Quantum channels showing superadditivity in classical capacity

Masahide Sasaki,[1] Kentaro Kato,[2] Masayuki Izutsu,[1] and Osamu Hirota[2]
[1]*Communications Research Laboratory, Ministry of Posts and Telecommunications, Koganei, Tokyo 184, Japan*
[2]*Research Center for Quantum Communications, Tamagawa University, Tamagawa-gakuen, Machida, Tokyo 194, Japan*
(Received 14 October 1997)

We consider a channel coding for sending classical information through a quantum channel with a given ensemble of quantum states (letter states). As is well known, it is generically possible in a quantum channel that the transmittable information in block coding of length n can exceed n times the maximum amount that can be sent without any coding scheme. This so-called superadditivity in classical capacity of a quantum channel is a distinct feature that cannot be found in a classical memoryless channel. In this paper, a practical model of channel coding that shows this property is presented. It consists of a simple code-word selection and the optimum decoding of the code words minimizing the average error probability. At first, optimization of decoding strategy is discussed. Then the channel coding that shows the superadditivity in classical capacity is demonstrated. [S1050-2947(98)05107-5]

PACS number(s): 03.67.Hk, 89.70.+c, 42.79.Sz, 89.80.+h

I. INTRODUCTION

Theory of quantum communication was initiated more than 30 years ago, in order to consider the quantum nature of the signal carrier in the optical frequency domain. In this region, one faces quite different features from rf band communication, due to quantum noise of the signal carrier itself. This theory was then developed in the 1970s, revealing new aspects of information transmission and signal detection. It now attracts much attention since new fields such as quantum computation and cryptography emerged. Being assisted by ideas and methods in these fields, significant progress was made in a basic and old issue on channel capacity. In particular, the theorem was established that the attainable maximum rate of asymptotically error free transmission for sending classical information by using a given source of quantum states (letter states) is precisely the Holevo bound [1,2] [let us call it the quantum channel coding (QCC) theorem] [3–6]. This rate is the asymptotic rate at infinite block length, $n \rightarrow \infty$, and is especially called the *classical capacity* of the quantum channel. The term *capacity* of the quantum channel is now used in various contexts of quantum information theory, including not only transmission of fixed classical alphabet but also sending intact quantum states. In this paper we confine ourselves to transmission of classical information by use of a given letter-state ensemble, and hereafter the term *capacity* is understood as *classical capacity* for this case.

The QCC theorem guarantees existence of codes that have the above asymptotic property, but does not tell directly how to construct such codes from given letter states. For practical applications, simple and systematic coding-decoding methods at finite block length are required. Such methods will immediately be applied to, for example, advanced schemes of satellite communication and ultrafast optical fiber communication. In these cases, signal power at the receiving end might be very weak due to long distance transmission or limited power supply so that a main source to cause error will be nonorthogonality among letter states, which is just the situation covered by the above theorem.

The purpose of this paper is to give some insights into practical aspects of quantum channel coding. In quantum channel coding, block sequences are made as direct product states of the letter states, some of them are selected and transmitted as *code-word states*, and they are then detected quantum mechanically. A quantum channel made of the code-word states of length n is called an n-product channel. There are two essential ingredients in using this n-product channel; one is the suitable selection of code-word states from all the possible sequences made of the letter states, and the other is a *collective decoding* that detects each code-word state as a single state vector rather than decoding the individual letter states separately. Especially, the latter fully utilizes superposition states of the code-word states, and brings an inseparable structure among the letter states, which is often called *entanglement*. This remarkable feature cannot be found in classical channel coding. As a consequence, the n-product channel can have a *memory effect* in the sense that the channel matrix cannot be factorized into the channel matrices corresponding to each letter, i.e., $P(y_1 y_2 \cdots y_n | x_1 x_2 \cdots x_n) \neq \Pi_{i=1}^{n} P(y_i | x_i)$. This is even so if neither a source system emitting letter states nor a physical process of transmission has memory effect. This effect can be used to increase the reliability of information transmission. In fact, if the code-word states are selected suitably, this effect makes it possible that more classical information can be sent through the n-product channel than n-times the amount that can be sent through a single use of the initial channel made merely of the letter states without any coding scheme. This so-called superadditivity in capacity is a generic nature of a quantum channel, and is indeed information theoretic quantum gain [1,2]. So the first step toward finding the ultimate channel coding for the Holevo bound might be to construct codes that attain this quantum gain.

In this paper quantum channels that show the superadditivity in capacity are described. We first consider optimization of decoding. The collective decoding used in the proof of the QCC theorem was the so-called *square-root measurement* [3]. This allows one to derive an explicit decoding observable systematically from given code-word states. In ad-

dition, this has been known to be almost optimum when the quantum states to be distinguished are equally likely and almost orthogonal [7–9], which is the case for the typical sequences obtained at very long block length. Therefore it played a sufficient role in evaluating the upper bound of the decoding error. But this measurement is actually more than that. In Sec. II it will be pointed out that the square-root measurement becomes precisely optimum in terms of the average error probability in certain cases of pure and linearly independent quantum states that are even neither equiprobable nor almost orthogonal. The optimality of decoding strategy should be pursued in order to achieve performance as high as possible, especially in a practical channel coding of finite block length. When the square-root measurement is not optimum, there is a method to construct the optimum one by modifying it. For a practical purpose, we present a basic scheme of the optimum collective decoding of code-word states in the case of the pure-state channel in Sec. III. As for physical realizations of this scheme, the readers are referred to the subsequent paper.

In order to quantify the superadditivity in capacity, the attainable maximum mutual information without any coding must be known. This quantity is usually denoted as C_1. The optimum solutions of the prior probabilities and the decoding observable that maximize the mutual information have been known only in a few cases [10–12]. In this paper the most basic case of binary and pure letter states is considered. In Sec. IV we define a *threshold point* where all the sequences are used as the code-word states and the accessible information (the maximum mutual information attained by optimizing the decoding observable with prior probabilities fixed) at block length n is exactly nC_1. At this point, the optimum decoding is not like a collective fashion but rather reduces to the separate measurement which detects each letter state individually, that is, there is no room for generating entanglement correlation among letter states. This threshold point will be a useful guide for quantitative discussions. Section V is devoted to concluding remarks.

II. DISTINGUISHING LINEARLY-INDEPENDENT QUANTUM STATES

To begin with, we shall describe the conditions for optimality in a general decision problem of M-ary quantum states. An ensemble of quantum states $\{\hat{\rho}_i\}$ is given with respective prior probabilities $\{\xi_i\}$. Decision process of these signal states $\{\hat{\rho}_i\}$ can be described by a probability operator measure (POM) $\{\hat{\Pi}_i\}$ satisfying the resolution of the identity $\Sigma_i \hat{\Pi}_i = \hat{I}$. The POM effecting the decision needs only M components $\hat{\Pi}_1, \hat{\Pi}_2, \ldots, \hat{\Pi}_M$ which are usually called *detection operators*. What we are seeking here are the optimum detection operators minimizing the average error probability. Defining the risk operators $\hat{W}_i \equiv \xi_i \hat{\rho}_i$ and the Lagrange operator $\hat{\Upsilon} \equiv \Sigma_i \hat{W}_i \hat{\Pi}_i$, the optimum conditions are written as [7,13]

$$\hat{\Pi}_i(\hat{W}_i - \hat{W}_j)\hat{\Pi}_j = 0 \quad \forall \ (i,j) \quad \text{(i)},$$

$$\hat{\Upsilon} - \hat{W}_i \geqslant 0 \quad \forall \ i \quad \text{(ii)}.$$

The minimum average error probability is given by

$$P_e(\text{opt}) = 1 - \text{Tr}\hat{\Upsilon}. \tag{1}$$

When the signal states are pure ($\hat{\rho}_i = |\rho_i\rangle\langle\rho_i|$) and linearly independent, the optimum detection operators can be given as the projection-valued measure (PVM) with rank 1 as $\hat{\Pi}_i = |\omega_i\rangle\langle\omega_i|$. The set $\{|\omega_i\rangle\}$ forms a complete orthonormal set in the Hilbert space $\mathcal{H}_s$ spanned by the signal states $\{|\rho_i\rangle\}$ and each of them is called a *measurement state*. Introducing a matrix $\mathbf{X} = (X_{ij}) \equiv (\langle\omega_i|\rho_j\rangle)$, the above conditions are rewritten as

$$\xi_i X_{ii} X_{ji}^* = \xi_j X_{ij} X_{jj}^* \quad \forall \ (i,j) \quad \text{(i}'\text{)},$$

$$\mathbf{T}^{(m)} \equiv (\xi_i X_{ii} X_{ji}^* - \xi_m X_{im} X_{jm}^*) \geqslant 0 \quad \text{(ii}'\text{)},$$

$$m = 1, \ldots, M.$$

In general, it is a complicated job to derive explicit expressions for the optimum measurement states satisfying the above conditions. Only in certain cases have they been known [14]. Otherwise, one has to rely on numerical simulations like the Bayes-cost-reduction algorithm [15]. The most tedious part in such a method is to check the second condition (ii′). But when the signal states are linearly independent, this is ensured more simply, if

$$\mathbf{\Upsilon}' \equiv (\xi_i X_{ii} X_{ji}^*) > 0 \quad \text{(ii}''\text{)}$$

is satisfied. Its proof was given in the Appendix of Ref. [15].

Now let us consider when the square-root measurement becomes optimum. The square-root measurement is defined as follows:

$$|\mu_i\rangle \equiv \hat{\rho}^{-1/2} |\tilde{\rho}_i\rangle, \tag{2a}$$

$$\hat{\rho} \equiv \sum_{i=1}^{M} |\tilde{\rho}_i\rangle\langle\tilde{\rho}_i|, \tag{2b}$$

$$|\tilde{\rho}_i\rangle \equiv \sqrt{\xi_i}\,|\rho_i\rangle. \tag{2c}$$

As is well known, the conditional probability based on this measurement $P(j|i) = |\langle\mu_j|\rho_i\rangle|^2$ can be calculated in the following way. First make the Gram matrix $\mathbf{\Gamma} \equiv (\langle\tilde{\rho}_i|\tilde{\rho}_j\rangle)$. Second diagonalize it as

$$\mathbf{\Gamma} = \mathbf{Q} \begin{pmatrix} g_1 & & \\ & \ddots & \\ & & g_M \end{pmatrix} \mathbf{Q}^\dagger, \tag{3}$$

where $\mathbf{Q}$ is a unitary matrix. Third and finally, calculate

$$\sqrt{\mathbf{\Gamma}} = \mathbf{Q} \begin{pmatrix} \sqrt{g_1} & & \\ & \ddots & \\ & & \sqrt{g_M} \end{pmatrix} \mathbf{Q}^\dagger. \tag{4}$$

Then the (i,j) components of $\sqrt{\mathbf{\Gamma}}$ are just $\langle\mu_i|\tilde{\rho}_j\rangle = \langle\mu_i|\rho_j\rangle\sqrt{\xi_j}$. Here we give a useful theorem in considering the optimum collective decoding.

Theorem 1

If $\{|\rho_i\rangle\}$ are linearly independent, the measurement by $\{|\mu_i\rangle\}$ becomes optimum when all of the diagonal components of $\sqrt{\mathbf{\Gamma}}$ are equal, that is, when probability of correctly identifying the letter is independent of the letter sent.

Proof

Define $\mathbf{Y}=(Y_{ij})\equiv(\langle\mu_i|\rho_j\rangle)$. The measurement by $\{|\mu_i\rangle\}$ is optimum if

$$\xi_i Y_{ii} Y_{ji}^* = \xi_j Y_{ij} Y_{jj}^* \quad \forall\ (i,j) \quad (\mathrm{i}'),$$

$$\mathbf{Y}' \equiv (\xi_i Y_{ii} Y_{ji}^*) > 0 \quad (\mathrm{ii}'')$$

are satisfied. Denoting $\sqrt{\mathbf{\Gamma}}=(\tilde{Y}_{ij})$, they can be rewritten as

$$\tilde{Y}_{ii} \tilde{Y}_{ji}^* = \tilde{Y}_{ij} \tilde{Y}_{jj}^* \quad \forall\ (i,j) \quad (\mathrm{i}'),$$

$$\mathbf{Y}' \equiv (\tilde{Y}_{ii} \tilde{Y}_{ji}^*) > 0 \quad (\mathrm{ii}'').$$

Since $\mathbf{\Gamma}$ is nonnegative and Hermitian, so is $\sqrt{\mathbf{\Gamma}}$. Therefore the above conditions reduce to

$$\tilde{Y}_{ii} = \tilde{Y}_{jj} \quad \forall\ (i,j) \quad (\mathrm{i}'),$$

$$\mathbf{Y}' \equiv (\tilde{Y}_{ii} \tilde{Y}_{ij}) > 0 \quad (\mathrm{ii}'').$$

Under the first condition, the second one further reduces to $\tilde{Y}_{11}\tilde{\mathbf{Y}}>0$. This is automatically satisfied since $\sqrt{\mathbf{\Gamma}}=\tilde{\mathbf{Y}}>0$. ($\tilde{\mathbf{Y}}>0$ implies $\tilde{Y}_{ii}>0\ \forall\ i$.) Thus for the square-root measurement, the first condition just above is enough for the optimum condition and this means the theorem. ■

Thus the square-root measurement plays a practical role not only in the case of equally probable and almost orthogonal states but also the case that the signal states satisfy the above condition. The related discussion was given by Ban *et al.* in the case of equally probable and symmetric states [16]. Even if the signal states do not satisfy the above condition, $\{|\mu_i\rangle\}$ can be good initial states in searching the optimum measurement states. At first, note the following remark.

Remark

If $\{|\rho_i\rangle\}$ are linearly independent, $\{|\mu_i\rangle\}$ are orthonormal.

Proof

The optimum measurement state $\{|\omega_i\rangle\}$ is a complete orthonormal set in $\mathcal{H}_s$. Define $\hat{X}\equiv\Sigma_{i,j}X_{ij}|\omega_i\rangle\langle\omega_j|$ so that $|\rho_i\rangle=\hat{X}|\omega_i\rangle$. Because of the linear independence of $\{|\rho_i\rangle\}$, $\hat{X}$ is nonsingular and $\hat{X}^{-1}$ exists. Then

$$\begin{aligned}\langle\mu_i|\mu_j\rangle &= \langle\tilde{\rho}_i|\hat{\rho}^{-1}|\tilde{\rho}_j\rangle = \sqrt{\xi_i}\langle\omega_i|\hat{X}^\dagger\hat{\rho}^{-1}\hat{X}|\omega_j\rangle\sqrt{\xi_j} \\ &= \sqrt{\xi_i}\langle\omega_i|\left(\sum_k \xi_k \hat{X}^{-1}|\rho_k\rangle\langle\rho_k|\hat{X}^{\dagger-1}\right)^{-1}|\omega_j\rangle\sqrt{\xi_j} \\ &= \sqrt{\xi_i}\langle\omega_i|\left(\sum_k \xi_k|\omega_k\rangle\langle\omega_k|\right)^{-1}|\omega_j\rangle\sqrt{\xi_j} = \delta_{ij}. \quad ■\end{aligned}$$

Thus, for linearly independent states, the set $\{|\mu_i\rangle\}$ is always a complete orthogonal set in $\mathcal{H}_s$. So it can be connected via a unitary operator $\hat{V}$ in $\mathcal{H}_s$ with the optimum measurement states $\{|\omega_i\rangle\}$ as $|\omega_i\rangle=\hat{V}|\mu_i\rangle$. Such an operator can be constructed, for example, as a series of two-dimensional rotations by applying the Bayes-cost-reduction algorithm [15]. This algorithm consists of steps of solving a binary decision problem of a chosen pair of signal states $\{|\rho_i\rangle,|\rho_j\rangle\}$ on the plane spanned by the corresponding pair of basis vectors $\{|\mu_i\rangle,|\mu_j\rangle\}$. At every step, two basis vectors are revised, and the average error would decrease or, at worst, remain the same. These two-dimensional rotations are continued till reaching the optimum point where the previous conditions (i') and (ii') are satisfied. The resulting series of their products is just the required unitary operator $\hat{V}$.

III. OPTIMUM COLLECTIVE DECODING OF CODE WORDS

Deriving the analytic expression of the optimum measurement basis vectors $\{|\omega_i\rangle\}$ is a difficult job, but these basis vectors can be constructed somehow as explained in the preceding section, and at the same time the channel matrix can be obtained. Thus only for evaluating performance, is it sufficient to derive these $\{|\omega_i\rangle\}$. From a practical point of view, however, the basis vectors $\{|\omega_i\rangle\}$ hardly imply a corresponding physical process. Although the set $\{|\omega_i\rangle\}$ forms a standard von Neumann measurement, its physical implementation usually remains a nontrivial problem. In this section we present a useful scheme for realizing the optimum collective decoding. In the case where the letter states are binary, this scheme naturally leads to an implementation based on a quantum circuit and a well-defined physical measurement.

Let binary letter states be $\{|+\rangle,|-\rangle\}$ whose state overlap $\kappa=\langle+|-\rangle$ is assumed to be real and to lie in $0\leqslant\kappa<1$. They span the two-dimensional Hilbert space $\mathcal{H}_l$. By nth extension, we pick up M-ary code-word states $\{|S_1\rangle,\ldots,|S_M\rangle\}$ $(M\leqslant 2^n)$ from the 2^n possible sequences of length n, in the nth extended Hilbert space $\mathcal{H}_l^{\otimes n}$ and use them with respective input probabilities $\{\zeta_1,\ldots,\zeta_M\}$. The rest of the sequences are denoted as $\{|S_{M+1}\rangle,\ldots,|S_{2^n}\rangle\}$. Since the code-word states are linearly independent, they span the M-dimensional Hilbert space $\mathcal{H}_s\subset\mathcal{H}_l^{\otimes n}$. The optimum collective decoding is described by the orthonormal basis vectors $\{|\omega_1\rangle,\ldots,|\omega_M\rangle\}$ in $\mathcal{H}_s$ derived in such a way as mentioned in the preceding section. An orthonormal set $\{|\omega_1\rangle,\ldots,|\omega_{2^n}\rangle\}$ in the extended space $\mathcal{H}_l^{\otimes n}$ can be made by adding the other basis vectors obtained by using the Schmidt orthogonalization,

$$|\omega_i\rangle = \frac{|S_i\rangle - \sum_{k=1}^{i-1}|\omega_k\rangle\langle\omega_k|S_i\rangle}{\sqrt{1-\sum_{k=1}^{i-1}|\langle\omega_k|S_i\rangle|^2}} \quad (i=M+1,\ldots,2^n). \quad (5)$$

We denote the expansion of all the sequences by the above basis vectors as

$$\begin{pmatrix} |S_1\rangle \\ \vdots \\ |S_{2^n}\rangle \end{pmatrix} = \mathbf{B} \begin{pmatrix} |\omega_1\rangle \\ \vdots \\ |\omega_{2^n}\rangle \end{pmatrix}, \tag{6a}$$

$$\mathbf{B} = (B_{ij}) = (\langle \omega_j | S_i \rangle). \tag{6b}$$

Making two orthonormal basis vectors $\{|a\rangle, |b\rangle\}$ from $\{|+\rangle, |-\rangle\}$, we introduce the 2^n product basis vectors,

$$\begin{aligned} |A_1\rangle &\equiv |a\rangle \otimes |a\rangle \otimes \cdots \otimes |a\rangle \otimes |a\rangle, \\ |A_2\rangle &\equiv |a\rangle \otimes |a\rangle \otimes \cdots \otimes |a\rangle \otimes |b\rangle, \\ &\vdots \\ |A_{2^n-1}\rangle &\equiv |b\rangle \otimes |b\rangle \otimes \cdots \otimes |b\rangle \otimes |a\rangle, \\ |A_{2^n}\rangle &\equiv |b\rangle \otimes |b\rangle \otimes \cdots \otimes |b\rangle \otimes |b\rangle, \end{aligned} \tag{7}$$

and denote another expansion by them as

$$\begin{pmatrix} |S_1\rangle \\ \vdots \\ |S_{2^n}\rangle \end{pmatrix} = \mathbf{C} \begin{pmatrix} |A_1\rangle \\ \vdots \\ |A_{2^n}\rangle \end{pmatrix}, \tag{8a}$$

$$\mathbf{C} = (C_{ij}) = (\langle A_j | S_i \rangle). \tag{8b}$$

The two basis sets are connected via a unitary operator $\hat{U}$ on $\mathcal{H}_l^{\otimes n}$ as

$$|\omega_i\rangle = \hat{U}^\dagger |A_i\rangle \quad (i = 1, \ldots, 2^n), \tag{9a}$$

where

$$\hat{U}^\dagger = \sum_{i,j}^{2^n} u_{ji} |A_j\rangle\langle A_i|, \quad u_{ji} = (\mathbf{B}^{-1}\mathbf{C})_{ij}. \tag{9b}$$

Here the optimum collective decoding can be described by the set $\{\hat{U}^\dagger|A_1\rangle, \ldots, \hat{U}^\dagger|A_M\rangle\}$. The minimum error probability is obtained as

$$P_e(\text{opt}) = 1 - \sum_{m=1}^{M} \zeta_m |\langle S_m | \hat{U}^\dagger | A_m \rangle|^2. \tag{10}$$

This clearly means that the optimum collective decoding $\{|\omega_m\rangle\}$ can be effected by (i) transforming the code-word states $\{|S_m\rangle\}$ by the unitary transformation $\hat{U}$, and (ii) applying the von Neumann measurement $\{|A_m\rangle\langle A_m|\}$ to the transformed code-word states. This type of detection scheme is called the received quantum state control [17,18]. The final measurement is actually a separate measurement distinguishing each output letter state as $|a\rangle$ or $|b\rangle$ sequentially.

Now the measurement basis vectors need not be the above combination but may be chosen as any combination of M distinct elements of product basis vectors $\{|A_{i_1}\rangle, \ldots, |A_{i_M}\rangle\}$. Depending on the choice, the matrix $\mathbf{C}$ should be redefined by rearranging the order of elements of the vectors of the right-hand side in Eq. (8a), as $\{|A_{i_1}\rangle, \ldots, |A_{i_M}\rangle, |A_{i_{M+1}}\rangle, \ldots, |A_{i_{2^n}}\rangle\}$. Then the unitary operator $\hat{U}$ constructed by Eq. (9) transforms the code-word states adaptively to the chosen basis vectors $\{|A_m\rangle\}$ such that the minimum average error probability is attained by the separate measurement. Note that $\hat{U}$ acts on the 2^n-dimensional Hilbert space $\mathcal{H}_l^{\otimes n}$ rather than the M-dimensional space $\mathcal{H}_s$. After the unitary transformation has been carried out, the resulting sequences at the final measurement always lie in the space spanned by $\{|A_{i_1}\rangle, \ldots, |A_{i_M}\rangle\}$. If the transformation is skipped, all the product basis vectors will come out. The channel model of this scheme is illustrated in the case of $n=3$ and $M=4$ (see Fig. 1).

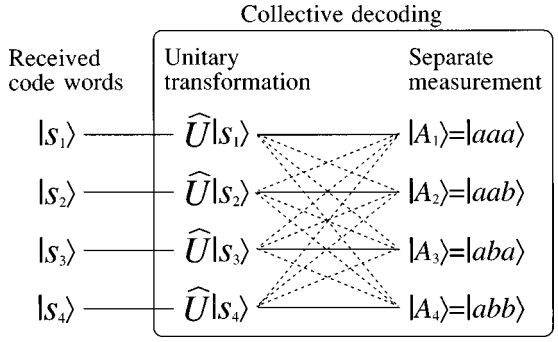

FIG. 1. The channel model obtained by decomposing the collective decoding into the unitary transformation and the separate measurement in the case of $n=3$ and $M=4$.

This kind of decomposition makes it easier to design the collective decoding systematically. As the final measurement on each letter-state system, each process of which is described by the set $\{|a\rangle, |b\rangle\}$, the most suitable and implementable method may be chosen. The main problem is the realization of the unitary transformation as an *adaptor* to the final measurement. Corresponding physical processes are sometimes subtle. The difficulty of finding them may be case by case depending on what kind of letter-state system is provided. However, if a 2-bit gate acting on quantum bits (qubits) made of the two basis vectors $\{|a\rangle, |b\rangle\}$ is available, the required unitary transformation can be, in principle, effected as a quantum circuit used in quantum computation.

Barenco *et al.* already showed that an exact *simulation* of any discrete unitary operator can be carried out by using a quantum computing network [19]. This story can be directly translated into the real operation of $\hat{U}$ on the code-word states. At first, $\hat{U}$ is decomposed into U(2) operators $\hat{T}_{[j,i]}$ by applying the algorithm proposed by Reck and others [20] as

$$\hat{U} = \hat{T}_{[2,1]} \hat{T}_{[3,1]} \cdots \hat{T}_{[2^n, 2^n-2]} \hat{T}_{[2^n, 2^n-1]}, \tag{11a}$$

where

$$\hat{T}_{[j,i]} = \exp[-\gamma_{ji}(|A_i\rangle\langle A_j| - |A_j\rangle\langle A_i|)]. \tag{11b}$$

Then the above two-dimensional rotations $\hat{T}_{[j,i]}$ are converted into quantum circuits by using the formula established by Barenco *et al.* [19]. Here the following point should be noted. Since qubits are letter states themselves constituting the code-word states, the gates should consist of the single physical species from which the letter states are made. Such

gates known so far are Sleator and Weinfurter's gate consisting of two-state atoms [21] and the quantum phase gate acting on two photon-polarization states [22]. In the following paper, examples of the required quantum circuits are described based on such gates.

IV. SUPERADDITIVITY IN CAPACITY OF QUANTUM CHANNEL

The purpose of this section is to demonstrate simple codes that show the superadditivity in capacity. Let us introduce some definitions. Prepare an ensemble of letter states $\hat{\mathbf{s}}=\{\hat{s}_1,\ldots,\hat{s}_l\}$ in a Hilbert space $\mathcal{H}_l$. They represent input letters $\{1,\ldots,l\}$. Let $\boldsymbol{\xi}=\{\xi_1,\ldots,\xi_l\}$ be corresponding prior probabilities. A decoding process is described by a probability operator measure on $\mathcal{H}_l$, $\hat{\boldsymbol{\pi}}=\{\hat{\pi}_1,\ldots,\hat{\pi}_{l'}\}$ representing output letters $\{1,\ldots,l'\}$. We call the mapping $\{1,\ldots,l\}\mapsto\{1,\ldots,l'\}$ the initial quantum channel. Fixing $\boldsymbol{\xi}$, $\hat{\mathbf{s}}$, and $\hat{\boldsymbol{\pi}}$, the mutual information is defined as

$$I_1(\boldsymbol{\xi},\hat{\mathbf{s}}:\hat{\boldsymbol{\pi}})=\sum_{i=1}^{l}\xi_i\sum_{j=1}^{l'}p(j|i)\log_2\frac{p(j|i)}{\sum_{k=1}^{l}\xi_k p(j|k)}, \quad (12)$$

where $p(j|i)=\mathrm{Tr}(\hat{\pi}_j\hat{s}_i)$ is a conditional probability that the letter j is chosen when the letter i has been sent. The maximum value of this quantity optimized with respect to $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\pi}}$ is usually denoted as C_1,

$$C_1(\hat{\mathbf{s}})\equiv\sup_{\boldsymbol{\xi},\hat{\boldsymbol{\pi}}}I_1(\boldsymbol{\xi},\hat{\mathbf{s}}:\hat{\boldsymbol{\pi}}). \quad (13)$$

A basic channel coding consists of (i) concatenation of the letter states into the l^n block sequences $\{\hat{s}_{i_1}\otimes\cdots\otimes\hat{s}_{i_n}\}$, (ii) pruning of them into M-ary code words $\hat{\mathbf{S}}=\{\hat{S}_1,\ldots,\hat{S}_M\}$ which can encode $\log_2 M$ classical bits, and (iii) finding an appropriate decoding POM $\hat{\mathbf{\Pi}}=\{\hat{\Pi}_1,\ldots,\hat{\Pi}_{M'}\}$ in the extended Hilbert space $\mathcal{H}_l^{\otimes n}$. The obtained channel is called the nth extended channel. Assigning input distribution $\boldsymbol{\zeta}=\{\zeta_1,\ldots,\zeta_M\}$ to the code words, the mutual information is defined also for this channel as

$$I_n(\boldsymbol{\zeta},\hat{\mathbf{S}}:\hat{\mathbf{\Pi}})=\sum_{i=1}^{M}\zeta_i\sum_{j=1}^{M'}P(j|i)\log_2\frac{P(j|i)}{\sum_{k=M}^{l}\zeta_k P(j|k)}, \quad (14)$$

where $P(j|i)=\mathrm{Tr}(\hat{\Pi}_j\hat{S}_i)$. Let us define the *nth order capacity* as

$$C_n(\hat{\mathbf{s}})\equiv\sup_{\boldsymbol{\zeta},\hat{\mathbf{\Pi}}}I_n(\boldsymbol{\zeta},\hat{\mathbf{S}}:\hat{\mathbf{\Pi}}). \quad (15)$$

Then generally, $C_n(\hat{\mathbf{s}})\geqslant nC_1(\hat{\mathbf{s}})$ holds for a quantum channel [1,2]. This property is the superadditivity in capacity. One can define the limit $C(\hat{\mathbf{s}})\equiv\lim_{n\to\infty}C_n(\hat{\mathbf{s}})/n$. The quantum channel coding theorem [3–6] says that this $C(\hat{\mathbf{s}})$ is just the attainable rate of asymptotically error free transmission, hence the *intrinsic capacity* of the initial quantum channel, and is exactly equal to the Holevo bound,

$$C(\hat{\mathbf{s}})=\sup_{\boldsymbol{\xi}}\left[H\left(\sum_i\xi_i\hat{s}_i\right)-\sum_i\xi_i H(\hat{s}_i)\right], \quad (16)$$

where $H(\hat{s}_i)\equiv-\mathrm{Tr}(\hat{s}_i\log_2\hat{s}_i)$ is the von Neumann entropy of the density operator $\hat{s}_i$. This theorem ensures that there exist such codes that the decoding error vanishes asymptotically as $n\to\infty$ if the transmission rate $R=(1/n)\log_2 M$ is kept below $C(\hat{\mathbf{s}})$.

A remaining big problem is to find such codes. For this purpose, the first thing to be understood is the superadditivity in capacity. It should be stressed that the strict superadditivity $C_n(\hat{\mathbf{s}})>nC_1(\hat{\mathbf{s}})$ is definitely impossible in a classical memoryless channel. In contrast, a quantum channel has a *memory effect* seen in channel matrix as $P(y_1y_2\cdots y_n|x_1x_2\cdots x_n)\neq\Pi_{i=1}^{n}P(y_i|x_i)$, even when the source of letter states and the physical transmission channel do not have any memory effects. This *memory effect* is caused by the decoding process itself. That is, when a *collective decoding* is applied to the code words, the entanglement structure among the letter states prevents, in general, the channel matrix from being factorized as above. For attaining the strict superadditivity, an appropriate *memory effect* needs to be generated by an appropriate selection of code words *and* a collective decoding for them. This property is thus indeed quantum gain in information transmission.

The essential role of entanglement for the information theoretic quantum gain can be stressed rigorously by the following theorem.

Theorem 2

Suppose that two ensembles of letter states $\hat{\mathbf{s}}^{(1)}=\{\hat{s}_i^{(1)}\}$ in $\mathcal{H}_l^{(1)}$ and $\hat{\mathbf{s}}^{(2)}=\{\hat{s}_j^{(2)}\}$ in $\mathcal{H}_l^{(2)}$ are given, and the first order capacities $C_1(\hat{\mathbf{s}}^{(1)})$ and $C_1(\hat{\mathbf{s}}^{(2)})$ are attained for the prior probabilities $\boldsymbol{\xi}^{(1)}$ and $\boldsymbol{\xi}^{(2)}$, and the detection operators $\hat{\boldsymbol{\pi}}^{(1)}$ and $\hat{\boldsymbol{\pi}}^{(2)}$, respectively. Then the accessible information of the channel with the inputs $\hat{\mathbf{s}}^{(1)}\otimes\hat{\mathbf{s}}^{(2)}$ and the prior probabilities $\boldsymbol{\xi}^{(1)}\otimes\boldsymbol{\xi}^{(2)}$ is given as

$$\sup_{\hat{\mathbf{\Pi}}}I(\boldsymbol{\xi}^{(1)}\otimes\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(1)}\otimes\hat{\mathbf{s}}^{(2)}:\hat{\mathbf{\Pi}})=C_1(\hat{\mathbf{s}}^{(1)})+C_1(\hat{\mathbf{s}}^{(2)}), \quad (17)$$

when

$$\hat{\mathbf{\Pi}}=\hat{\boldsymbol{\pi}}^{(1)}\otimes\hat{\boldsymbol{\pi}}^{(2)}.$$

The proof is given in Appendix A [23].

Now suppose that the supremum in Eq. (13) is attained when $\boldsymbol{\xi}=\boldsymbol{\xi}^*$ and $\hat{\boldsymbol{\pi}}=\hat{\boldsymbol{\pi}}^*$. If all of the l^n sequences $\{\hat{s}_{i_1}\otimes\cdots\otimes\hat{s}_{i_n}\}$ are used as the code words with fixed prior probabilities $\{\xi_{i_1}^*\times\cdots\times\xi_{i_n}^*\}$, according to the above theorem, the optimum decoding $\hat{\mathbf{\Pi}}$ maximizing the mutual information $I_n(\boldsymbol{\xi}^{*\otimes n},\hat{\mathbf{s}}^{\otimes n},\hat{\mathbf{\Pi}})$ is

$$\hat{\Pi}_{i_1,\ldots,i_n}=\hat{\pi}_{i_1}^{*}\otimes\cdots\otimes\hat{\pi}_{i_1}^{*}. \tag{18}$$

The accessible information is simply n times $C_1(\hat{\mathbf{s}})$,

$$\sup_{\hat{\Pi}} I_n(\boldsymbol{\xi}^{*\otimes n},\hat{\mathbf{s}}^{\otimes n},\hat{\Pi})=nC_1(\hat{\mathbf{s}}). \tag{19}$$

Thus in this restricted case, there is no room for entanglement to be generated. So this case provides a *threshold point* for the information theoretic quantum gain. Once the input probabilities are redistributed so as to reduce weights of some code words, the entanglement becomes possible, and the quantum gain can be obtained by an appropriate decoding $\hat{\Pi}$. Concerning the threshold point, it might be worth mentioning a similar theorem in terms of the average error probability.

Theorem 3

For given states $\{\hat{s}_i\}$ with prior probabilities $\{\xi_i\}$, let $\{\hat{\pi}_i\}$ be the optimum POM minimizing the average error probability. Then in distinguishing the product states $\{\hat{s}_{i_1}\otimes\cdots\otimes\hat{s}_{i_n}\}$ associated with the prior probabilities $\{\xi_{i_1}\times\cdots\times\xi_{i_n}\}$, the optimum POM minimizing the average error probability is

$$\hat{\Pi}_{i_1,\ldots,i_n}=\hat{\pi}_{i_1}\otimes\cdots\otimes\hat{\pi}_{i_n} \tag{20}$$

and the minimum average error probability is given as

$$P_e(\mathrm{opt})=1-(\mathrm{Tr}\ \hat{\upsilon})^n, \tag{21}$$

where $\hat{\upsilon}\equiv\Sigma_i\xi_i\hat{\pi}_i$ is a Lagrange operator.

Its proof is given in Appendix B [23]. It should be noted that the optimum POM for the letter states $\{\hat{\pi}_i\}$ need not, in general, coincide with the one for the mutual information, $\{\hat{\pi}_i^{*}\}$ in Eq. (18), even for the same ensemble $\{\hat{s}_i\}$. The result of this theorem will be discussed later with the example of the channel coding.

The simple example of a channel coding showing the superadditivity in capacity was already given by the authors in the case of the third extension of the binary pure-state channel [24]. Here we generalize this example into nth order extension, and demonstrate the relation

$$\sup_{\hat{\Pi}} I_n(\boldsymbol{\zeta},\hat{\mathbf{s}}^{\otimes n},\hat{\Pi})>nC_1(\hat{\mathbf{s}}), \tag{22}$$

which ensures the strict superadditivity. The initial channel is made of binary letters $\hat{\mathbf{s}}=\{|+\rangle\langle+|,|-\rangle\langle-|\}$ whose inner product $\kappa=\langle+|-\rangle$ is assumed to be real. It is well known that the first order capacity is achieved by the symmetric channel with the detection operators $\{\hat{\pi}_1,\hat{\pi}_2\}$ which minimizes the average error probability [10–12]. These operators are given as $\hat{\pi}_i=|\omega_i\rangle\langle\omega_i|$ with

$$|\omega_1\rangle=\sqrt{\frac{1-p}{1-\kappa^2}}|+\rangle-\sqrt{\frac{p}{1-\kappa^2}}|-\rangle, \tag{23a}$$

$$|\omega_2\rangle=-\sqrt{\frac{p}{1-\kappa^2}}|+\rangle+\sqrt{\frac{1-p}{1-\kappa^2}}|-\rangle, \tag{23b}$$

where $p=(1-\sqrt{1-\kappa^2})/2$ is the minimum average error probability. The first order capacity is given simply as

$$C_1(\hat{\mathbf{s}})=1+(1-p)\log_2(1-p)+p\ \log_2 p. \tag{24}$$

In the nth order extension, half of all the 2^n sequences are used as the code-word states and are input to the channel with equal prior probabilities. Such 2^{n-1} code-word states are generated in a recursive manner from the four code-word states $\{|+++\rangle,|+--\rangle,|--+\rangle,|-+-\rangle\}$ in the third order extension, where $|+--\rangle\equiv|+\rangle\otimes|-\rangle\otimes|-\rangle$, etc. That is, defining vectors consisting of bra-state vectors of code-word states,

$$\boldsymbol{\gamma}^{(3)}\equiv\begin{pmatrix}\langle+++|\\ \langle+--|\\ \langle-+-|\\ \langle--+|\end{pmatrix},\quad \boldsymbol{\lambda}^{(3)}\equiv\begin{pmatrix}\langle---|\\ \langle-++|\\ \langle+-+|\\ \langle++-|\end{pmatrix}, \tag{25}$$

they are given as

$$\boldsymbol{\gamma}^{(n)}=\begin{pmatrix}\langle S_1|\\ \langle S_2|\\ \vdots\\ \langle S_{2^{n-1}}|\end{pmatrix}\equiv\begin{pmatrix}\langle+|\otimes\boldsymbol{\gamma}^{(n-1)}\\ \langle-|\otimes\boldsymbol{\lambda}^{(n-1)}\end{pmatrix}. \tag{26}$$

This code-word selection can be specified by the notation $[[n,n-1,2]]$ according to the nomenclature of coding theory.

These code words are decoded by the square-root measurement. The measurement basis vectors are defined as

$$|\mu_i\rangle\equiv\hat{\rho}^{-1/2}|S_i\rangle,\quad \hat{\rho}\equiv\sum_{i=1}^{2^{n-1}}|S_i\rangle\langle S_i|, \tag{27}$$

where the prior probabilities are not included in the density matrix $\hat{\rho}$ unlike Eq. (2), simply for mathematical convenience. As it will become clear soon, these basis vectors effect the *optimum collective decoding* for the above code words. We have to evaluate the channel matrix $P(j|i)=|\langle\mu_j|S_i\rangle|^2$. The Gram matrix of the code words is given by

$$\boldsymbol{\Gamma}^{(n)}=\boldsymbol{\gamma}^{(n)}\cdot\boldsymbol{\gamma}^{(n)\dagger}=\begin{pmatrix}\boldsymbol{\Gamma}^{(n-1)} & \kappa^2\boldsymbol{\Lambda}^{(n-1)}\\ \kappa^2\boldsymbol{\Lambda}^{(n-1)} & \boldsymbol{\Gamma}^{(n-1)}\end{pmatrix}, \tag{28a}$$

where

$$\boldsymbol{\Lambda}^{(n-1)}=\frac{1}{\kappa}\boldsymbol{\gamma}^{(n-1)}\cdot\boldsymbol{\lambda}^{(n-1)\dagger}=\begin{pmatrix}\boldsymbol{\Gamma}^{(n-2)} & \boldsymbol{\Lambda}^{(n-2)}\\ \boldsymbol{\Lambda}^{(n-2)} & \boldsymbol{\Gamma}^{(n-2)}\end{pmatrix}, \tag{28b}$$

and

$$\mathbf{\Gamma}^{(3)}=\begin{pmatrix} 1 & \kappa^2 & \kappa^2 & \kappa^2 \\ \kappa^2 & 1 & \kappa^2 & \kappa^2 \\ \kappa^2 & \kappa^2 & 1 & \kappa^2 \\ \kappa^2 & \kappa^2 & \kappa^2 & 1 \end{pmatrix}, \tag{28c}$$

$$\mathbf{\Lambda}^{(3)}=\begin{pmatrix} \kappa^2 & 1 & 1 & 1 \\ 1 & \kappa^2 & 1 & 1 \\ 1 & 1 & \kappa^2 & 1 \\ 1 & 1 & 1 & \kappa^2 \end{pmatrix}. \tag{28d}$$

$\mathbf{\Gamma}^{(n)}$ and $\mathbf{\Lambda}^{(n)}$ can be diagonalized by a $2^{n-1}\times 2^{n-1}$ matrix

$$\mathbf{Q}^{(n)}=\mathbf{H}_{2^{n-1}}/\sqrt{2^{n-1}}, \tag{29a}$$

where $\mathbf{H}_{2^{n-1}}$ is the Hadamard matrix defined by

$$\mathbf{H}_{2k}=\begin{pmatrix} \mathbf{H}_k & \mathbf{H}_k \\ \mathbf{H}_k & -\mathbf{H}_k \end{pmatrix}, \quad \mathbf{H}_2=\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{29b}$$

The diagonalized matrices,

$$\mathbf{G}^{(n)}\equiv\mathbf{Q}^{(n)\dagger}\mathbf{\Gamma}^{(n)}\mathbf{Q}^{(n)}, \tag{30a}$$

$$\mathbf{F}^{(n)}\equiv\mathbf{Q}^{(n)\dagger}\mathbf{\Lambda}^{(n)}\mathbf{Q}^{(n)}, \tag{30b}$$

can be decomposed into $2^{n-2}\times 2^{n-2}$ matrices $\mathbf{G}^{(n-1)}$ and $\mathbf{F}^{(n-1)}$ as

$$\mathbf{G}^{(n)}=\begin{pmatrix} \mathbf{G}^{(n-1)}+\kappa^2\mathbf{F}^{(n-1)} & 0 \\ 0 & \mathbf{G}^{(n-1)}-\kappa^2\mathbf{F}^{(n-1)} \end{pmatrix}, \tag{31a}$$

$$\mathbf{F}^{(n)}=\begin{pmatrix} \mathbf{G}^{(n-1)}+\mathbf{F}^{(n-1)} & 0 \\ 0 & \mathbf{G}^{(n-1)}-\mathbf{F}^{(n-1)} \end{pmatrix}. \tag{31b}$$

After recursive decompositions, they can be represented as

$$\mathbf{G}^{(n)}=\begin{pmatrix} \mathbf{A}(n,1) & & \\ & \ddots & \\ & & \mathbf{A}(n,2^{n-3}) \end{pmatrix}, \tag{32a}$$

$$\mathbf{F}^{(n)}=\begin{pmatrix} \mathbf{B}(n,1) & & \\ & \ddots & \\ & & \mathbf{B}(n,2^{n-3}) \end{pmatrix}, \tag{32b}$$

where $\mathbf{A}(n,k)$ and $\mathbf{B}(n,k)$ $(k=1,\ldots,2^{n-3})$ are 4×4 block matrices defined by

$$\mathbf{A}(n,k)=a(n,k)\mathbf{G}^{(3)}+b(n,k)\mathbf{F}^{(3)}, \tag{33a}$$

$$\mathbf{B}(n,k)=c(n,k)\mathbf{G}^{(3)}+d(n,k)\mathbf{F}^{(3)}, \tag{33b}$$

with

$$\mathbf{G}^{(3)}=\begin{pmatrix} 1+3\kappa^2 & & & \\ & 1-\kappa^2 & & \\ & & 1-\kappa^2 & \\ & & & 1-\kappa^2 \end{pmatrix}, \tag{34a}$$

$$\mathbf{F}^{(3)}=\begin{pmatrix} 3+\kappa^2 & & & \\ & -1+\kappa^2 & & \\ & & -1+\kappa^2 & \\ & & & -1+\kappa^2 \end{pmatrix}. \tag{34b}$$

The coefficients in Eq. (33) are determined by the following recursive formula for $k=1,\ldots,2^{n-4}$,

$$\begin{aligned} a(n,k)&=a(n-1,k)+\kappa^2 c(n-1,k), \\ a(n,2^{n-4}+k)&=a(n-1,k)-\kappa^2 c(n-1,k), \\ b(n,k)&=b(n-1,k)+\kappa^2 d(n-1,k), \\ b(n,2^{n-4}+k)&=b(n-1,k)-\kappa^2 d(n-1,k), \end{aligned} \tag{35a}$$

$$\begin{aligned} c(n,k)&=a(n-1,k)+c(n-1,k), \\ c(n,2^{n-4}+k)&=a(n-1,k)-c(n-1,k), \\ d(n,k)&=b(n-1,k)+d(n-1,k), \\ d(n,2^{n-4}+k)&=b(n-1,k)-d(n-1,k), \end{aligned}$$

with the initial values

$$\begin{aligned} a(4,1)&=a(4,2)=1, \\ b(4,1)&=-b(4,2)=\kappa^2, \\ c(4,1)&=c(4,2)=d(4,1)=-d(4,2)=1. \end{aligned} \tag{35b}$$

Thus the diagonal matrices $\mathbf{A}(n,k)$ are obtained, and the square root of the Gram matrix is given as

$$\sqrt{\mathbf{\Gamma}^{(n)}}=\mathbf{Q}^{(n)}\begin{pmatrix} \sqrt{\mathbf{A}(n,1)} & & \\ & \ddots & \\ & & \sqrt{\mathbf{A}(n,2^{n-3})} \end{pmatrix}\mathbf{Q}^{(n)\dagger}. \tag{36}$$

For representing the result, let us define

$$\alpha(n,k)\equiv\sqrt{(1+3\kappa^2)a(n,k)+(3+\kappa^2)b(n,k)}, \tag{37a}$$

$$\beta(n,k)\equiv\sqrt{(1-\kappa^2)[a(n,k)-b(n,k)]}, \tag{37b}$$

and

$$\mathbf{D}(n,k)\equiv\mathbf{Q}^{(3)}\sqrt{\mathbf{A}(n,k)}\mathbf{Q}^{(3)\dagger} \tag{37c}$$

$$=\begin{pmatrix} \mu(n,k) & \nu(n,k) & \nu(n,k) & \nu(n,k) \\ \nu(n,k) & \mu(n,k) & \nu(n,k) & \nu(n,k) \\ \nu(n,k) & \nu(n,k) & \mu(n,k) & \nu(n,k) \\ \nu(n,k) & \nu(n,k) & \nu(n,k) & \mu(n,k) \end{pmatrix}, \tag{37d}$$

where

$$\mu(n,k)\equiv\tfrac{1}{4}[\alpha(n,k)+3\beta(n,k)], \tag{37e}$$

$$\nu(n,k)\equiv\tfrac{1}{4}[\alpha(n,k)-\beta(n,k)]. \tag{37f}$$

Then $\sqrt{\mathbf{\Gamma}^{(n)}}$ can be represented as

$$\sqrt{\mathbf{\Gamma}^{(n)}}=\begin{pmatrix} \mathbf{R}(n,1) & \mathbf{R}(n,2) & \mathbf{R}(n,3) & \mathbf{R}(n,4) & \cdots \\ \mathbf{R}(n,2) & \mathbf{R}(n,1) & \mathbf{R}(n,4) & \mathbf{R}(n,3) & \cdots \\ \mathbf{R}(n,3) & \mathbf{R}(n,4) & \mathbf{R}(n,1) & \mathbf{R}(n,2) & \cdots \\ \mathbf{R}(n,4) & \mathbf{R}(n,3) & \mathbf{R}(n,2) & \mathbf{R}(n,1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{R}(n,2^{n-3}-1) & \mathbf{R}(n,2^{n-3}) & \mathbf{R}(n,2^{n-3}-3) & \mathbf{R}(n,2^{n-3}-2) & \cdots \\ \mathbf{R}(n,2^{n-3}) & \mathbf{R}(n,2^{n-3}-1) & \mathbf{R}(n,2^{n-3}-2) & \mathbf{R}(n,2^{n-3}-3) & \cdots \end{pmatrix}, \tag{38a}$$

where

$$\begin{pmatrix} \mathbf{R}(n,1) \\ \vdots \\ \mathbf{R}(n,2^{n-3}) \end{pmatrix}\equiv\frac{1}{2^{n-3}}\mathbf{H}_{2^{n-1}}\begin{pmatrix} \mathbf{D}(n,1) \\ \vdots \\ \mathbf{D}(n,2^{n-3}) \end{pmatrix}. \tag{38b}$$

$\mathbf{R}(n,k)$ can be further arranged in the form

$$\mathbf{R}(n,k)=\begin{pmatrix} u(n,k) & v(n,k) & v(n,k) & v(n,k) \\ v(n,k) & u(n,k) & v(n,k) & v(n,k) \\ v(n,k) & v(n,k) & u(n,k) & v(n,k) \\ v(n,k) & v(n,k) & v(n,k) & u(n,k) \end{pmatrix}. \tag{39}$$

The two kinds of components $u(n,k)$ and $v(n,k)$ can be calculated by

$$\begin{pmatrix} u(n,1) \\ \vdots \\ u(n,2^{n-3}) \end{pmatrix}\equiv\frac{1}{2^{n-3}}\mathbf{H}_{2^{n-3}}\begin{pmatrix} \mu(n,1) \\ \vdots \\ \mu(n,2^{n-3}) \end{pmatrix}, \tag{40a}$$

$$\begin{pmatrix} v(n,1) \\ \vdots \\ v(n,2^{n-3}) \end{pmatrix}\equiv\frac{1}{2^{n-3}}\mathbf{H}_{2^{n-3}}\begin{pmatrix} \nu(n,1) \\ \vdots \\ \nu(n,2^{n-3}) \end{pmatrix}. \tag{40b}$$

After squaring each component of $\sqrt{\mathbf{\Gamma}^{(n)}}$, the channel matrix $P(j|i)$ can be obtained. According to the symmetry seen in Eq. (38), it is easy to see that the mutual information is given as

$$I_n(S:\mu)=n-1+\sum_{k=1}^{2^{n-3}}[u(n,k)^2\log_2 u(n,k)^2 + 3v(n,k)^2\log_2 v(n,k)^2]. \tag{41}$$

In order to see the quantum gain, the difference between the mutual information per letter state $I_n(S:\mu)/n$ and $C_1(\hat{\mathbf{s}})$ is plotted as functions of κ in Fig. 2. The dashed line corresponds to the case of $n=2$ where the two code-word states $\{|++\rangle,|--\rangle\}$ are sent with the same prior probabilities and are detected by the optimum measurement minimizing the average error probability. In this case, the positive quantum gain was not found in the whole region of κ. For $n=3-13$ (solid lines), the difference becomes positive at the larger side of κ. This positive gain clearly shows the superadditivity in capacity. Let κ_* be the value of $\kappa(<1)$ for which the difference becomes zero. Then for $\kappa_*<\kappa<1$ the difference is always positive, and as n increases, κ_* decreases so that the positive gain appears in a wider region of κ. This relation is plotted in Fig. 3. The circles represent the points (n,κ_*). The solid line is just a guide for the eye. The dashed line corresponds to the curve of $n=2\kappa^{-1.5}$. This figure may provide a rough estimate for n in order to obtain the positive gain. That is, for a given κ, one may guess that the superadditivity will appear when the order of extension for our $[[n,n-1,2]]$ code is taken as an integer n larger than $2\kappa^{-1.5}$. Unfortunately we did not succeed in giving a more

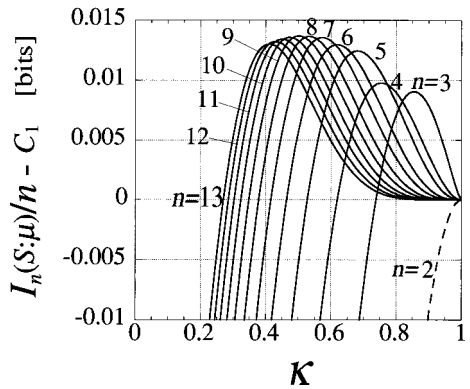

FIG. 2. The difference between the mutual information per letter state $I_n(S:\mu)/n$ and $C_1(\hat{\mathbf{s}})$ for $n=2-13$. The dashed line corresponds to the case of $n=2$, while the solid lines correspond to the case of $n=3-13$.

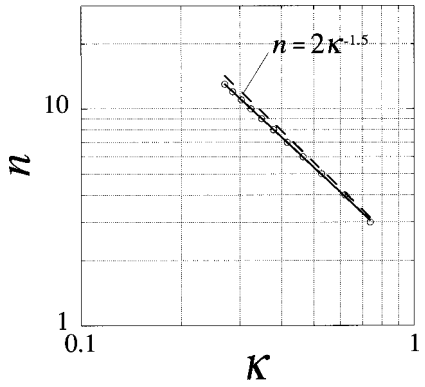

FIG. 3. The relation between n and κ_* (<1) at which $I_n(S:\mu)/n=C_1(\hat{\mathbf{s}})$ holds. κ_*'s are denoted by the circles. The solid line is just a guide for the eye. The dashed line corresponding to the curve of $n=2\kappa^{-1.5}$.

rigorous condition. The maximum amount of the positive gain is still quantitatively unsatisfactory compared with the gap between the intrinsic capacity $C(\hat{\mathbf{s}})$ and the first order capacity $C_1(\hat{\mathbf{s}})$. Actually it is less than 10% of the maximum gap. In the case of $n=9$ for which the maximum gain was obtained, $C(\hat{\mathbf{s}})$, $I_9(S:\mu)/9$ and $C_1(\hat{\mathbf{s}})$ versus κ are plotted in Fig. 4.

As far as the minimum average error probability is concerned, the square-root measurement we used [Eq. (27)] is the optimum for our $[[n,n-1,2]]$ code [Eq. (26)] because all of the diagonal components of $\sqrt{\mathbf{\Gamma}^{(n)}}$ are equal to $u(n,1)$ as seen from Eqs. (38) and (39), for which Theorem 1 holds. The minimum average error probabilities versus κ are plotted for $n=3$, 5, 7, 9, 11, 13 by the solid lines in Fig. 5. For a fixed κ, they increase as with n. The dotted lines represent the minimum average error probabilities corresponding to the *threshold points*, that is, Eq. (21). Although the error probabilities of our code $[[n,n-1,2]]$ are smaller than those of the *threshold points*, they are still larger than $p=(1-\sqrt{1-\kappa^2})/2$ (the minimum bit error) at the larger side of κ. In spite of this, the quantum gain $I_n(S:\mu)/n>C_1(\hat{\mathbf{s}})$ reveals itself within such regions.

The same tendency could be seen in other codes. Let us consider the so-called simplex code $[[2^r-1,r,2^{r-1}]]$, for example. All the code words are the same distance apart. Let $n=2^r-1$ and $M=2^r$. Suppose that M-ary code words are

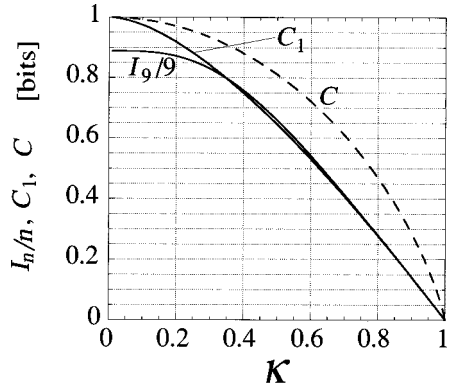

FIG. 4. $C(\hat{\mathbf{s}})$, $I_9(S:\mu)/9$, and $C_1(\hat{\mathbf{s}})$ as functions of κ.

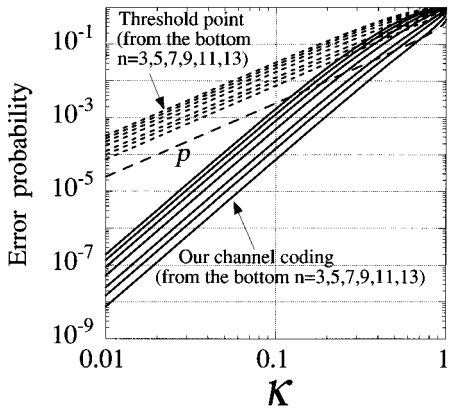

FIG. 5. The minimum average error probabilities corresponding to the code $[[n,n-1,2]]$ (solid lines), the initial channel, i.e., p (dashed line), and the *threshold points* (dotted lines) as functions of κ.

used with equal prior probabilities. Then the square-root measurement is again the optimum collective decoding for them. Defining

$$\alpha(n)\equiv\sqrt{1+(M-1)\kappa^{M/2}}, \tag{42a}$$

$$\beta(n)\equiv\sqrt{1-\kappa^{M/2}}, \tag{42b}$$

it is straightforward to see

$$(\sqrt{\mathbf{\Gamma}^{(n)}})_{ii}=u(n)\equiv\frac{1}{M}[\alpha(n)+(M-1)\beta(n)], \tag{43a}$$

$$(\sqrt{\mathbf{\Gamma}^{(n)}})_{ij}=v(n)\equiv\frac{1}{M}[\alpha(n)-\beta(n)], \quad i\neq j. \tag{43b}$$

The mutual information is then given by

$$I_n(S:\mu)=\log_2 M+u(n)^2\log_2 u(n)^2 +(M-1)v(n)^2\log_2 v(n)^2. \tag{44}$$

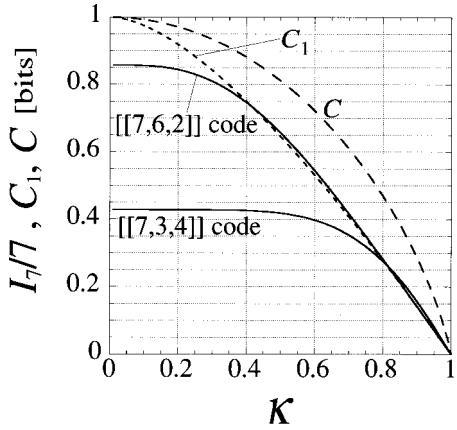

FIG. 6. $C(\hat{\mathbf{s}})$, $I_7(S:\mu)/7$ for both the [[7,3,4]] simplex code and the [[7,6,2]] code, and $C_1(\hat{\mathbf{s}})$ as functions of κ.

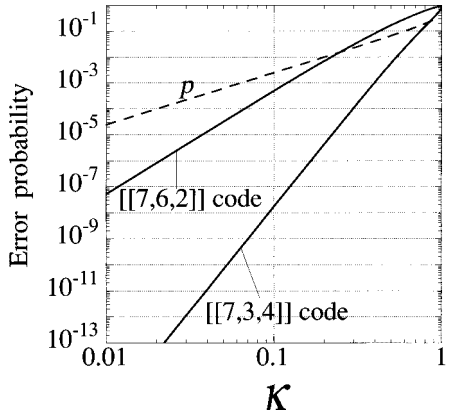

FIG. 7. The minimum average error probabilities corresponding to the [[7,3,4]] simplex code, the [[7,6,2]] code, and the initial channel, i.e., p (dashed line) as functions of κ.

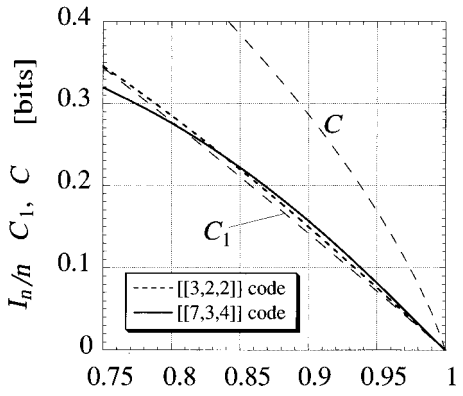

FIG. 8. The mutual information per letter corresponding to the [[7,3,4]] simplex code (solid line) and the code [[3,2,2]] (dashed line) as functions of κ.

This code is compared with the previous one at $n=7$ in terms of both the mutual information per letter state and the minimum average error probability in Figs. 6 and 7, respectively. The [[7,3,4]] simplex code has higher distinguishability of the code words than the [[7,6,2]] code so that the minimum average error is much smaller, while its mutual information is not necessarily larger than the latter. The former overcomes the latter only in the region $0.82<\kappa<1$. Around this region, the minimum average error probability is, again, larger than the minimum bit error p.

This tendency may be understood as a result from the facts that in order to produce the quantum gain a quantum interference among the code-word states must occur to reduce certain components of the channel matrix, and that such a quantum interference occurs more drastically when the nonorthogonality of the code-word states is larger, hence on the larger side of κ. On the other hand, the nonorthogonality causes a certain amount of decoding error as well. This decreases the amount of transmittable information, while the quantum gain may appear if the quantum interference reduces certain components of the channel matrix in a proper manner. Thus at shorter block length, the quantum gain as the difference $I_n/n-C_1>0$ is likely to appear on the larger side of κ being accompanied by a certain amount of decoding error. A rough guide in order to construct a channel attaining the quantum gain for a given κ is as follows: the ratio of the number of the message bits k to the block length n should be taken larger than $C_1(\hat{\mathbf{s}})$ at this κ first, and then code-word states should be selected with a distance as equally apart as possible. As κ becomes closer to the unity, it becomes more effective in obtaining the quantum gain to take the ratio k/n small and to select the code-word states being distant. As an example, the simplex code [[7,3,4]] is compared with the code [[3,2,2]] in the region $0.75<\kappa<1$ in Fig. 8. For $\kappa>0.85$, the [[7,3,4]] code (solid line) is more efficient in terms of the mutual information than the [[3,2,2]] code (dotted line). For constructing codes such that the decoding error can be as small as possible and the rate can reach the Holevo bound, a larger block length at which the typical subspace can be well defined is necessary. Practical methods for obtaining larger quantum gain must be studied in great detail along this direction.

V. CONCLUDING REMARKS

The initial channel considered in this paper is the binary symmetric pure-state channel which is the simplest quantum channel. When the binary letter states are orthogonal, the channel is error free, and there is no quantum regime, that is, nth order capacity obviously satisfies the strict additivity $C_n=nC_1$. When they are nonorthogonal, i.e., $0<\kappa<1$, the strict *superadditivity* $C_n>nC_1$, in turn, reveals itself. What we have shown in this paper is a demonstration of $I_n>nC_1$ that ensures the strict *superadditivity*.

The first part of this paper was devoted to the optimum collective decoding of the code-word states at the minimum average error probability. The scheme we proposed consists of the unitary transformation and the separate measurement. The unitary transformation generates appropriate superposition states among the code-word states such that the minimum average error is attained at the separate measurement. These output states are not separable into the letter states any more. Minimization of decoding error is just a manifestation of the optimum use of quantum interference associated with this kind of superposition states. The required unitary transformation is essentially a conditional dynamics in a higher-dimensional Hilbert space, and is realized by a quantum circuit which is capable of manipulating each letter state in a conditional manner depending on the other letter states. Thus our scheme suggests a state-of-the-art quantum decoder structure.

Quantum channels involving the above collective decoding have a *memory effect*, i.e., $P(y_1y_2\cdots y_n|x_1x_2\cdots x_n)\neq\Pi_{i=1}^{n}P(y_i|x_i)$. This inseparability of quantum channel is a direct origin of the quantum gain $I_n>nC_1$. Only when code-word states are selected suitably does this inseparability lead to the quantum gain. We gave a heuristic approach to attain this gain for a given letter-state ensemble. Our examples are always accompanied by a larger amount of decoding error than the minimum average error in the initial channel. As mentioned in the preceding section, this dilemma is because both decoding error and the quantum gain are originated from the nonorthogonality of the letter states.

Although some basic aspects for realizing the quantum gain were clarified by this paper, practical codes that transmit classical alphabet faithfully at the maximum rate are still completely unknown. Even in classical information theory,

realization of such codes that achieve asymptotically error free transmission at the rate C_1 is very difficult. It might be an interesting problem to consider an application of some conventional error correcting codes to the n-product quantum channels described in this paper. This will lead to realization of asymptotically error free transmission at the rate I_n/n $(>C_1)$. For the ultimate quantum channel coding, typicality of the Hilbert space spanned by the code-word states and sophisticated quantum error correction might be considered together.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Holevo of Steklov Mathematical Institute and Dr. Usuda of the Nagoya Institute of Technology for giving crucial comments on this work. They would also like to thank Dr. C. A. Fuchs of the California Institute of Technology, Dr. C. H. Bennett of the IBM T. J. Watson research center, Dr. M. Ban of Hitachi Advanced Research Laboratory, and Dr. K. Yamazaki and Dr. M. Osaki of Tamagawa University, for their helpful discussions.

APPENDIX A: PROOF OF THEOREM 2

We first prove the following lemma.

Lemma

$$\sup_{\hat{\mathbf{\Pi}}} I(\boldsymbol{\xi}^{(1)}\otimes\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(1)}\otimes\hat{\mathbf{s}}^{(2)}:\hat{\mathbf{\Pi}})=\sup_{\hat{\boldsymbol{\pi}}^{(1)}} I(\boldsymbol{\xi}^{(1)},\hat{\mathbf{s}}^{(1)}:\hat{\boldsymbol{\pi}}^{(1)})+\sup_{\hat{\boldsymbol{\pi}}^{(2)}} I(\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(2)}:\hat{\boldsymbol{\pi}}^{(2)}). \tag{A1}$$

Proof of lemma

(1) Clearly,

$$\sup_{\hat{\mathbf{\Pi}}} I(\boldsymbol{\xi}^{(1)}\otimes\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(1)}\otimes\hat{\mathbf{s}}^{(2)}:\hat{\mathbf{\Pi}})\geqslant \sup_{\hat{\mathbf{\Pi}}=\hat{\boldsymbol{\pi}}^{(1)}\otimes\hat{\boldsymbol{\pi}}^{(2)}} I(\boldsymbol{\xi}^{(1)}\otimes\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(1)}\otimes\hat{\mathbf{s}}^{(2)}:\hat{\mathbf{\Pi}})=\sup_{\hat{\boldsymbol{\pi}}^{(1)}} I(\boldsymbol{\xi}^{(1)},\hat{\mathbf{s}}^{(1)}:\hat{\boldsymbol{\pi}}^{(1)})+\sup_{\hat{\boldsymbol{\pi}}^{(2)}} I(\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(2)}:\hat{\boldsymbol{\pi}}^{(2)}). \tag{A2}$$

(2) Representing $P(j|i_1,i_2)=\mathrm{Tr}(\hat{\Pi}_j\hat{s}_{i_1}^{(1)}\otimes\hat{s}_{i_2}^{(2)})$, where $\hat{\Pi}_j$ is a POM on $\mathcal{H}_l^{(1)}\otimes\mathcal{H}_l^{(2)}$,

$$I(\boldsymbol{\xi}^{(1)}\otimes\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(1)}\otimes\hat{\mathbf{s}}^{(2)}:\hat{\mathbf{\Pi}})=\sum_j\sum_{i_1,i_2}\xi_{i_1}^{(1)}\xi_{i_2}^{(2)}P(j|i_1,i_2)\log_2\left(\frac{P(j|i_1,i_2)}{\sum_{k_1,k_2}\xi_{k_1}^{(1)}\xi_{k_2}^{(2)}P(j|k_1,k_2)}\right)$$

$$=\sum_j\sum_{i_1,i_2}\xi_{i_1}^{(1)}\xi_{i_2}^{(2)}P(j|i_1,i_2)\left[\log_2\left(\frac{P(j|i_1,i_2)}{\sum_{k_1}\xi_{k_1}^{(1)}P(j|k_1,i_2)}\right)+\log_2\left(\frac{\sum_{k_1}\xi_{k_1}^{(1)}P(j|k_1,i_2)}{\sum_{k_1,k_2}\xi_{k_1}^{(1)}\xi_{k_2}^{(2)}P(j|k_1,k_2)}\right)\right]$$

$$=\sum_{i_2}\xi_{i_2}^{(2)}\left[\sum_j\sum_{i_1}\xi_{i_1}^{(1)}P(j|i_1,i_2)\log_2\left(\frac{P(j|i_1,i_2)}{\sum_{k_1}\xi_{k_1}^{(1)}P(j|k_1,i_2)}\right)\right]+\sum_j\sum_{i_2}\xi_{i_2}^{(2)}$$

$$\times\left[\sum_{i_1}\xi_{i_1}^{(1)}P(j|i_1,i_2)\right]\log_2\left(\frac{\left[\sum_{k_1}\xi_{k_1}^{(1)}P(j|k_1,i_2)\right]}{\sum_{k_2}\xi_{k_2}^{(2)}\left[\sum_{k_1}\xi_{k_1}^{(1)}P(j|k_1,i_2)\right]}\right). \tag{A3}$$

We introduce two kinds of POM on $\mathcal{H}_l^{(1)}$ and $\mathcal{H}_l^{(2)}$ as

$$\hat{\Pi}_{(i_2)j}^{(1)}\equiv\mathrm{Tr}^{(2)}(\hat{\Pi}_j\hat{I}^{(1)}\otimes\hat{s}_{i_2}^{(2)}), \tag{A4}$$

$$\overline{\Pi}_j^{(2)}\equiv\sum_{i_1}\xi_{i_1}^{(1)}\,\mathrm{Tr}^{(1)}(\hat{\Pi}_j\hat{s}_{i_1}^{(1)}\otimes\hat{I}^{(2)}), \tag{A5}$$

where $\mathrm{Tr}^{(i)}$ means taking the trace over the space $\mathcal{H}_l^{(i)}$, and $\hat{I}^{(i)}$ is the identity operator in $\mathcal{H}_l^{(i)}$. Representing the conditional probabilities in Eq. (A3) as

$$P(j|i_1,i_2)=\mathrm{Tr}^{(1)}(\hat{\Pi}_{(i_2)j}^{(1)}\hat{s}_{i_1}^{(1)}), \tag{A6}$$

$$\sum_{i_1}\xi_{i_1}^{(1)}P(j|i_1,i_2)\equiv\mathrm{Tr}^{(2)}(\overline{\Pi}_j^{(2)}\hat{s}_{i_2}^{(2)}), \tag{A7}$$

we see that Eq. (A3) is equivalent to

$$I(\boldsymbol{\xi}^{(1)}\otimes\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(1)}\otimes\hat{\mathbf{s}}^{(2)}:\hat{\mathbf{\Pi}})=\sum_{i_2}\xi^{(2)}_{i_2}I(\boldsymbol{\xi}^{(1)},\hat{\mathbf{s}}^{(1)}:\hat{\mathbf{\Pi}}^{(1)}_{(i_2)})+I(\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(2)}:\overline{\mathbf{\Pi}}^{(2)}). \quad \text{(A8)}$$

Hence

$$\begin{aligned}&\sup_{\hat{\mathbf{\Pi}}} I(\boldsymbol{\xi}^{(1)}\otimes\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(1)}\otimes\hat{\mathbf{s}}^{(2)}:\hat{\mathbf{\Pi}})\\ &=\sum_j \xi^{(2)}_j \sup_{\hat{\mathbf{\Pi}}} I(\boldsymbol{\xi}^{(1)},\hat{\mathbf{s}}^{(1)}:\hat{\mathbf{\Pi}}^{(1)}_j)+\sup_{\hat{\mathbf{\Pi}}} I(\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(2)}:\overline{\mathbf{\Pi}}^{(2)})\\ &\leqslant \sup_{\hat{\boldsymbol{\pi}}^{(1)}} I(\boldsymbol{\xi}^{(1)},\hat{\mathbf{s}}^{(1)}:\hat{\boldsymbol{\pi}}^{(1)})+\sup_{\hat{\boldsymbol{\pi}}^{(2)}} I(\boldsymbol{\xi}^{(2)},\hat{\mathbf{s}}^{(2)}:\hat{\boldsymbol{\pi}}^{(2)}). \quad \text{(A9)}\end{aligned}$$

The two inequalities (A2) and (A9) prove the lemma.

Now suppose that $I(\boldsymbol{\xi}^{(i)},\hat{\mathbf{s}}^{(i)}:\hat{\boldsymbol{\pi}}^{(i)})$ is maximized when $\boldsymbol{\xi}^{(i)}=\boldsymbol{\xi}^{(i)}_*$ and $\hat{\boldsymbol{\pi}}^{(i)}=\hat{\boldsymbol{\pi}}^{(i)}_*$. Then the lemma means that

$$\sup_{\hat{\mathbf{\Pi}}} I(\boldsymbol{\xi}^{(1)}_*\otimes\boldsymbol{\xi}^{(2)}_*,\hat{\mathbf{s}}^{(1)}\otimes\hat{\mathbf{s}}^{(2)}:\hat{\mathbf{\Pi}})=C_1(\hat{\mathbf{s}}^{(1)})+C_1(\hat{\mathbf{s}}^{(2)}). \quad \text{(A10)}$$

On the other hand,

$$\begin{aligned}&I(\boldsymbol{\xi}^{(1)}_*\otimes\boldsymbol{\xi}^{(2)}_*,\hat{\mathbf{s}}^{(1)}\otimes\hat{\mathbf{s}}^{(2)}:\hat{\boldsymbol{\pi}}^{(1)}_*\otimes\hat{\boldsymbol{\pi}}^{(2)}_*)\\ &=I(\boldsymbol{\xi}^{(1)}_*,\hat{\mathbf{s}}^{(1)}:\hat{\boldsymbol{\pi}}^{(1)}_*)+I(\boldsymbol{\xi}^{(2)}_*,\hat{\mathbf{s}}^{(2)}:\hat{\boldsymbol{\pi}}^{(2)}_*)\\ &=C_1(\hat{\mathbf{s}}^{(1)})+C_1(\hat{\mathbf{s}}^{(2)}). \quad \text{(A11)}\end{aligned}$$

These two equations prove the theorem.

APPENDIX B: PROOF OF THEOREM 3

The necessary and sufficient condition that $\{\hat{\pi}_i\}$ is the optimum POM are

$$\hat{\pi}_j(\hat{w}_j-\hat{w}_k)\hat{\pi}_k=0 \quad \forall\ (j,k) \quad \text{(i)},$$

$$\hat{v}-\hat{w}_j\geqslant 0 \quad \forall\ j \quad \text{(ii)},$$

where $\hat{w}_i=\xi_i\hat{s}_i$ and $\hat{v}=\Sigma_j\hat{w}_j\hat{\pi}_j$ are the risk operators and the Lagrange operator, respectively. We would like to prove that $\hat{\Pi}_{j_1,\ldots,j_l}\ (=\hat{\pi}_{j_1}\otimes\cdots\otimes\hat{\pi}_{j_l})$ satisfy

$$\hat{\Pi}_{j_1\cdots j_l}(\hat{W}_{j_1\cdots j_l}-\hat{W}_{k_1\cdots k_l})\hat{\Pi}_{k_1\cdots k_l}=0 \quad \text{(i}'\text{)},$$

$$\hat{\Upsilon}-\hat{W}_{j_1\cdots j_l}\geqslant 0 \quad \text{(ii}'\text{)},$$

where $\hat{W}_{j_1,\ldots,j_l}=\hat{w}_{j_1}\otimes\cdots\otimes\hat{w}_{j_l}$ and $\hat{\Upsilon}=\Sigma_{j_1\cdots j_l}\hat{W}_{j_1,\ldots,j_l}\hat{\Pi}_{j_1,\ldots,j_l}=\hat{v}^{\otimes n}$. Here note that the following formulas:

$$\begin{aligned}A_1\otimes\cdots\otimes A_l-B_1\otimes\cdots\otimes B_l&=(A_1-B_1)\otimes A_2\otimes\cdots\otimes A_l+B_1\otimes(A_2-B_2)\otimes\cdots\otimes A_l+\cdots\\ &\quad+B_1\otimes B_2\otimes\cdots\otimes(B_l-A_l).\end{aligned} \quad \text{(B1)}$$

Then to ensure (i′), we rearrange the left-hand side as

$$\begin{aligned}&\hat{\Pi}_{j_1,\ldots,j_l}(\hat{W}_{j_1,\ldots,j_l}-\hat{W}_{k_1,\ldots,k_l})\hat{\Pi}_{k_1,\ldots,k_l}\\ &=\hat{\pi}_{j_1}\hat{w}_{j_1}\hat{\pi}_{k_1}\otimes\cdots\otimes\hat{\pi}_{j_n}\hat{w}_{j_n}\hat{\pi}_{k_n}\\ &\quad-\hat{\pi}_{j_1}\hat{w}_{k_1}\hat{\pi}_{k_1}\otimes\cdots\otimes\hat{\pi}_{j_n}\hat{w}_{k_n}\hat{\pi}_{k_n},\end{aligned} \quad \text{(B2)}$$

and set $A_m=\hat{\pi}_{j_m}\hat{w}_{j_m}\hat{\pi}_{k_m}$ and $B_m=\hat{\pi}_{j_m}\hat{w}_{k_m}\hat{\pi}_{k_m}$. Since (i) is equivalent to $A_m-B_m=0$, after Eq. (B1) is applied to Eq. (B2) we obtain (i′). Similarly, to show (ii′), we set $A_m=\hat{v}$ and $B_m=\hat{w}_{j_m}$. From the definitions, $A_m\geqslant 0$ and $B_m\geqslant 0$. In addition, (ii) is nothing but $A_m\geqslant B_m$. So when $\hat{\Upsilon}-\hat{W}_{j_1,\ldots,j_l}$ is decomposed by Eq. (B1), its nonnegative definiteness is obvious.

[1] A. S. Holevo, Probl. Peredachi Inf. **9**, 3111 (1973).
[2] A. S. Holevo, Probl. Peredachi Inf. **15**, 3 (1979).
[3] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, Phys. Rev. A **54**, 1869 (1996).
[4] A. S. Holevo, e-print quant-ph/9611023; IEEE Trans. Inf. Theory **IT-44**, 269 (1998).
[5] B. Schumacher and M. Westmoreland, Phys. Rev. A **56**, 131 (1997).
[6] A. S. Holevo, e-print quant-ph/9708046.
[7] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[8] A. S. Holevo, Theor. Probab. Appl. **23**, 411 (1978).
[9] P. Hausladen and W. K. Wootters, J. Mod. Opt. **41**, 2385 (1994).
[10] C. A. Fuchs and A. Peres, Phys. Rev. A **53**, 2038 (1996).
[11] M. Ban, K. Yamazaki, and O. Hirota, Phys. Rev. A **55**, 22 (1997).
[12] M. Osaki, M. Ban, and O. Hirota, J. Mod. Opt. **45**, 269 (1998).
[13] A. S. Holevo, J. Multivar. Anal. **3**, 337 (1973).
[14] M. Osaki and O. Hirota, in *Quantum Communication and Measurement*, edited by V. P. Belavkin, O. Hirota, and R. L. Hudson (Plenum Publishing, New York, 1995), pp. 401–406; M. Osaki, M. Ban, and O. Hirota, Phys. Rev. A **54**, 1691 (1996).
[15] C. W. Helstrom, IEEE Trans. Inf. Theory **IT-28**, 359 (1982).
[16] M. Ban, K. Kurokawa, R. Momose, and O. Hirota, Int. J. Theor. Phys. **36**, 1269 (1997).
[17] O. Hirota, Opt. Commun. **67**, 204 (1988).
[18] M. Sasaki and O. Hirota, Phys. Lett. A **210**, 21 (1996); **224**, 213 (1997); M. Sasaki, T. S. Usuda, O. Hirota, and A. S.

Holevo, Phys. Rev. A **53**, 1273 (1996); M. Sasaki and O. Hirota, *ibid.* **54**, 2728 (1996).
[19] A. Barenco, C. H. Bennett, R. Cleve, D. P. M. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Phys. Rev. A **52**, 3457 (1995).
[20] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Phys. Rev. Lett. **73**, 58 (1994).
[21] T. Sleator and H. Weinfurter, Phys. Rev. Lett. **74**, 4087 (1995).
[22] Q. A. Turchette, C. J. Hood, W. Lange, H. Mabuchi, and H. J. Kimble, Phys. Rev. Lett. **75**, 4710 (1995).
[23] The authors are indebted to a private communication from A. S. Holevo for the proofs of Theorems 2 and 3.
[24] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, Phys. Lett. A **236**, 1 (1997).