal{H}_{K,l}} P_l \rho P_l]$. Koashi and Imoto also proved the opposite inequality, i.e., proved the equation

$$R_p = H\left( \sum_\rho p(\rho) \sum_l \mathrm{Tr}_{\mathcal{H}_{K,l}} P_l \rho P_l \right), \quad (3)$$

where the right-hand side (RHS) of Eq. (3) is given by the finest decomposition satisfying (i) and (ii). Following their proof, we can understand that if $\limsup (1/n) \log \dim \mathcal{H}_n < R_p$,

$$\liminf \epsilon_{n,p}(E^n, D^n) > 0, \quad (4)$$

which is called the weak converse. When the support of $p$ consists of pure states, if $\limsup(1/n)\log\dim\mathcal{H}_n < R_p = H(\overline{\rho_p})$, we obtain

$$\lim \epsilon_{n,p}(E^n, D^n) = 1, \tag{5}$$

which is called the strong converse, and was proven by Winter [15]. A more simple proof was given by Hayashi [16]. However, the strong converse in the mixed states case is an open problem. Moreover, in the pure states case, the optimal exponent of average error was treated by Hayashi [16].

## III. QUANTUM UNIVERSAL VARIABLE-LENGTH SOURCE CODING

In the variable-length case, we need to describe a quantum measurement with state evolution, by using *an instrument* consisting of a decomposition $\mathbf{E}' = \{\mathbf{E}'_\omega\}_{\omega\in\Omega}$, by CP maps from $\mathcal{S}(\mathcal{H})$ to $\mathcal{S}(\mathcal{H})$ under the condition $\Sigma_{\omega\in\Omega}\mathrm{Tr}\mathbf{E}'_\omega(\rho) = 1, \forall \rho \in \mathcal{S}(\mathcal{H})$. When we perform the instrument $\mathbf{E}' = \{\mathbf{E}'_\omega\}_{\omega\in\Omega}$ for an initial state $\rho$, we get the data $\omega$ and the final state $\mathbf{E}'_\omega(\rho)/\mathrm{Tr}\mathbf{E}'_\omega(\rho)$ with the probability $\mathrm{Tr}\mathbf{E}'_\omega(\rho)$. A quantum variable-length encoder $\mathbf{E}$ is given by a measurement process $\mathbf{E}'$ and encoding process $E''_\omega$ depending on the data $\omega$, which is a TP-CP map from $\mathcal{S}(\mathcal{H})$ to $\mathcal{S}(\mathcal{H}_\omega)$, where the Hilbert space $\mathcal{H}_\omega$ depends on the data $\omega$, as

$$\mathbf{E}_\omega = E''_\omega \circ \mathbf{E}'_\omega .$$

Therefore, any quantum variable-length encoder $\mathbf{E}$ consists of a decomposition $\mathbf{E} = \{\mathbf{E}_\omega\}_{\omega\in\Omega}$, by CP maps from $\mathcal{S}(\mathcal{H})$ to $\mathcal{S}(\mathcal{H}_\omega)$ under the condition $\Sigma_{\omega\in\Omega}\mathrm{Tr}\mathbf{E}_\omega(\rho) = 1, \forall \rho \in \mathcal{S}(\mathcal{H})$. For details about instruments, see Ozawa [11–13].

The decoder is given by a set of TP-CP maps $\mathbf{D} = \{\mathbf{D}_\omega\}_{\omega\in\Omega}$, which presents the decoding process depending on the data $\omega$. A pair of an encoder $\mathbf{E} = \{\mathbf{E}_\omega\}_{\omega\in\Omega}$ and a decoder $\mathbf{D} = \{\mathbf{D}_\omega\}_{\omega\in\Omega}$ is called a *quantum variable-length source code* on $\mathcal{H}$. The coding length is described by $\log|\Omega| + \log\dim\mathcal{H}_\omega$, which is a random variable obeying the probability $\mathrm{P}^{\mathbf{E}}_\rho(\omega) := \mathrm{Tr}\mathbf{E}_\omega(\rho)$ when the input state is $\rho$. Of course, any quantum variable-length source code can be regarded as a quantum fixed-length source code whose length is the maximum of $\log|\Omega| + \log\dim\mathcal{H}_\omega$.

When the state $\vec{\rho}_n$ on $\mathcal{H}^{\otimes n}$ obeys the i.i.d. distribution $p^n$ of the probability $p$ on $\mathcal{S}(\mathcal{H})$, the error of decoding for a variable-length code $(\mathbf{E}^n, \mathbf{D}^n)$ on $\mathcal{H}^{\otimes n}$ is evaluated by Bures' distance as

$$\sum_{\omega_n\in\Omega_n} \mathrm{Tr}\mathbf{E}^n_{\omega_n}(\vec{\rho}_n) b^2\left(\vec{\rho}_n, \mathbf{D}^n_{\omega_n}\left(\frac{\mathbf{E}^n_{\omega_n}(\vec{\rho}_n)}{\mathrm{Tr}\mathbf{E}^n_{\omega_n}(\vec{\rho}_n)}\right)\right),$$

and the average error is given by

$$\epsilon_{n,p}(\mathbf{E}^n, \mathbf{D}^n) := \sum_{\vec{\rho}_n\in\mathcal{S}(\mathcal{H}^{\otimes n})} p^n(\vec{\rho}_n) \sum_{\omega_n\in\Omega_n} \mathrm{Tr}\mathbf{E}^n_{\omega_n}(\vec{\rho}_n) \times b^2\left(\vec{\rho}_n, \mathbf{D}^n_{\omega_n}\left(\frac{\mathbf{E}^n_{\omega_n}(\vec{\rho}_n)}{\mathrm{Tr}\mathbf{E}^n_{\omega_n}(\vec{\rho}_n)}\right)\right). \tag{6}$$

In this case, the data $\omega_n$ obey the probability

$$\mathrm{P}^{\mathbf{E}^n}_{p^n}(\omega_n) := \sum_{\vec{\rho}_n\in\mathcal{S}(\mathcal{H}^{\otimes n})} p^n(\vec{\rho}_n)\mathrm{Tr}\mathbf{E}^n_{\omega_n}(\vec{\rho}_n) = \mathrm{Tr}\mathbf{E}^n_{\omega_n}(\overline{\rho_p}^{\otimes n}). \tag{7}$$

A sequence $\{(\mathbf{E}^n, \mathbf{D}^n)\}$ of quantum variable-length source code is called *universal* for a probability family $\mathcal{P}$ on $\mathcal{S}(\mathcal{H})$ if

$$\epsilon_{n,p}(\mathbf{E}^n, \mathbf{D}^n) \to 0$$

for any probability $p \in \mathcal{P}$.

As guaranteed by Theorem 1, we can reduce the coding rate to the admissible rate $H(\overline{\rho_p})$ with a sufficiently small error and a probability infinitely close to 1, asymptotically, i.e., there exists a quantum universal variable-length source code $\{(\mathbf{E}^n, \mathbf{D}^n)\}$ satisfying that

$$\lim \mathrm{P}^{\mathbf{E}^n}_{p^n}\left\{\frac{1}{n}(\log|\Omega_n| + \log\dim\mathcal{H}_{\omega_n}) \geqslant H(\overline{\rho_p}) + \epsilon\right\} = 0, \qquad \forall \epsilon > 0, \forall p \in \mathcal{P}. \tag{8}$$

Conversely, if a quantum variable-length source code $\{(\mathbf{E}^n, \mathbf{D}^n)\}$ is universal for a family $\mathcal{P}$ and

$$\lim \mathrm{P}^{\mathbf{E}^n}_{p^n}\left\{\frac{1}{n}(\log|\Omega_n| + \log\dim\mathcal{H}_{\omega_n}) \geqslant R\right\} = 0, \tag{9}$$

then $R \geqslant R_p$ because the inequality (9) implies the existence of a fixed-length code with the rate $R$ and an asymptotically small error. When two probabilities $p, q \in \mathcal{P}$ satisfy that $\overline{\rho_p} = \overline{\rho_q}$, Eq. (7) guarantees that $\mathrm{P}^{\mathbf{E}^n}_{p^n} = \mathrm{P}^{\mathbf{E}^n}_{q^n}$. Thus, any quantum universal variable-length source code $\{(\mathbf{E}^n, \mathbf{D}^n)\}$ satisfies the inequality

$$\inf\left\{R \,\middle|\, \lim \mathrm{P}^{\mathbf{E}^n}_{p^n}\left\{\frac{1}{n}(\log|\Omega_n| + \log\dim\mathcal{H}_{\omega_n}) \geqslant R\right\} = 0\right\} \geqslant \sup_{q\in\mathcal{P}:\overline{\rho_p}=\overline{\rho_q}} R_q .$$

Therefore, the inequalities

$$H(\overline{\rho_p}) \geqslant \sup_{\{(\mathbf{E}^n,\mathbf{D}^n)\}:\text{univ. for } \mathcal{P}} \inf\left\{ R \left| \lim \mathrm{P}_{p^n}^{\mathbf{E}^n}\left\{ \frac{1}{n}(\log|\Omega_n| + \log\dim \mathcal{H}_{\omega_n}) \geqslant R \right\} = 0 \right\} \geqslant \sup_{q\in\mathcal{P}:\overline{\rho_p}=\overline{\rho_q}} R_q \right. \tag{10}$$

hold. When the support of $p$ consists of pure states, since the admissible rate $R_p$ equals $H(\overline{\rho_p})$, the RHS of (10) equals $H(\overline{\rho_p})$, i.e., our code is optimal. However, in the mixed states case, the admissible rate $R_p$ of a probability $p$ is rarely less than $H(\rho_p)$. [See Eq. (3).] In this rare case, our code cannot go up to the admissible rate $R_p$. When for any $\rho \in \mathcal{S}(\mathcal{H})$ there exists a probability $q \in \mathcal{P}$ such that $\overline{\rho_q} = \rho$ and $R_q = H(\overline{\rho_q})$, the RHS of (10) equals $H(\overline{\rho_p})$, although the admissible rate $R_p$ is less than $H(\overline{\rho_p})$. In this case, our code is optimal for any probability $p \in \mathcal{P}$.

Next, we discuss the exponent of the overflow probability: $\mathrm{P}_{p^n}^{\mathbf{E}^n}\{1/n(\log|\Omega_n| + \log\dim \mathcal{H}_{\omega_n}) \geqslant R\}$.

*Theorem 1.* For any family $\mathcal{P}$, there exists a quantum variable-length source code $\{(\mathbf{E}^n,\mathbf{D}^n)\}$ on $\mathcal{H}^{\otimes n}$ which satisfies the condition that $\epsilon_{n,p}(\mathbf{E}^n,\mathbf{D}^n)$ tends to 0 uniformly for $p \in \mathcal{P}$ and that

$$\lim \frac{-1}{n} \log \mathrm{P}_{p^n}^{\mathbf{E}^n}\left\{ \frac{1}{n}(\log|\Omega_n| + \log\dim \mathcal{H}_{\omega_n}) \geqslant R \right\} = \inf_{q\in\mathcal{P}:H(\overline{\rho_q})\geqslant R} \min_{V:\text{unitary}} D(\overline{\rho_q} \| V\overline{\rho_p}V^*), \tag{11}$$

where $D(\rho\|\sigma)$ is quantum relative entropy $\mathrm{Tr}\,\rho(\log\rho - \log\sigma)$.

Of course, when the set $\mathcal{S} := \{\overline{\rho_p} | p \in \mathcal{P}\}$ is unitary invariant, the RHS equals $\inf_{q\in\mathcal{P}:H(\overline{\rho_q})\geqslant R} D(\overline{\rho_q}\|\overline{\rho_p})$. We construct a quantum variable-length source code satisfying Eq. (11) in Sec. IV. Indeed, as is guaranteed by the following theorem, our code is optimal in the sense of the exponent of the decreasing rate of the overflow probability when $\inf_{q\in\mathcal{P}:H(\overline{\rho_q})\geqslant R} \min_{V:\text{unitary}} D(\overline{\rho_q}\|V\overline{\rho_p}V^*) = \inf_{q\in\mathcal{P}:R_q>R} D(\overline{\rho_q}\|\overline{\rho_p})$.

*Theorem 2.* If a sequence $\{(\mathbf{E}^n,\mathbf{D}^n)\}$ of quantum variable-length source codes on $\mathcal{H}^{\otimes n}$ is universal for a family $\mathcal{P}$, then

$$\limsup \frac{-1}{n} \log \mathrm{P}_{p^n}^{\mathbf{E}^n}\left\{ \frac{1}{n}(\log|\Omega_n| + \log\dim \mathcal{H}_{\omega_n}) \geqslant R \right\} \leqslant \inf_{q\in\mathcal{P}:R_q>R} D(\overline{\rho_q}\|\overline{\rho_p}). \tag{12}$$

Of course, when the family consists of all probabilities on $\mathcal{S}(\mathcal{H})$, the RHS of Eq. (11) and the RHS of (12) coincide, i.e., our code is optimal in the sense of the exponent of the overflow probability.

## IV. CONSTRUCTION OF A QUANTUM VARIABLE-LENGTH SOURCE CODE

First, we construct a universal quantum variable-length source code that achieves the optimal rate (11) for the family of all probabilities on $\mathcal{S}(\mathcal{H})$. This family is covariant for the actions of the $d$-dimensional special unitary group $\mathrm{SU}(d)$, and any $n$-i.i.d. distribution $p^n$ is invariant for the action of the $n$th symmetric group $S_n$ on the tensored space $\mathcal{H}^{\otimes n}$. Thus, our code should satisfy the invariance for these actions on $\mathcal{H}^{\otimes n}$.

Now, we focus on the irreducible decomposition of the tensored space $\mathcal{H}^{\otimes n}$ concerning the representations of $S_n$ and $\mathrm{SU}(d)$, and define the Young index $\mathbf{n}$ as

$$\mathbf{n} := (n_1, \ldots, n_d), \qquad \sum_{i=1}^{d} n_i = n, n_i \geqslant n_{i+1}$$

and denote the set of Young indices $\mathbf{n}$ by $Y_n$. Young index $\mathbf{n}$ uniquely corresponds to the irreducible unitary representation of $S_n$ and the one of $\mathrm{SU}(d)$. Now, we denote the representation space of the irreducible unitary representation of $S_n$ [$\mathrm{SU}(d)$] corresponding to $\mathbf{n}$ by $\mathcal{V}_\mathbf{n}$ ($\mathcal{U}_\mathbf{n}$), respectively. In particular, regarding a unitary representation of $\mathrm{SU}(d)$, Young index $\mathbf{n}$ gives the highest weight of the corresponding representation. Then, the tensored space $\mathcal{H}^{\otimes n}$ is decomposed as follows; i.e., $\mathcal{H}^{\otimes n}$ is equivalent with the following direct sum space under the representation of $S_n$ and $\mathrm{SU}(d)$:

$$\mathcal{H}^{\otimes n} = \bigoplus_{\mathbf{n}} \mathcal{W}_\mathbf{n}, \qquad \mathcal{W}_\mathbf{n} := \mathcal{U}_\mathbf{n} \otimes \mathcal{V}_\mathbf{n}.$$

For details, see Weyl [18], Goodman and Wallach [19], and Iwahori [20]. The efficiency of this representation method was discussed from several viewpoints. Regarding fixed-length source coding, it was discussed by Jozsa *et al.* [4]. Regarding quantum relative entropy, it was discussed by Hayashi [21]. Regarding quantum hypothesis testing, it was discussed by Hayashi [22]. Regarding estimation of the spectrum, it was discussed by Keyl and Werner [8].

In the following, for an intuitive explanation of our construction, we naively construct a good variable-length code in the case $\mathcal{H} = \mathbb{C}^2$. For this construction, we fixed a strictly increasing sequence $\vec{a} := \{a_i\}_{i=1}^{l+1}$ of real numbers such that $\frac{1}{2} = a_1 < a_2 < \cdots < a_l < a_{l+1} = 1$. We define the encoder $\mathbf{E}^{\vec{a},n}$ with the data set $\{1, \ldots, l\}$ by

$$\mathcal{H}_i^{\vec{a},n} := \bigoplus_{n_1/n \in [a_i, a_{i+1})} \mathcal{W}_\mathbf{n}, \qquad i = 1, \ldots, l-1,$$

$$\mathcal{H}_l^{\vec{a},n} := \bigoplus_{n_1/n \in [a_l, a_{l+1}]} \mathcal{W}_\mathbf{n},$$

$$\mathbf{E}_i^{\vec{a},n}(\rho_n) := P_i^{\vec{a},n}\rho_n P_i^{\vec{a},n}, \qquad \rho_n \in \mathcal{S}(\mathcal{H}^{\otimes n})$$

and define the decoder $\mathbf{D}_i^{\vec{a},n}$ as the embedding from $\mathcal{H}_i^{\vec{a},n}$ to $\mathcal{H}^{\otimes n}$, where we denote the projection to $\mathcal{H}_i^{\vec{a},n}$ by $P_i^{\vec{a},n}$. Assume that the larger eigenvalue of the mixture $\overline{\rho_p}$ belongs to the interval $[a_i, a_{i+1})$. As is guaranteed by Lemma 5 in Appendix A, if the larger eigenvalue of the mixture $\overline{\rho_p}$ does not lie on the boundary on the interval $[a_i, a_{i+1})$, the probability $\mathrm{Tr}\overline{\rho_p}^{\otimes n} P_i^{\vec{a},n}$ tends to 1. Thus, we can prove $\epsilon_{n,p}(\mathbf{E}^{\vec{a},n}, \mathbf{D}^{\vec{a},n}) \to 0$. Its speed depends on the divergence between the probability and the boundary. Of course, if we choose $a_{i+1} - a_i$ to be sufficiently small, the coding length is close to the entropy $H(\overline{\rho_p})$ with almost probability 1. However, when the larger eigenvalue lies on the boundary, the state is demolished, as is caused by the same reason of Lemma 2. In this case, similarly to Lemma 2, we can prove

$$\lim \epsilon_{n,p}(\mathbf{E}^{\vec{a},n}, \mathbf{D}^{\vec{a},n}) > 0.$$

Now, we assume that the interval $a_{i+1} - a_i$ $(i = 2, \ldots, l-1)$ is $\delta := 1/[2(l-1)]$ and that $a_2 - a_1, a_{l+1} - a_l < \delta$. Then, our code is uniquely defined by the choice of $a_2 \in (\frac{1}{2}, \frac{1}{2} + \delta)$. For the nondemolition of initial states, we construct a variable-length code, by choosing $a_2 \in \{k/n | k/n \in (\frac{1}{2}, \frac{1}{2} + \delta), k \in \mathbb{Z}\}$ at random. In this protocol, we can expect that the average error tends to 0 for any probability $p$ on $\mathcal{S}(\mathbb{C}^2)$. Note that the set $\{k/n | k/n \in (\frac{1}{2}, \frac{1}{2} + \delta), k \in \mathbb{Z}\} \times \{1, 2, \ldots, l, l+1\}$ corresponds to the data set $\Omega_n$. In order to achieve the rate $H(\overline{\rho_p})$, we need to choose the set $\Omega_n$ so that $(1/n)\log|\Omega_n| \to 0$. It is essential in our code to restrict $a_2$ to this lattice $\{k/n | k \in \mathbb{Z}\}$.

Moreover, when $\delta$ is large for a fixed number $n$, the demolition of initial state seems small and the coding length seems long. Therefore, roughly speaking, in this code for a finite number $n$, by choosing $\delta$, we can treat the trade-off between the coding length and the nondemolition of the input state.

Next, we generalize the above code to the $d$-dimensional case, and evaluate its average error. In order to satisfy the universality and the condition (11), we need to choose $\delta$ depending on $n$ more carefully. For $\delta > 0$, we define a subset $Y_{\delta,n}$ of $\mathbb{Z}^d$ as

$$Y_{\delta,n} := \left\{ \mathbf{k} \in \mathbb{Z}^d \middle| \sum_{i=1}^d k_i = n, \exists \ \mathbf{n} \in Y_n \cap U_{\mathbf{k},n\delta} \right\},$$

and define an operator $M_{\mathbf{k}}^{\delta,n}$ for any element $\mathbf{k} \in Y_{\delta,n}$ as

$$M_{\mathbf{k}}^{\delta,n} := \frac{1}{C_{1,d}(n\delta)} P_{\mathbf{k}}^{\delta,n},$$

$$P_{\mathbf{k}}^{\delta,n} := \sum_{\mathbf{n} \in Y_n \cap U_{\mathbf{k},n\delta}} P_{\mathbf{n}},$$

$$U_{\mathbf{p},\delta} := \{\mathbf{q} \in \mathbb{R}^d | \|\mathbf{p} - \mathbf{q}\| \leqslant \delta\},$$

$$C_{1,d}(x) := \mathcal{N}\left\{ \mathbf{k} \in \mathbb{Z}^d \middle| \|\mathbf{k}\| \leqslant x, \sum_{i=1}^d k_i = 0 \right\},$$

where $P_{\mathbf{n}}$ denotes the projection to $\mathcal{W}_{\mathbf{n}}$ and $\mathcal{N}$ denotes the member of elements.

The number $\mathcal{N}\{\mathbf{k} \in \mathbb{Z}^d \cap U_{\mathbf{n},n\delta} | \Sigma_{i=1}^d k_i = n\}$ is independent of $\mathbf{n} \in Y_n$ and equals $C_{1,d}(n\delta)$. Thus, we have the relations

$$P_{\mathbf{n}} \sum_{\mathbf{k} \in Y_{\delta,n}} M_{\mathbf{k}}^{\delta,n} P_{\mathbf{n}} = \frac{\mathcal{N}\{\mathbf{k} \in Y_{\delta,n} | \mathbf{n} \in Y_n \cap U_{\mathbf{k},n\delta}\}}{C_{1,d}(n\delta)} P_{\mathbf{n}} = P_{\mathbf{n}},$$

which implies the condition

$$\sum_{\mathbf{k} \in Y_{\delta,n}} M_{\mathbf{k}}^{\delta,n} = I.$$

The encoder $\mathbf{E}^{\delta,n}$ whose data set is $Y_{\delta,n}$ is defined by

$$\mathcal{H}_{\mathbf{k}}^{\delta,n} := \bigoplus_{\mathbf{n} \in Y_n : \|\mathbf{n} - \mathbf{k}\| \leqslant n\delta} \mathcal{W}_{\mathbf{n}},$$

$$\mathbf{E}_{\mathbf{k}}^{\delta,n}(\rho_n) := \sqrt{M_{\mathbf{k}}^{\delta,n}} \rho_n \sqrt{M_{\mathbf{k}}^{\delta,n}}, \qquad \forall \rho_n \in \mathcal{S}(\mathcal{H}^{\otimes n}),$$

and the decoder $\mathbf{D}_{\mathbf{k}}^{\delta,n}$ is defined as the embedding from $\mathcal{H}_{\mathbf{k}}^{\delta,n}$ to $\mathcal{H}^{\otimes n}$.

As is proven in Appendixes B and C, the quantum variable-length source code $(\mathbf{E}^{\delta,n}, \mathbf{D}^{\delta,n})$ on $\mathcal{H}^{\otimes n}$ satisfies

$$\epsilon_{n,p}(\mathbf{E}^{\delta,n}, \mathbf{D}^{\delta,n}) \leqslant \inf_{\delta_1 : 0 < \delta_1 < \delta} 1 - \frac{C_{2,d}(n\delta_1)}{C_{1,d}(n\delta)} [1 - (n+d)^{4d} \exp(-n C_{3,d}(\delta - \delta_1)^2)]^{3/2}, \tag{13}$$

$$\frac{-1}{n} \log \mathrm{P}_{p^n}^{\mathbf{E}^{\delta,n}} \left\{ \frac{1}{n} (\log|Y_{\delta,n}| + \log \dim \mathcal{H}_{\mathbf{k}}^{\delta,n}) \geqslant R \right\} \geqslant \frac{-5d}{n} \log(n+d) + \inf_{\mathbf{q} \in \mathbb{R}_+^{d,1} : H(\mathbf{q}) \geqslant R - (4d/n)\log(n+d)} \left( \inf_{\mathbf{q}' \in \mathbb{R}_+^{d,1} : \|\mathbf{q} - \mathbf{q}'\| \leqslant 2\delta} D(\mathbf{q}' \| \mathbf{p}) \right), \tag{14}$$

where

$$C_{2,d}(x) := \min_{\mathbf{p}\in\mathbb{R}^d:\sum_i p_i=0} \mathcal{N}\left\{ \mathbf{k}\in\mathbb{Z}^d \middle| \|\mathbf{k}-\mathbf{p}\|\leqslant x, \sum_{i=1}^d k_i=0 \right\},$$

$$C_{3,d} := \min_{\mathbf{q},\mathbf{p}\in\mathbb{R}_+^{d,1}} \frac{D(\mathbf{q}\|\mathbf{p})}{\|\mathbf{p}-\mathbf{q}\|^2}, \tag{15}$$

$$\mathbb{R}_+ := \{x\in\mathbb{R} | x\geqslant 0\}, \qquad \mathbb{R}_+^{d,1} := \left\{ \mathbf{p}\in\mathbb{R}_+^d \middle| \sum_i p_i = 1 \right\},$$

and $\mathbf{p}\in\mathbb{R}_+^{d,1}$ denotes the probability $(p_1,p_2,\ldots,p_d)$, where $p_i$ is the eigenvalue of $\overline{\rho_p}$ and $p_1\geqslant p_2\geqslant\cdots\geqslant p_d$. In this paper, we use an italic letter $p$ to denote a probability on $\mathcal{S}(\mathcal{H})$ while we use a bold letter $\mathbf{p}$ to denote a probability $(p_1,\ldots,p_d)$ on $\{1,\ldots,d\}$. Note that the RHS of (13) is independent of $p$. Our main point is simultaneously reducing $\epsilon_{n,p}(\mathbf{E}^{\delta,n},\mathbf{D}^{\delta,n})$ and $\mathrm{P}_{p^n}^{\mathbf{E}^{\delta,n}}\{1/n(\log|Y_{\delta,n}|+\log\dim\mathcal{H}_{\mathbf{k}}^{\delta,n})\geqslant R\}$. The RHS of (14) decreases as $\delta$ increases while the relation between the RHS of (13) and $\delta$ is not necessarily simple. However, letting $\delta := n^{-1/4}$ and $\delta_1 := n^{-1/4}-n^{-1/3}$, we can check that the RHS of (13) tends to 0, and that the RHS of (14) tends to the RHS of (11). Thus, we obtain theorem 1 when $\mathcal{P}$ consists of all probabilities on $\mathcal{S}(\mathcal{H})$.

If we adopt another criterion,

$$\epsilon''_{n,p}(\mathbf{E}^n,\mathbf{D}^n) := \sum_{\vec{\rho}_n\in\mathcal{S}(\mathcal{H}^{\otimes n})} p^n(\vec{\rho}_n) \sum_{\omega_n\in\Omega_n} \mathrm{Tr}\mathbf{E}^n_{\omega_n}(\vec{\rho}_n)\left\{ 1-\left[ \mathrm{Tr}\left| \vec{\rho}_n D_{\omega_n}\left( \frac{\mathbf{E}^n_{\omega_n}(\vec{\rho}_n)}{\mathrm{Tr}\mathbf{E}^n_{\omega_n}(\vec{\rho}_n)} \right) \right| \right]^2 \right\},$$

we have the following inequality instead of (13):

$$\epsilon''_{n,p}(\mathbf{E}^{\delta,n},\mathbf{D}^{\delta,n}) \leqslant \inf_{\delta_1:0<\delta_1<\delta} 1-\frac{C_{2,d}(n\delta_1)}{C_{1,d}(n\delta)}[1-(n+d)^{4d}\exp(-nC_{3,d}(\delta-\delta_1)^2)]^2, \tag{16}$$

which is proven in Appendix C.

Next, deforming the code $(\mathbf{E}^{\delta,n},\mathbf{D}^{\delta,n})$, we construct a universal quantum variable-length source code that achieves the optimal rate in the general case with no trivial redundancy. Define the set $Y_{\delta,\delta_1,n}(\mathcal{S})$ as

$$Y_{\delta,\delta_1,n}(\mathcal{S}) := \left\{ \mathbf{k}\in Y_{\delta,n} \middle| \exists\ \ \rho\in\mathcal{S}, \qquad \left\| \mathbf{p}(\rho)-\frac{\mathbf{k}}{n} \right\| \leqslant \delta_1 \right\},$$

where $\mathbf{p}(\rho)$ consists of eigenvalues of $\rho$ such that $p_1(\rho)\geqslant\cdots\geqslant p_d(\rho)$. In particular, $\mathbf{p}=\mathbf{p}(\overline{\rho_p})$. Note that $\mathcal{S}$ is defined after theorem 1, and is different from $\mathcal{S}(p)$. When the data $\mathbf{k}$ belong to $Y_{\delta,\delta_1,n}(\mathcal{S})$, we send the state $\mathbf{E}^{\delta,n}_{\mathbf{k}}(\vec{\rho}_n)/\mathrm{Tr}\mathbf{E}^{\delta,n}_{\mathbf{k}}(\vec{\rho}_n)$. Otherwise, we send only the classical information 0, except for $Y_{\delta,\delta_1,n}(\mathcal{S})$. Then, the data set of the encoder is $Y_{\delta,\delta_1,n,+}(\mathcal{S}) := Y_{\delta,\delta_1,n}(\mathcal{S})\cup\{0\}$. The decoder is defined as

$$\mathbf{D}^{\delta,\delta_1,n,\mathcal{S}}_{\mathbf{k}} := \mathbf{D}^{\delta,n}_{\mathbf{k}}, \qquad \forall\mathbf{k}\in Y_{\delta,\delta_1,n}(\mathcal{S}).$$

As is proven in Appendixes B and C, the quantum variable-length source code $(\mathbf{E}^{\delta,\delta_1,n,\mathcal{S}}\mathbf{D}^{\delta,\delta_1,n,\mathcal{S}})$ on $\mathcal{H}^{\otimes n}$ satisfies

$$\epsilon_{n,p}(\mathbf{E}^{\delta,\delta_1,n,\mathcal{S}},\mathbf{D}^{\delta,\delta_1,n,\mathcal{S}}) \leqslant 1-\frac{C_{2,d}(n\delta_1)}{C_{1,d}(n\delta)}[1-(n+d)^{4d}\exp(-nC_{3,d}(\delta-\delta_1)^2)]^{3/2}, \tag{17}$$

$$\frac{-1}{n}\log \mathrm{P}_{p^n}^{\mathbf{E}^{\delta,\delta_1,n,\mathcal{S}}}\left\{ \frac{1}{n}[\log|Y_{\delta,\delta_1,n,+}(\mathcal{S})|+\log\dim\mathcal{H}_{\mathbf{k},n,\delta}]\geqslant R \right\}$$
$$\geqslant \frac{-5d}{n}\log(n+d)+\min_{\mathbf{q}\in\mathbb{R}_+^{d,1}:H(\mathbf{q})\geqslant R-(4d/n)\log(n+d)\exists\rho\in\mathcal{S},\|\mathbf{q}-\mathbf{p}(\rho)\|\leqslant\delta_1}\left( \min_{\mathbf{q}'\in\mathbb{R}_+^{d,1}:\|\mathbf{q}-\mathbf{q}'\|\leqslant 2\delta} D(\mathbf{q}'\|\mathbf{p}) \right) \tag{18}$$

for $\forall p\in\mathcal{P}$. Note that $D(\mathrm{p}(\rho)\|\mathrm{p}(\sigma))=\min_{V:\text{unitary}}D(\rho\|V^*\sigma V)$. Letting $\delta := n^{-1/4}$ and $\delta_1 := n^{-1/4}-n^{-1/3}$, we can show that the RHS of (17) tends to 0, and that the RHS of (18) tends to the RHS of (11).

### V. OPTIMALITY OF THE EXPONENT OF THE OVERFLOW PROBABILITY

Next, we prove Theorem 2. When the support of any element $p$ of $\mathcal{P}$ consists of pure states, i.e., the pure states case, we can prove Theorem 2 by using the monotonicity of quantum relative entropy because the strong converse (5) holds in quantum fixed-length pure state source coding, as is explained in Sec. II. However, in the mixed states case, we cannot use this strategy, and we need the following lemma called the strong converse part of quantum Stein's lemma in quantum hypothesis testing proven by Ogawa and Nagaoka [23] as an alternative. Its other proof was given by Hayashi [22].

*Lemma 1.* Let $\rho$ and $\sigma$ be density operators on $\mathcal{H}$. If any sequence of operators $\vec{T}=\{T_n\}$ on $\mathcal{H}^{\otimes n}$ satisfies that $0\leqslant T_n \leqslant I$ and that $\liminf \mathrm{Tr}\rho^{\otimes n}T_n>0$, then the inequality

$$\limsup \frac{-1}{n}\log \mathrm{Tr}\sigma^{\otimes n}T_n \leqslant D(\rho\|\sigma)$$

holds.

Since the monotonicity of quantum relative entropy corresponds to the weak converse part of quantum Stein's lemma, the former strategy can be regarded as the combination of the strong converse part (5) of quantum fixed-length pure state source coding and the weak converse part of quantum Stein's lemma, and the latter proof can be regarded as the combination of the weak converse part (4) of quantum fixed-length source coding and the strong converse part of quantum Stein's lemma.

First, for the reader's convenience, we give the former proof which is simpler than the latter but is applied only in the pure states case. After this proof, we give a more sound proof which can be used in the general case. Let $p$ and $q$ be an arbitrary elements of $\mathcal{P}$, and $R$ be an arbitrary real number such that $R$ is less than the minimum admissible rate of $q$, i.e., $R<R_q$. In particular, we assume that the support of $q$ consists of pure states. For a quantum variable-length source code $\{(\mathbf{E}^n,\mathbf{D}^n)\}$ for a family $\mathcal{P}$, deforming the code $(\mathbf{E}^n,\mathbf{D}^n)$, we define the fixed-length code $(E^{R,n},D^{R,n})$ as follows. When the data $\omega_n$ satisfy

$$\log|\Omega_n|+\log\dim\mathcal{H}_{\omega_n}\geqslant nR, \tag{19}$$

we send classical information which indicates condition (19). Otherwise, we send the data $\omega_n$ and the state $\mathbf{E}^n_{\omega_n}(\vec{\rho}_n)/\mathrm{Tr}\mathbf{E}^n_{\omega_n}(\vec{\rho}_n)$. In the decoding process, if we receive the classical information which indicates condition (19), we regard a quantum state $\rho_R$ out of the original space $\mathcal{H}^{\otimes n}$ as the decoded state. Note that $b(\rho_R,\rho)\leqslant 1$ for any state $\rho \in \mathcal{S}(\mathcal{H}^{\otimes n})$. Otherwise, we perform the operation $\mathbf{D}^n_{\omega_n}$ as the decoding process. Since the maximum of this code is less than $nR$, we can regard it as a fixed-length code whose length is $nR$.

From the construction of the fixed-length code $(E^{R,n},D^{R,n})$, we can easily check that

$$\epsilon_{n,q}(E^{R,n},D^{R,n})\leqslant \mathrm{P}^{\mathbf{E}^n}_{q^n}\left\{\frac{1}{n}(\log|\Omega_n|+\log\dim\mathcal{H}_{\omega_n})<R\right\}\epsilon_{n,q}\left(\mathbf{E}^n,\mathbf{D}^n\middle|\frac{1}{n}(\log|\Omega_n|+\log\dim\mathcal{H}_{\omega_n})<R\right)$$
$$+\mathrm{P}^{\mathbf{E}^n}_{q^n}\left\{\frac{1}{n}(\log|\Omega_n|+\log\dim\mathcal{H}_{\omega_n})\geqslant R\right\},$$

where $\epsilon_{n,q}(\mathbf{E}^n,\mathbf{D}^n|1/n(\log|\Omega_n|+\log\dim\mathcal{H}_{\omega_n})<R)$ denotes the conditional average of the total error under the condition $1/n(\log|\Omega_n|+\log\dim\mathcal{H}_{\omega_n})<R$. Thus, we have the inequality

$$\epsilon_{n,q}(E^{R,n},D^{R,n})-\epsilon_{n,q}(\mathbf{E}^n,\mathbf{D}^n)\leqslant \mathrm{P}^{\mathbf{E}^n}_{q^n}\left\{\frac{1}{n}(\log|\Omega_n|+\log\dim\mathcal{H}_{\omega_n})\geqslant R\right\}\left[1-\epsilon_{n,q}\left(\mathbf{E}^n,\mathbf{D}^n\middle|\frac{1}{n}(\log|\Omega_n|+\log\dim\mathcal{H}_{\omega_n})\geqslant R\right)\right]$$
$$\leqslant \mathrm{P}^{\mathbf{E}^n}_{q^n}\left\{\frac{1}{n}(\log|\Omega_n|+\log\dim\mathcal{H}_{\omega_n})\geqslant R\right\}=\mathrm{P}^{\mathbf{E}^n}_{\overline{\rho_q}^{\otimes n}}\left\{\frac{1}{n}(\log|\Omega_n|+\log\dim\mathcal{H}_{\omega_n})\geqslant R\right\}. \tag{20}$$

Since the support of $q$ consists of pure states and $H(\overline{\rho_q})=R_q>R$, we obtain the relation

$$\epsilon_{n,q}(E^{R,n},D^{R,n})\to 1,$$

which is called the strong converse part of the quantum fixed-length pure state source coding [15]. Since the universality guarantees the relation

$$\epsilon_{n,q}(\mathbf{E}^n,\mathbf{D}^n)\to 0, \tag{21}$$

we have

$$\mathrm{P}_{n,q} := \mathrm{P}^{\mathbf{E}^n}_{\overline{\rho}_q^{\otimes n}}\left\{\frac{1}{n}(\log|\Omega_n| + \log\dim\mathcal{H}_{\omega_n}) \geqslant R\right\} \to 1. \quad (22)$$

Using the monotonicity of quantum relative entropy, we have

$$nD(\overline{\rho}_q\|\overline{\rho}_p) = D(\overline{\rho}_q^{\otimes n}\|\overline{\rho}_p^{\otimes n}) \geqslant \mathrm{P}_{n,q}\log\frac{\mathrm{P}_{n,q}}{\mathrm{P}_{n,p}} + (1-\mathrm{P}_{n,q})\log\frac{1-\mathrm{P}_{n,q}}{1-\mathrm{P}_{n,p}},$$

where we define $\mathrm{P}_{n,p}$ similarly to Eq. (22). Since $-(1-\mathrm{P}_{n,q})\log(1-\mathrm{P}_{n,p}) \geqslant 0$,

$$-\frac{\log \mathrm{P}_{n,p}}{n} \leqslant \frac{nD(\overline{\rho}_q\|\overline{\rho}_p) + h(\mathrm{P}_{n,q})}{n\mathrm{P}_{n,q}} \to D(\overline{\rho}_q\|\overline{\rho}_p),$$

where $h(x)$ is the binary entropy $-x\log x-(1-x)\log(1-x)$. Now, we obtain inequality (12) in the pure states case.

Next, we proceed the general case. It follows from (4) and the inequality $R<R_q$ that we have

$$\liminf \epsilon_{n,q}(E^{R,n}, D^{R,n}) > 0. \quad (23)$$

From (20) and (21), the relation $\liminf \mathrm{P}_{n,q} > 0$ holds. There exists a POVM (positive operator valued measure, i.e., a partition of the unity $I$ into positive Hermitian matrices) $M^n = \{M^n_{\omega_n}\}_{\omega_n}$ such that

$$\mathrm{Tr}\rho_n M^n_{\omega_n} = \mathrm{Tr}\mathbf{E}^n_{\omega_n}(\rho_n), \qquad \forall \rho_n \in \mathcal{S}(\mathcal{H}^{\otimes n}).$$

Letting

$$T_n := \sum_{\omega_n : 1/n(\log|\Omega_n| + \log\dim\mathcal{H}_{\omega_n}) \geqslant R} M^n_{\omega_n},$$

we have $\mathrm{P}_{n,q} = \mathrm{Tr}\overline{\rho}_q^{\otimes n} T_n$ and $\mathrm{P}_{n,p} = \mathrm{Tr}\overline{\rho}_p^{\otimes n} T_n$. Thus, Lemma 1 guarantees that

$$\limsup -\frac{1}{n}\log \mathrm{P}_{n,p} \leqslant D(\overline{\rho}_q\|\overline{\rho}_p).$$

Now, the proof is completed.

## VI. DISCUSSION

In our code, the nonzero number $\delta$ is essential. One may expect that the quantum variable-length source code $\{(\mathbf{E}^{0,n}, \mathbf{D}^{0,n})\}$ is universal. However, this code destroys the input state by a quantum measurement as follows.

*Lemma 2.* Assume that $d=2$ and $\{|e_1\rangle, |e_2\rangle\}$ is a CONS of $\mathbb{C}^2$. If the support of $p$ is pure states $\{|e_1\rangle\langle e_1|, |e_2\rangle\langle e_2|\}$, the average error $\epsilon_{n,p}(\mathbf{E}^{0,n}, \mathbf{D}^{0,n})$ does not tends to 0.

As is understood from our proof of theorem 1, bound (11) cannot be achieved unless $\delta$ tends to 0. It seems essential to approximate the nonzero number $\delta > 0$ to 0.

If we discuss quantum universal coding under another error $\epsilon'_{n,p}(\mathbf{E}^n, \mathbf{D}^n)$ instead of $\epsilon_{n,p}(\mathbf{E}^n, \mathbf{D}^n)$ [cf. (6)],

$$\epsilon'_{n,p}(\mathbf{E}^n, \mathbf{D}^n) := \sum_{\vec{\rho}_n \in \mathcal{S}(\mathcal{H})} p^n(\vec{\rho}_n)\frac{1}{n}\sum_{i=1}^n \sum_{\omega_n \in \Omega_n} \mathrm{Tr}\mathbf{E}^n_{\omega_n}(\vec{\rho}_n) \times b^2(\rho_i, \mathbf{E}^n_{\omega_n}(\vec{\rho}_n)_i), \quad (24)$$

$$\vec{\rho}_n = \rho_1 \otimes \cdots \otimes \rho_n,$$

$$\mathbf{E}^n_{\omega_n}(\vec{\rho}_n)_i = \mathrm{Tr}_{\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_{i-1} \otimes \mathcal{H}_{i+1} \otimes \cdots \otimes \mathcal{H}_n} \mathbf{E}^n_{\omega_n}(\vec{\rho}_n),$$

we can use several strategies for quantum universal coding. For example, if we use $n\epsilon$ states only for the estimation of $H(\overline{\rho}_p)$, we can reduce the error $\epsilon'_{n,p}$ to zero, asymptotically, by use of the Jozsa *et al.* protocol [4]. However, in this strategy, we cannot reduce the error $\epsilon_{n,p}$ because the demolition of the first $n\epsilon$ states is crucial for this criterion.

Next, we discuss how rapidly the average error $\epsilon_{n,p}$ tends to 0 in our code. Assume that $d=2$ and $\{|e_1\rangle, |e_2\rangle\}$ is a CONS (complete orthonormal system) of $\mathbb{C}^2$. Unless $\delta_n > 0$ satisfies $|\delta_n| < 1$, the coding length always equals $2n$. Then, we can assume that $|\delta_n| < 1$.

*Lemma 3.* If the support of $p$ is pure states $\{|e_1\rangle\langle e_1|, |e_2\rangle\langle e_2|\}$, the relation

$$\lim \frac{-1}{n}\log \epsilon_{n,p}(\mathbf{E}^{\delta_n, n}, \mathbf{D}^{\delta_n, n}) = 0 \quad (25)$$

holds for any sequence $\{\delta_n\}$ satisfying $|\delta_n| < 1$.

Therefore, it seems impossible to construct a universal code whose average error $\epsilon_{n,p}$ exponentially tends to 0.

In general, even if $R_p = H(\overline{\rho}_p)$ for $\forall\ p \in \mathcal{P}$, the RHS of (11) does not necessarily coincide with the RHS of (12). For example, when

$$\mathcal{P} = \{p_t | t \in (0, 1/2)\}, \qquad H(\overline{\rho}_{p_t}) = R_{p_t},$$

$$\overline{\rho}_{p_t} = \begin{pmatrix} t\cos^2\theta(t) + (1-t)\sin^2\theta(t) & (1-2t)\cos\theta(t)\sin\theta(t) \\ (1-2t)\cos\theta(t)\sin\theta(t) & (1-t)\cos^2\theta(t) + t\sin^2\theta(t) \end{pmatrix},$$

and $\theta$ is continuous and one-to-one, both sides of (10) coincide with $H(\overline{\rho_{p_t}})$ while the RHS of (11) is strictly smaller than the RHS of (12) as follows. For $t_1, t_0 \in (0,1/2)$, we can calculate as

$$\inf_{t\in(0,1/2):H(\overline{\rho_{p_t}})\geqslant h(t_1)} \min_{V:\text{unitary}} D(\overline{\rho_{p_t}}\|V\overline{\rho_{p_{t_0}}}V^*) = d(t_1,t_0),$$

$$\inf_{t\in(0,1/2):R_{p_t}>h(t_1)} D(\overline{\rho_{p_t}}\|\overline{\rho_{p_{t_0}}}) = \cos^2(\theta(t)-\theta(t))d(t_1,t_0)+\sin^2(\theta(t)-\theta(t)) \times d(t_1,1-t_0),$$

where

$$h(t) := -t\log t-(1-t)\log(1-t),$$

$$d(t,t') := t\log\frac{t}{t'}+(1-t)\log\frac{1-t}{1-t'}.$$

Thus, its difference equals $\sin^2(\theta(t)-\theta(t))[d(t_1,1-t_0) - d(t_1,t_0)]>0$. This gap is closely related to the ambiguity of the large deviation-type bounds in quantum estimation [17]. It seems very hard to match the upper bound and the lower bound concerning the exponent of the overflow probability in the general case.

## VII. CONCLUSION

We construct a quantum variable-length code satisfying Eq. (11). This is optimal in the sense of Theorem 2 when the family $\mathcal{P}$ consists of probabilities on $\mathcal{S}(\mathcal{H})$ with no trivial redundancies. However, in our code the average error does not exponentially vanish. The construction of such a code seems to be difficult.

## ACKNOWLEDGMENTS

The authors wish to thank Professor H. Nagaoka and Dr. A. Winter for useful comments.

## APPENDIX A: REPRESENTATION THE THEORETICAL TYPE METHOD

For our proof, we need the following two lemmas.

*Lemma 4.* The relation

$$\dim \mathcal{V}_{\mathbf{n}} \leqslant C(\mathbf{n})(n+d)^{2d} \leqslant (n+d)^{2d}\exp\left[nH\left(\frac{\mathbf{n}}{n}\right)\right], \tag{A1}$$

$$\mathcal{N}\{\mathbf{n}|\mathbf{n}\in Y_n\} \leqslant (n+1)^d, \tag{A2}$$

$$\dim \mathcal{U}_{\mathbf{n}} \leqslant (n+1)^d \tag{A3}$$

holds, where $C(\mathbf{n})$ is defined as

$$C(\mathbf{n}) := \frac{n!}{n_1!n_2!\cdots n_d!}.$$

*Proof.* Inequality (A2) is trivial. Using Young index $\mathbf{n}$, the basis of $\mathcal{U}_{\mathbf{n}}$ is described by $\{e_{\mathbf{n}'}\}_{\mathbf{n}'\in Y^{\mathbf{n}}}$, where $Y^{\mathbf{n}}$ is defined as

$$Y^{\mathbf{n}} := \left\{ \mathbf{n}'=\{n_i'\}\in\mathbb{Z}^d \left| \begin{array}{l} \sum_i n_i' = \sum_i n_i, \\ \sum_{i=1}^m n_{s(i)}' \leqslant \sum_{i=1}^m n_i, \\ \qquad 1\leqslant \forall m \leqslant d-1, \\ s \text{ is any permutation} \end{array} \right. \right\}.$$

Thus, we obtain (A3). Note that the correspondence $\mathbf{n}'$ and $e_{\mathbf{n}'}$ depends on the choice of Cartan subalgebra, i.e., the choice of basis of $\mathcal{H}$.

According to Weyl [18] and Iwahori [20], the following equation holds, and it is evaluated as

$$\dim \mathcal{V}_{\mathbf{n}} = \frac{n!}{(n_1+d-1)!(n_2+d-2)!\cdots n_d!} \times \prod_{j>i}(n_i-n_j-i+j) \leqslant \frac{n!}{n_1!n_2!\cdots n_d!}\prod_{j>i}(n_i-n_j-i+j) \leqslant C(\mathbf{n})(n+d)^{2d} \leqslant (n+d)^{2d}\exp\left[nH\left(\frac{\mathbf{n}}{n}\right)\right]. \quad \blacksquare$$

The following is essentially equivalent to Keyl and Werner's result [8]. For the reader's convenience, we give a simpler proof.

*Lemma 5.* Assume that $\mathbf{p}$ is the spectrum of $\rho$ such that $p_1\geqslant p_2\geqslant\cdots\geqslant p_d$. The relations

$$\operatorname{Tr} P_{\mathbf{n}}\rho^{\otimes n} \leqslant (n+d)^{3d}\exp\left[-nD\left(\frac{\mathbf{n}}{n}\middle\|\mathbf{p}\right)\right], \tag{A4}$$

$$\sum_{\mathbf{n}/n\notin\mathcal{R}} \operatorname{Tr} P_{\mathbf{n}}\rho^{\otimes n} \leqslant (n+d)^{4d}\exp\left[-n\min_{q\notin\mathcal{R}} D(\mathbf{q}\|\mathbf{p})\right] \tag{A5}$$

hold for a subset $\mathcal{R}$ of $\mathbb{R}_+^{d,1}$.

*Proof.* Let $\mathcal{U}_{\mathbf{n}}'$ be an irreducible representation of SU($d$) in $\mathcal{H}^{\otimes n}$, which is equivalent to $\mathcal{U}_{\mathbf{n}}$. We denote its projection by $P_{\mathbf{n}}'$. Now, we choose the basis $\{e_{\mathbf{n}'}\}_{\mathbf{n}'\in Y^{\mathbf{n}}}$ of $\mathcal{U}_{\mathbf{n}}'$ depending on the basis $\{e_i\}$ of $\mathcal{H}$. The base $e_{\mathbf{n}'}$ is the eigenvector of $\rho^{\otimes n}$ with the eigenvalue $\Pi_{i=1}^d a_i^{n_i'}$. Since $\mathbf{n}'$ is majorized by $\mathbf{n}$, we can calculate the operator norm by

$$\|P_{\mathbf{n}}'\rho^{\otimes n}P_{\mathbf{n}}'\| = \prod_{i=1}^d a_i^{n_i}, \tag{A6}$$

where $\|X\| := \sup_{x\in\mathcal{H}}\|Xx\|$. From (A1), (A3), and (A6), the relations

$$\mathrm{Tr}\, P_{\mathbf{n}}\rho^{\otimes n} = \dim \mathcal{V}_{\mathbf{n}} \times \mathrm{Tr}\, P'_{\mathbf{n}}\rho^{\otimes n} \leqslant (n+d)^{3d} C(\mathbf{n}) \prod_{i=1}^{d} a_i^{n_i}$$

$$= (n+d)^{3d} \mathrm{Mul}(\mathbf{a},\mathbf{n})$$

hold, where we denote the multinomial distribution of $\mathbf{a}$ by $\mathrm{Mul}(\mathbf{a},\bullet)$. Thus, we obtain (A4). Inequality (A2) guarantees

$$\sum_{\mathbf{n}/n \notin \mathcal{R}} \mathrm{Tr}\, P_{\mathbf{n}}\rho^{\otimes n} \leqslant (n+d)^{4d} \exp\left( -n \inf_{\mathbf{q} \notin \mathcal{R}} D(\mathbf{q}\|\mathbf{p}) \right). \quad \blacksquare$$

## APPENDIX B: PROOF OF (14) AND (18)

First, we prove inequality (14). For a sufficiently large integer $n$, the relations

$$|Y_{\delta,n}| \leqslant \mathcal{N}\{\mathbf{k} \in \mathbb{Z}^d | k_i \geqslant 0\} \leqslant (n+1)^d$$

hold. Since $\dim \mathcal{U}_{\mathbf{n}} \leqslant (n+d)^d$, for any $\mathbf{k} \in Y_{\delta,n}$, we have

$$\log |Y_{\delta,n}| + \log \dim \mathcal{H}_{\mathbf{k}}^{\delta,n} \leqslant d \log(n+1) + \max_{\mathbf{n} \in Y_{\delta,n} \cap U_{\mathbf{k},n\delta}} \log \dim \mathcal{U}_{\mathbf{n}} + \log \dim \mathcal{V}_{\mathbf{n}} \leqslant 4d \log(n+d) + \max_{\mathbf{n} \in Y_{\delta,n} \cap U_{\mathbf{k},n\delta}} nH\left(\frac{\mathbf{n}}{n}\right).$$

From (A5), we have

$$\mathrm{Tr} M_{\mathbf{k}}^{\delta,n} \overline{\rho_p}^{\otimes n} \leqslant \frac{|Y_{\delta,n}|}{C_{1,d}(n\delta)} (n+d)^{3d} \max_{\mathbf{n}' \in Y_n \cap U_{\mathbf{k},n\delta}} \exp\left[ -nD\left(\frac{\mathbf{n}'}{n} \middle\| \mathbf{p}\right)\right] \leqslant (n+d)^{4d} \max_{\mathbf{n}' \in Y_n \cap U_{\mathbf{k},n\delta}} \exp\left[ -nD\left(\frac{\mathbf{n}'}{n} \middle\| \mathbf{p}\right)\right]$$

$$\leqslant (n+d)^{4d} \max_{\mathbf{q} \in U_{\mathbf{k}/n,\delta} \cap \mathbb{R}_+^{d,1}} \exp[-nD(\mathbf{q}\|\mathbf{p})].$$

Thus,

$$\mathrm{P}_{\overline{\rho_p}^{\otimes n}}^{\mathbf{E}^n}\left\{ \frac{1}{n}(\log |Y_{\delta,n}| + \log \dim \mathcal{H}_{\mathbf{k}}^{\delta,n}) \geqslant R \right\}$$

$$\leqslant \sum_{\mathbf{k} \in Y_{\delta,n} : \max_{\mathbf{n} \in Y_n \cap U_{\mathbf{k},n\delta}} H(\mathbf{n}/n) \geqslant R - (4d/n)\log(n+d)} \mathrm{Tr} M_{\mathbf{k}}^{\delta,n} \overline{\rho_p}^{\otimes n}$$

$$\leqslant |Y_{\delta,n}| (n+d)^{4d} \max_{\mathbf{n} \in Y_n : H(\mathbf{n}/n) \geqslant R - (4d/n)\log(n+d)} \left[ \max_{\mathbf{n}' \in Y_n : \|\mathbf{n}-\mathbf{n}'\| \leqslant 2\delta n} \exp\left( -nD\left(\frac{\mathbf{n}'}{n} \middle\| \mathbf{p}\right)\right)\right]$$

$$\leqslant (n+d)^{5d} \max_{\mathbf{q} \in \mathbb{R}_+^{d,1} : H(\mathbf{q}) \geqslant R - (4d/n)\log(n+d)} \left[ \max_{\mathbf{q}' \in \mathbb{R}_+^{d,1} : \|\mathbf{q}-\mathbf{q}'\| \leqslant 2\delta} \exp[-nD(\mathbf{q}'\|\mathbf{p})]\right].$$

Then, we obtain (14).

Next, we proceed to (18). Since $|Y_{\delta,\delta_1,n,+}(\mathcal{S})| \leqslant |Y_{\delta,n}|$, we have

$$\mathrm{P}_{\overline{\rho_p}^{\otimes n}}^{\mathbf{E}^n}\left\{ \frac{1}{n}(\log |Y_{\delta,\delta_1,n,+}(\mathcal{S})| + \log \dim \mathcal{H}_{\mathbf{k}}^{\delta,n}) \geqslant R \right\} \leqslant \sum_{\mathbf{k} \in Y_{\delta,\delta_1,n,+}(\mathcal{S}) : \max_{\mathbf{n} \in Y_n \cap U_{\mathbf{k},n\delta}} H(\mathbf{n}/n) \geqslant R - (4d/n)\log(n+d)} \mathrm{Tr} M_{\mathbf{k}}^{\delta,n} \overline{\rho_p}^{\otimes n}$$

$$\leqslant |Y_{\delta,n}| (n+d)^{4d} \max_{\substack{\mathbf{n} \in Y_n : H(\mathbf{n}/n) \geqslant R - (4d/n)\log(n+d) \\ \exists \rho \in \mathcal{S}, \|(\mathbf{n}/n) - \mathbf{p}(\rho)\| \leqslant \delta_1}} \left\{ \max_{\mathbf{n}' \in Y_n : \|\mathbf{n}-\mathbf{n}'\| \leqslant 2\delta n} \exp\left[ -nD\left(\frac{\mathbf{n}'}{n} \middle\| \mathbf{p}\right)\right]\right\}$$

$$\leqslant (n+d)^{5d} \max_{\substack{\mathbf{q} \in \mathbb{R}_+^{d,1} : H(\mathbf{q}) \geqslant R - (4d/n)\log(n+d) \\ \exists \rho \in \mathcal{S}, \|\mathbf{q} - \mathbf{p}(\rho)\| \leqslant \delta_1}} \left( \max_{\mathbf{q}' \in \mathbb{R}_+^{d,1} : \|\mathbf{q}-\mathbf{q}'\| \leqslant 2\delta} \exp[-nD(\mathbf{q}'\|\mathbf{p})]\right).$$

Then, we obtain (18).

## APPENDIX C: PROOF OF (13), (16), AND (17)

We can evaluate the average error as

$$\epsilon_{n,p}(\mathbf{E}^{\delta,n},\mathbf{D}^{\delta,n}) = \sum_{\vec{\rho}_n \in \mathcal{S}(\mathcal{H}^{\otimes n})} p^n(\vec{\rho}_n) \sum_{\mathbf{k}\in Y_{\delta,n}} \mathrm{Tr} M_{\mathbf{k}}^{\delta,n} \vec{\rho}_n \left( 1 - \mathrm{Tr} \left| \sqrt{\vec{\rho}_n} \sqrt{\frac{\sqrt{M_{\mathbf{k}}^{\delta,n}} \vec{\rho}_n \sqrt{M_{\mathbf{k}}^{\delta,n}}}{\mathrm{Tr} M_{\mathbf{k}}^{\delta,n} \vec{\rho}_n}} \right| \right)$$

$$= 1 - \sum_{\vec{\rho}_n \in \mathcal{S}(\mathcal{H}^{\otimes n})} p^n(\vec{\rho}_n) \sum_{\mathbf{k}\in Y_{\delta,n}} \sqrt{\mathrm{Tr} M_{\mathbf{k}}^{\delta,n} \vec{\rho}_n} \mathrm{Tr} \sqrt{\sqrt{\vec{\rho}_n} \sqrt{M_{\mathbf{k}}^{\delta,n}} \vec{\rho}_n \sqrt{M_{\mathbf{k}}^{\delta,n}} \sqrt{\vec{\rho}_n}}$$

$$= 1 - \sum_{\vec{\rho}_n \in \mathcal{S}(\mathcal{H}^{\otimes n})} p^n(\vec{\rho}_n) \sum_{\mathbf{k}\in Y_{\delta,n}} \sqrt{\mathrm{Tr} M_{\mathbf{k}}^{\delta,n} \vec{\rho}_n} \mathrm{Tr} \sqrt{\vec{\rho}_n} \sqrt{M_{\mathbf{k}}^{\delta,n}} \sqrt{\vec{\rho}_n}$$

$$= 1 - \sum_{\mathbf{k}\in Y_{\delta,n}} \frac{1}{C_{1,d}(n\delta)} \sum_{\vec{\rho}_n \in \mathcal{S}(\mathcal{H}^{\otimes n})} p^n(\vec{\rho}_n) (\mathrm{Tr} P_{\mathbf{k}}^{\delta,n} \vec{\rho}_n)^{3/2}$$

$$\leqslant 1 - \sum_{\mathbf{k}\in Y_{\delta,n}} \frac{1}{C_{1,d}(n\delta)} \left( \sum_{\vec{\rho}_n \in \mathcal{S}(\mathcal{H}^{\otimes n})} p^n(\vec{\rho}_n) \mathrm{Tr} P_{\mathbf{k}}^{\delta,n} \vec{\rho}_n \right)^{3/2}, \tag{C1}$$

$$= 1 - \sum_{\mathbf{k}\in Y_{\delta,n}} \frac{1}{C_{1,d}(n\delta)} (\mathrm{Tr} \overline{\rho_p^{\otimes n}} P_{\mathbf{k}}^{\delta,n})^{3/2}, \tag{C2}$$

where inequality (C1) follows from Jensen's inequality concerning the convex function $x \mapsto x^{3/2}$.

The relations

$$C_{2,d}(n\delta_1) \leqslant \mathcal{N}(Y_{\delta,n} \cap U_{n\mathbf{p},n\delta_1}), \qquad 0 < \delta_1 < \delta, \tag{C3}$$

$$P_{\mathbf{k}}^{\delta,n} \geqslant \sum_{\mathbf{n}\in Y_n \cap U_{n\mathbf{p},n(\delta-\delta_1)}} P_{\mathbf{n}}, \qquad \forall \mathbf{k} \in Y_{\delta,n} \cap U_{n\mathbf{p},n\delta_1} \tag{C4}$$

hold. Using Lemma 5 and Eqs. (C4) and (15), we have

$$\mathrm{Tr} P_{\mathbf{k}}^{\delta,n} \overline{\rho_p^{\otimes n}} \geqslant 1 - (n+d)^{4d} \exp(-n \min_{q \notin U_{\mathbf{p},\delta-\delta_1}} D(\mathbf{q}\|\mathbf{p})) \geqslant 1 - (n+d)^{4d} \exp[-nC_{3,d}(\delta-\delta_1)^2]. \tag{C5}$$

It follows from (C3) and (C5) that

$$\sum_{\mathbf{k}\in Y_{\delta,n}} \frac{1}{C_{1,d}(n\delta)} (\mathrm{Tr} \overline{\rho_p^{\otimes n}} P_{\mathbf{k}}^{\delta,n})^{3/2} \geqslant \frac{1}{C_{1,d}(n\delta)} \sum_{\mathbf{k}\in Y_{\delta,n}\cap U_{n\mathbf{p},n\delta_1}} (\mathrm{Tr} \overline{\rho_p^{\otimes n}} P_{\mathbf{k}}^{\delta,n})^{3/2}$$

$$\geqslant \frac{C_{2,d}(n\delta_1)}{C_{1,d}(n\delta)} \{1 - (n+d)^{4d} \exp[-nC_{3,d}(\delta-\delta_1)^2]\}^{3/2}. \tag{C6}$$

Inequality (13) follows from (C2) and (C6).

Similarly to (C2), we can prove

$$\epsilon''_{n,p}(\mathbf{E}^{\delta,n},\mathbf{D}^{\delta,n}) \leqslant 1 - \sum_{\mathbf{k}\in Y_{\delta,n}} \frac{1}{C_{1,d}(n\delta)} (\mathrm{Tr} \overline{\rho_p^{\otimes n}} P_{\mathbf{k}}^{\delta,n})^2,$$

which implies (16).

In the general case, similarly to (C2), we can prove that

$$\epsilon_{n,p}(\mathbf{E}^{\delta,n},\mathbf{D}^{\delta,n}) \leqslant 1 - \sum_{\mathbf{k}\in Y_{\delta,\delta_1,n}(\mathcal{S})} \frac{1}{C_{1,d}(n\delta)} (\mathrm{Tr} \overline{\rho_p^{\otimes n}} P_{\mathbf{k}}^{\delta,n})^{3/2}. \tag{C7}$$

Since $Y_{\delta,\delta_1,n,+}(\mathcal{S}) \cap U_{n\mathbf{p},n\delta_1} = Y_{\delta,n} \cap U_{n\mathbf{p},n\delta_1}$, we can prove that

$$\sum_{\mathbf{k}\in Y_{\delta,\delta_1,n,+}(\mathcal{S})} \frac{1}{C_{1,d}(n\delta)} (\mathrm{Tr}\overline{\rho_p}^{\otimes n} P_{\mathbf{k}}^{\delta,n})^{3/2} \geqslant \frac{1}{C_{1,d}(n\delta)} \sum_{\mathbf{k}\in Y_{\delta,n}\cap U_{n\mathbf{p},n\delta_1}} (\mathrm{Tr}\overline{\rho_p}^{\otimes n} P_{\mathbf{k}}^{\delta,n})^{3/2}$$

$$\geqslant \frac{C_{2,d}(n\delta_1)}{C_{1,d}(n\delta)} \{1-(n+d)^{4d}\exp[-nC_{3,d}(\delta-\delta_1)^2]\}^{3/2}. \tag{C8}$$

Inequality (17) follows from (C7) and (C8).

## APPENDIX D: PROOF OF LEMMA 2

In this case, the average error is calculated as

$$\epsilon_{n,p}(\mathbf{E}^{0,n},\mathbf{D}^{0,n}) = 1-\sum_{\mathbf{n}\in Y_n}\sum_{\vec{e}_n} p(\vec{e}_n)(\langle\vec{e}_n|P_{\mathbf{n}}|\vec{e}_n\rangle)^{3/2} = 1-\sum_{\vec{e}_n} p(\vec{e}_n)\sum_{\mathbf{n}\in Y_n}(\langle\vec{e}_n|P_{\mathbf{n}}|\vec{e}_n\rangle)^{3/2},$$

where $\vec{e}_n := e_{i_1}\otimes e_{i_2}\otimes\cdots\otimes e_{i_n}\in\mathcal{H}^{\otimes n}$. We define $\mathbf{n}(\vec{e}_n) := (n_1(\vec{e}_n), n_2(\vec{e}_n))$ by

$$n_i(\vec{e}_n) := \mathcal{N}\{j\in[1,n]|e_{i_j}=e_i\}. \tag{D1}$$

Now, we focus a typical element $\vec{e}_n$, i.e., $[n_i(\vec{e}_n)/n]\cong p_i$. The number satisfying Eq. (D1) is $\binom{n_2(\vec{e}_n)}{n(\vec{e}_n)}$, and $\dim\mathcal{V}_{\mathbf{n}'} = [\binom{n_2(\vec{e}_n)}{n} - \binom{n_2(\vec{e}_n)-1}{n}]$, where $\mathbf{n}(\vec{e}_n) = (n_1(\vec{e}_n), n_2(\vec{e}_n)) \in Y_n$. Then,

$$\langle\vec{e}_n|P_{\mathbf{n}(\vec{e}_n)}|\vec{e}_n\rangle = \binom{n}{n_2(\vec{e}_n)}^{-1}\left[\binom{n}{n_2(\vec{e}_n)} - \binom{n}{n_2(\vec{e}_n)-1}\right] = 1-\frac{n_2(\vec{e}_n)}{n_1(\vec{e}_n)+1} = \frac{\frac{n_1(\vec{e}_n)}{n}+\frac{1}{n}-\frac{n_2(\vec{e}_n)}{n}}{\frac{n_1(\vec{e}_n)}{n}+\frac{1}{n}} \cong \frac{p_1-p_2}{p_1}.$$

Since $x^{3/2}+y^{3/2}\leqslant(x+y)^{3/2}$ for $0<x,y<1$, we can evaluate

$$\sum_{\mathbf{n}\in Y_n}(\langle\vec{e}_n|P_{\mathbf{n}}|\vec{e}_n\rangle)^{3/2} \leqslant \left(\sum_{\mathbf{n}\in Y_n\setminus\{\mathbf{n}'\}}\langle\vec{e}_n|P_{\mathbf{n}}|\vec{e}_n\rangle\right)^{3/2} + (\langle\vec{e}_n|P'_{\mathbf{n}}|\vec{e}_n\rangle)^{3/2} \cong \left(1-\frac{p_1-p_2}{p_1}\right)^{3/2} + \left(\frac{p_1-p_2}{p_1}\right)^{3/2}$$

$$= \left(\frac{p_2}{p_1}\right)^{3/2} + \left(\frac{p_1-p_2}{p_1}\right)^{3/2} < 1.$$

Therefore,

$$\lim\epsilon_{n,p}(\mathbf{E}^{0,n},\mathbf{D}^{0,n}) \geqslant 1-\left[\left(\frac{p_2}{p_1}\right)^{3/2} + \left(\frac{p_1-p_2}{p_1}\right)^{3/2}\right] > 0.$$

## APPENDIX E: PROOF OF LEMMA 3

For any $\mathbf{n}\in Y_n, \delta_n>0$, we denote $([n_1-(1/\sqrt{2})\delta_n], n-[n_1-(1/\sqrt{2})\delta_n]) \in Y_{\delta,n}$ by $\mathbf{k}(\mathbf{n},\delta_n)$, where $[x]$ is defined as the maximum integer $n$ satisfying $n\leqslant x$. The element $\mathbf{k}(\mathbf{n},\delta_n)$ satisfies

$$\mathbf{n}=(n_1,n_2)\in U_{\mathbf{k}(\mathbf{n},\delta_n),\delta_n}, \quad (n_1+1,n_2-1)\notin U_{\mathbf{k}(\mathbf{n},\delta_n),\delta_n}.$$

For any $\delta>0$, we have

$$\epsilon_{n,p}(\mathbf{E}^{\delta,n},\mathbf{D}^{\delta,n}) = \sum_{\vec{e}_n} p^n(\vec{e}_n)\left(1 - \sum_{\mathbf{k}\in Y_{\delta,n}} \frac{1}{C_{1,d}(n\delta)}(\mathrm{Tr} P_{\mathbf{k}}^{\delta,n}\vec{\rho}_n)^{3/2}\right)$$

$$\geqslant \sum_{\vec{e}_n} p^n(\vec{e}_n)\left(1 - \sum_{\mathbf{k}\neq \mathbf{k}(\mathbf{n},\delta_n)\in Y_{\delta,n}} \frac{1}{C_{1,d}(n\delta)}(\mathrm{Tr}\, P_{\mathbf{k}}^{\delta,n}\vec{\rho}_n) - \frac{1}{C_{1,d}(n\delta)}(\mathrm{Tr}\, P_{\mathbf{k}(\mathbf{n},\delta_n)}^{\delta,n}\vec{\rho}_n)^{3/2}\right)$$

$$= \sum_{\vec{e}_n} p^n(\vec{e}_n)\left(\frac{1}{C_{1,d}(n\delta)}(\mathrm{Tr}\, P_{\mathbf{k}(\mathbf{n},\delta_n)}^{\delta,n}\vec{\rho}_n) - \frac{1}{C_{1,d}(n\delta)}(\mathrm{Tr}\, P_{\mathbf{k}(\mathbf{n},\delta_n)}^{\delta,n}\vec{\rho}_n)^{3/2}\right)$$

$$= \sum_{\vec{e}_n} p^n(\vec{e}_n)\frac{1}{C_{1,d}(n\delta)}[\mathrm{Tr}\, P_{\mathbf{k}(\mathbf{n},\delta_n)}^{\delta,n}\vec{\rho}_n - (\mathrm{Tr}\, P_{\mathbf{k}(\mathbf{n},\delta_n)}^{\delta,n}\vec{\rho}_n)^{3/2}]$$

$$\geqslant \sum_{\vec{e}_n : n_i(\vec{e}_n)/n \cong p_i} p^n(\vec{e}_n)\frac{1}{C_{1,d}(n\delta)}[\mathrm{Tr}\, P_{\mathbf{k}(\mathbf{n},\delta_n)}^{\delta,n}\vec{\rho}_n - (\mathrm{Tr}\, P_{\mathbf{k}(\mathbf{n},\delta_n)}^{\delta,n}\vec{\rho}_n)^{3/2}]$$

$$\cong \sum_{\vec{e}_n : \frac{n_i(\vec{e}_n)}{n} \cong p_i} p^n(\vec{e}_n)\frac{1}{C_{1,d}(n\delta)}\left[\frac{p_1-p_2}{p_1} - \left(\frac{p_1-p_2}{p_1}\right)^{3/2}\right]$$

$$\cong \frac{1}{C_{1,d}(n\delta)}\left[\frac{p_1-p_2}{p_1} - \left(\frac{p_1-p_2}{p_1}\right)^{3/2}\right] \geqslant \frac{1}{2|Y_n|}\left[\frac{p_1-p_2}{p_1} - \left(\frac{p_1-p_2}{p_1}\right)^{3/2}\right].$$

Note that the RHS is independent of $\delta > 0$. Thus,

$$\frac{-1}{n}\log \epsilon_{n,p}(\mathbf{E}^{\delta,n},\mathbf{D}^{\delta,n}) \leqslant \frac{-1}{n}\left\{\log\frac{1}{2|Y_n|} + \log\left[\frac{p_1-p_2}{p_1} - \left(\frac{p_1-p_2}{p_1}\right)^{3/2}\right]\right\} \leqslant \frac{1}{n}\left\{\log 2(n+1)^2 - \log\left[\frac{p_1-p_2}{p_1} - \left(\frac{p_1-p_2}{p_1}\right)^{3/2}\right]\right\} \to 0.$$

Therefore, we obtain Eq. (25).


[1] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).

[2] R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).

[3] M. Koashi and N. Imoto, Phys. Rev. Lett. **87**, 017902 (2001); LANL e-print quant-ph/0103128.

[4] R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **81**, 1714 (1998); LANL e-print quant-ph/9805017.

[5] T.J. Lynch, Proc. IEEE **54**, 1490 (1966).

[6] L.D. Davisson, Proc. IEEE **54**, 2010 (1966).

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Academic Press, New York, 1981).

[8] M. Keyl and R.F. Werner, Phys. Rev. A **64**, 052311 (2001); LANL e-print quant-ph/0102027.

[9] C. Krattenthanler and P.B. Slater, IEEE Trans. Inf. Theory **IT-46**, 801 (2000).

[10] B. Schumacher and M.D. Westmoreland, Phys. Rev. A **64**, 042304 (2001); LANL e-print quant-ph/0011014.

[11] M. Ozawa, J. Math. Phys. **25**, 79 (1984).

[12] M. Ozawa, *Mathematical Characterizations of Measurement Statistics, in Quantum Communication and Measurement*, edited by V. P. Belavkin *et al.* (Plenum Press, New York, 1995).

[13] M. Ozawa, Phys. Rev. A **62**, 062101 (2000).

[14] H. Barnum, C.A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. A **54**, 4707 (1996).

[15] A. Winter, Ph.D. dissertation, Universität Bielefeld (2000); LANL e-print quant-ph/9907077.

[16] M. Hayashi, Phys. Rev. A (to be published); LANL e-print quant-ph/0202002.

[17] M. Hayashi, J. Phys. A (to be published); LANL e-print quant-ph/0202003.

[18] H. Weyl, *The Classical Groups, Their Invariants and Representations* (Princeton University Press, Princeton, NJ, 1939).

[19] R. Goodman and N. Wallach, *Representations and Invariants of the Classical Groups* (Cambridge University Press, Cambridge, 1998).

[20] N. Iwahori, *Taishougun to Ippansenkeigun no Hyougenron* (Iwanami, Tokyo, 1978) (in Japanese).

[21] M. Hayashi, J. Phys. A **34**, 3413 (2001); LANL e-print quant-ph/9704040.

[22] M. Hayashi, LANL e-print quant-ph/0107004.

[23] T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory IT-**46**, 2428 (2000); LANL e-print quant-ph/9906090.