Vol 11 No 5, May 2002
1009-1963/2002/11(05)/0450-06
Chinese Physics
© 2002 Chin. Phys. Soc.
and IOP Publishing Ltd

Average mutual information of the random block-message ensembles for many-letters*

Hao San-Ru(郝三如)[a)b)†] and Hou Bo-Yu(侯伯宇)[a)]

[a)] *Institute of Modern Physics, Northwest University, Xi'an 710069, China*

[b)] *Department of Computer Teaching, Hunan Normal University, Changsha 410081, China*

(Received 15 September 2001; revised manuscript received 14 January 2002)

By making use of the theoretical framework presented by Bostroem (K. J. Bostroem, LANL quant-ph/0009052), we generalize the standard quantum information theory of block messages with fixed block length to the variable one. We show that the states belonging to a sufficiently large Hilbert space are the highly distinguishable states. We also consider the collection states (product states of more than one qubit state) and seek a "pretty good measurement"(PGM) with measurement vectors to improve the mutual information. The average mutual information over random block-message ensembles with variable block length n is discussed in detail.

Keywords: mutual information, random block message, many-letter ensemble
PACC: 0365, 0367, 8970

1. Introduction

The mutual information $I(X:Y)$[1−6] is one of the first information-theoretic quantities investigated with respect to quantum systems.[7−12] It quantifies how correlated two messages are. How much do we know about a message drawn from X^n when we have read a message drawn from Y^n? The maximum mutual information of an ensemble of quantum states obtainable between the states of the ensemble and the outcomes of a positive operator-valued measurement (POVM)[13−16] on these states is called accessible information.[17−19]

Information theory is the theory of messages composed from letters. Classical information theory can deal with the messages of not only the fixed-length blocks but also the variable-length blocks. It can be easily generalized to describe the messages of variable-length blocks. Unlike classical information theory, since quantum systems obey the laws of quantum mechanics and are usually based on the Hilbert space of fixed dimension, quantum information theory cannot directly use the results of classical information theory to deal with the messages of variable-length blocks. However, as we cannot directly deal with quantum block messages of variable lengths, we can construct an infinite direct sum of Hilbert spaces with different dimension number n. Since a quantum message block with length n is a quantum state in the Hilbert space with dimension number n, we have many quantum block messages with different length n in the infinite direct sum of Hilbert spaces. By making use of this infinite direct sum of Hilbert spaces, we can describe the quantum block messages with different block lengths. Bostroem[7] has presented a theoretical framework which allows the treatment of quantum messages with components of variable lengths.

The description of the direct sum of Hilbert spaces may appear rather unphysical to the reader, since everything in quantum mechanics is usually described in terms of direct products. But without explicitly using it, the direct sum is always present. For example, the fundamental space of quantum information theory, the two-dimensional complex number space $\mathcal{C}^2$ of a single qubit is in fact the direct sum of two $\mathcal{C}$. The reason is that the Cartesian product $\mathcal{C}^2 = \mathcal{C} \times \mathcal{C}$ can be embedded into the direct sum $\mathcal{C} \oplus \mathcal{C}$ by preserving the Hilbert space structure, since both components of $\mathcal{C}^2$ are mutually orthogonal. Hence the Cartesian product and the direct sum are just different representations of the same principle; adding separate levels of an observable, i.e. combining properties of a system by a quantum mechanical "OR". In many-letter theory the distinct degenerate levels of the length operator are added.

*Project supported in part by the National Natural Science Foundation of China (Grant No 19975036), and the Foundation of the Science and Technology Committee of Hunan Province, China (Grant No 21000205).
†Author to whom correspondence should be addressed. E-mail:srhao@sparc2.hunnu.edu.cn

As standard quantum information theory describes block messages of a fixed block length, we would go further now and allow quantum messages to have an arbitrary length. With the theoretical framework described by Bostroem,[7] in this paper, we discuss the mutual information about the many-letter random block ensemble with a variable length n of the block-message ensemble and study the average mutual information over the whole random block-message ensemble. In section 2 we review briefly some basic concepts of the many-letter quantum information theory in order to motivate the corresponding notions presented in section 3, which is dedicated to the varying-length many-letter quantum information theory. In section 3, we generalize the standard quantum information theory of the block message of a fixed length to the variable length. We show that the states belonging to a sufficiently large Hilbert space are the highly distinguishable states, and we discuss the average mutual information of the many-letter quantum system. A concluding remark is given in section 4.

2. Description of many-letter quantum information systems

In order to discuss the variable-length quantum block message, one should seek a quantum analogue of the set $\mathcal{A}^+$ of classical block messages of arbitrary lengths. This can be done by mapping each classical message $x^n \in \mathcal{A}^+$ to a product Hilbert vector $|x^n\rangle \in \mathcal{H}^n$. Regarding the set of the product block message of length n, we have

$$\mathcal{Q}^n \equiv \{|x^n\rangle \in \mathcal{H}^n \quad | \, x^n \in \mathcal{A}^n\}. \tag{1}$$

With $\mathcal{Q}^0 \equiv \{|\cdot\rangle\}$ being defined as the set formed by an empty message $|\cdot\rangle$, one can construct the Hilbert space that is spanned by members of $\mathcal{Q}^n$ as $\mathcal{H}_Q^n \equiv \mathrm{Span}(\mathcal{Q}^n) = \mathcal{H}_Q \otimes \cdots \otimes \mathcal{H}_Q$, with its dimension being given by $\dim\mathcal{H}_Q^n = (\dim\mathcal{H}_Q)^n \leq |\ \mathcal{Q}\ |^n$ and the subscript Q indicating that the Hilbert space $\mathcal{H}_Q^n$ is spanned by the elements of $\mathcal{Q}^n$. By making use of Eq.(1), one can construct the infinite set $\mathcal{Q}_+ \equiv \cup_{n=1}^{\infty}\mathcal{Q}^n$. The space spanned by the elements of $\mathcal{Q}_+$ is called the many-letter space: $\mathcal{M}_Q \equiv \oplus_{n=0}^{\infty}\mathcal{H}_Q^n$. Here the direct sum of two Hilbert spaces is defined as the orthogonal sum of their elements. Of course, the many-letter space can be restricted to a maximum number of letters, $\mathcal{M}_Q^N \equiv \oplus_{n=0}^{N}\mathcal{H}_Q^n$, due to a finite reservoir of available qubits. $\mathcal{M}_Q^N$ is a subspace of the total many-letter space $\mathcal{M}_Q$, i.e. $\mathcal{M}_Q^N \subset \mathcal{M}_Q$, and each many-letter message can be truncated to this subspace by an appropriate projector.

Alice can construct a random block-message ensemble $|\Phi\rangle$, from which she chooses variable block messages $|\varphi\rangle$ to communicate with Bob. This random block-message ensemble can be expressed as

$$|\Phi\rangle = \{[|\varphi\rangle, p(\varphi)] \quad | \quad |\varphi\rangle \in \Gamma\}, \tag{2}$$

where Γ is the set of block messages chosen with nonzero probability $p(\varphi)$:

$$\Gamma \equiv \{|\varphi\rangle \in \mathcal{H}_Q^n \quad | \quad p(\varphi) > 0, n = 0, 1, 2, \cdots\}. \tag{3}$$

The corresponding message matrix is $\sigma = \sum_{\varphi\in\Gamma} p\ (\varphi)|\varphi\rangle\langle\varphi|$.

In order to discuss block messages of different lengths, one should define a length operator for quantum block messages. Since the length of a quantum block message is an observable quantity, there is a self-adjoint length operator $\hat{L}$ acting on the many-letter space $\mathcal{M}_Q$ with a spectral decomposition of mutually orthogonal projectors Π_n on $\mathcal{M}_Q$, such that

$$\hat{L} = \sum_{n=0}^{\infty} n\Pi_n, \Pi_n\Pi_m = \delta_{nm}\Pi_n, \sum_{n=0}^{\infty}\Pi_n = \mathbf{1}. \tag{4}$$

Every message $|\varphi\rangle$ drawn from the ensemble has a well-defined length $L(\varphi)$ because it is in one of the eigenspaces of the length operator, i.e. $\hat{L}|\varphi\rangle = L(\varphi)|\varphi\rangle$. Thus the message matrix σ of random block messages can be block-diagonalized into a convex combination of block-message matrices σ_n, i.e. $\sigma = \sum_{n=0}^{\infty}\lambda_n\sigma_n$. Here we have $\sigma_n \in \mathcal{H}_Q^n$ and the probability λ_n given by $\lambda_n \equiv \sum_{L(\varphi)=n} p(\varphi)$, such that $\sum_{n=0}^{\infty}\lambda_n = 1$.

3. Average mutual information

In the general communication theory, the Shannon mutual information is defined in terms of the conditional probability $P(j|i)$ of obtaining output y_j, provided that the letter sent was x_i, which can be expressed as follows

$$I(X:Y) = \sum_i p_i \sum_j P(j|i)\log\left[\frac{P(j|i)}{\sum_k p_k P(j|k)}\right], \tag{5}$$

where log denotes the logarithm with base number 2, $\{x_i \in X\}$ are the input letters and $\{p_i\}$ are their prior probabilities; in addition, $\{y_i \in Y\}$ denote the output letters. In classical information theory, the channel matrix $[P(j|i)]$ depends only on the communication channel, with which the classical information communicates, and it is a constant for a specific communication channel and characterizes the noise in the channel. The mutual information is a convex function of the prior probability p_i of the input letters and depends only on p_i. In contrast, in a quantum information theoretic context where signal carriers are quantum states to be transmitted without noise, the channel matrix $[P(b|a)]$ (see Eqs.(7) and (16)) generally becomes a variable, which is usually called conditional probability. This is because the act of quantum detection itself generally has a probabilistic output, so the channel matrix is dependent on the choice of quantum detection strategy. As a result, the mutual information should depend on both the prior probability p_i and the choice of the quantum detection strategy.

A quantum detection strategy is described by a POVM on a Hilbert space $\mathcal{H}$. A POVM is any set $\hat{F}_a$ of Hermitian positive operators forming a resolution of the identity:

$$\hat{F}_a^+ = \hat{F}_a, \quad \hat{F}_a \geq 0 \quad \forall a, \quad \sum_a \hat{F}_a = I. \tag{6}$$

The detection operator $\hat{F}_a$ corresponds to the output letter y_j and the conditional probability is given by

$$P(b|a) = \mathrm{Tr}(\hat{F}_a \hat{\rho}_a). \tag{7}$$

Thus, in the quantum context the value of the mutual information $I(A{:}B)$ is carried out with respect to the choice of POVM $\{\hat{F}_a\}$ for a fixed ensemble $\varepsilon = \{\hat{\rho_a}, p_a\}$, and the mutual information $I(A{:}B)$ is a convex function on the set of all POVMs[20] provided that the input ensemble $\varepsilon = \{\hat{\rho_a}, p_a\}$ is fixed.

There is a maximum value for the mutual information at a group of POVM. In general, it is no longer obvious what Bob's optimal measurement will be. But Bob can invoke a general procedure that, while not guaranteed to be optimal, is usually at least pretty good. The POVM constructed by this procedure is generally called a "pretty good measurement" (PGM).[21]

For this procedure, we want the overlaps between the different signal states (code words) to be sufficiently small provided that the code words are very long. The smaller the overlaps, the better the detection strategy, and hence the higher the mutual information. We see that the average overlap of two typical n-letter code words becomes zero[6] if $n \longrightarrow \infty$.

For a typical code, each typical code word has a large overlap with a typical subspace $\wedge^{(n)}$ that has a dimension $\dim\wedge^{(n)} > 2^{n(S(\rho)-\delta)}$, where $S(\rho)$ is the von Neumann entropy of a random ensemble $\varepsilon = \{|\varphi_x\rangle, \rho_x\}$. Furthermore, for a typical code, the marginal ensemble governing each letter is close to ε.

Because the typical subspace is very large for a large n, Alice can choose many code words, yet is assured that the typical overlap of two typical code words is very small. Heuristically, the typical code words are randomly distributed in a typical subspace and, on average, two random unit vectors in a space of dimension D have overlap $1/D$. Therefore, if $|u\rangle$ and $|v\rangle$ are two code words, we have

$$< |\langle u|v\rangle|^2 >_\wedge \; < \; 2^{-n(s(\rho)-\delta)}. \tag{8}$$

Here $< \cdot >_\wedge$ denotes an average over random typical code words. One can convince oneself that the typical code words really are uniformly distributed in the typical subspace as follows. Averaged over the ensemble, the overlap of random code words $|\varphi\rangle_n$ and $|\psi\rangle_n$ is

$$\begin{aligned} |_n\langle\varphi|\psi\rangle_n|^2 &= \sum_{L(\varphi)=n, L(\psi)=n} p(\varphi)p(\psi)|\langle\varphi|\psi\rangle|^2 \\ &= \mathrm{Tr}(\rho^{(n)})^2, \end{aligned} \tag{9}$$

where the subscript n indicates that $|\varphi\rangle$ has n qubits (or a quantum state in a Hilbert space of 2^n dimensions). $\rho^{(n)} = \rho \otimes \rho \otimes \cdots \otimes \rho$, and $\rho = \sum_x p_x |\varphi_x\rangle\langle\varphi_x|$ is the density matrix of ensemble $\varepsilon = \{|\varphi_x\rangle, p_x\}$. $L(\varphi)$ (or $L(\psi)$) is the length of the message blocks presented by quantum states $|\varphi\rangle$ (or $|\psi\rangle$).

We restrict the trace in Eq.(9) to the typical subspace $\wedge^{(n)}$, which has $\dim\wedge^{(n)} < 2^{n(S+\delta)}$, and the eigenvalue τ of $\rho^{(n)}$ to $\wedge^{(n)}$ satisfies $\tau < 2^{-n(S-\delta)}$. Thus we obtain

$$\begin{aligned} < |_n\langle\varphi|\psi\rangle_n|^2 >_\wedge &= \mathrm{Tr}_\wedge[\rho^{(n)}]^2 < 2^{n(S+\delta)}[2^{-n(S-\delta)}]^2 \\ &= 2^{-n(S-3\delta)}, \end{aligned} \tag{10}$$

where $\mathrm{Tr}_\wedge$ denotes the trace in the typical subspace.

Now we further suppose that $2^{n(S-\delta)}$ random code words $\{|\varphi_i\rangle\}$ are selected. If $|\varphi_j\rangle$ is any fixed code word,

$$\begin{aligned} &\sum_{i\neq j} < |\langle\varphi_i|\varphi_j\rangle|^2 > \; < \; 2^{n(S-\delta)} \cdot 2^{-n(S-\delta')} + \epsilon \\ &= 2^{-n(\delta-\delta')} + \epsilon, \end{aligned} \tag{11}$$

where the summation is taken over all code words, and the average is no longer restricted to the typical code words, and the ϵ on the right-hand side arises from the typical case. Now for any fixed δ, we can choose δ' and ϵ as small as we please for n sufficiently large. Then we conclude that when we average over both code and code words with a code, the code words become highly distinguishable as $n \longrightarrow \infty$, i.e. the code words are almost orthogonal to each other for the length n of code words sufficiently large.

For the random block-message ensemble (2), the n-letter states drawn from random block-message ensemble $|\Phi\rangle$ are expressed as

$$|\Phi_n\rangle = \{[|\varphi\rangle_n, p_n(|\varphi\rangle_n)] \mid |\varphi\rangle_n \in \mathcal{H}_Q^n, p_n(|\varphi\rangle_n) > 0, L(|\varphi\rangle_n) = n\}, \tag{12}$$

where the $|\varphi\rangle_n$ states are the eigenstates of the length operator $\hat{L}$ corresponding to the eigenvalue n, i.e. $\hat{L}|\varphi\rangle_n = L(|\varphi\rangle_n)|\varphi\rangle_n$ with $L(|\varphi\rangle_n) = n$. As is demonstrated above, those states chosen from Eq.(12) are highly distinguishable provided that n is sufficiently large.

Now if Alice send the signal states chosen from Eq.(12), which are not assumed to be orthogonal or normalized, then Bob can exploit the aforementioned high distinguishability in his choice of measurement, i.e. PGM.[21] We first construct an auxiliary matrix G

$$G = \sum_{a=1} [|\varphi_a\rangle_{nn}\langle\varphi_a|]_{L(|\varphi\rangle_n)=n}, \tag{13}$$

where $|\varphi_a\rangle_n$ is the ath block message with length n. It is easy to check that G is a positive operator on the subspace $\mathcal{H}_Q^n$ of the total random block space Γ. Therefore, on this subspace $\mathcal{H}_Q^n$, G has an inverse, G^{-1}, and the inverse has a positive square root $G^{-1/2}$.

By defining the POVM operator

$$F_a = \left[G^{-1/2}(|\varphi_a\rangle_{nn}\langle\varphi_a|)G^{-1/2}\right]_{L(|\varphi\rangle_n)=n}, \tag{14}$$

one can easily find that

$$\begin{aligned} &\sum_{a=1} [F_a]_{L(|\varphi\rangle_n)=n} \\ &= \left[G^{-1/2}\left(\sum_a |\varphi_a\rangle_{nn}\langle\varphi_a|\right)G^{-1/2}\right]_{L(|\varphi\rangle_n)=n} \\ &= \left[G^{-1/2}GG^{-1/2}\right]_{L(|\varphi\rangle_n)=n} \\ &= 1, \end{aligned} \tag{15}$$

on the space spanned by $|\varphi_a\rangle_n$s. Performing a PGM, Bob will obtain, if Alice chooses preparation $|\varphi_a\rangle_n$, the measurement outcome b with conditional probability, which comes from Eq.(7),

$$p(b|a) =_n \langle\varphi_a|F_b|\varphi_a\rangle_n = \mathrm{Tr}\left(F_b|\varphi_a\rangle_{nn}\langle\varphi_a|\right). \tag{16}$$

These conditional probabilities, together with the ensemble (12), determine the amount of mutual information that Bob can gain from the n-letter message.

For the random block-message systems (2), the n-letter probability $p_n(|\varphi\rangle_n)$ in Eq.(12) becomes

$$p_n(|\varphi\rangle_n) = \left[\frac{p(\varphi)}{\lambda_n}\right]_{L(\varphi)=n}. \tag{17}$$

This depends only on the n-letter states $|\varphi\rangle$ which satisfy $L(\varphi) = n$, and the symbol λ_n in Eq.(17) is the total probability of the random block-message quantum state $|\varphi\rangle$ with the block length equal to n. Substituting Eqs.(16) and (17) into Eq.(5), one can obtain the mutual information for the random n-letter block message, which is expressed as follows

$$I_n(A:B) = \left[\sum_a [p(\varphi_a)/\lambda_n] \sum_b \langle\varphi_a|F_b|\varphi_a\rangle \times \log\left[\frac{\langle\varphi_a|F_b|\varphi_a\rangle}{\sum_c (p(\varphi_c)/\lambda_n)\langle\varphi_c|F_b|\varphi_c\rangle}\right]\right]_{L(\varphi)=n}. \tag{18}$$

For a mixed state, the n-letter ensemble (12) can be rewritten as

$$|\Phi_n\rangle = \{\rho_{n,a}, p_{n,a}\}, \tag{19}$$

and the density operator of this ensemble reads

$$\rho_n = \sum_a p_{n,a}\rho_{n,a}. \tag{20}$$

The conditional probability $p(b|a)$ (Eq.(16)) becomes

$$p(b|a) = \mathrm{Tr}(F_b\rho_{n,a}), \tag{21}$$

and the *a priori* probability $p_{n,a}$ in Eq.(19) satisfies the following relation

$$\sum_a p_{n,a} = \sum_{L(\varphi_n)=n} p_n(\varphi_n) = \lambda_n. \tag{22}$$

Now the mutual information for the random mixed n-letter-state block ensemble can be written as

$$I_n(A:B) = \sum_a p_{n,a} \sum_b \mathrm{Tr}(F_b\rho_{n,a}) \times \log\left[\frac{\mathrm{Tr}(F_b\rho_{n,a})}{\sum_c p_{n,c}\mathrm{Tr}(F_b\rho_{n,c})}\right]. \tag{23}$$

As demonstrated above, the overlap of the two code words approaches zero as $n \longrightarrow \infty$. Alice can choose highly distinguishable code words which almost become mutual-orthogonal code words as n is sufficiently large, and Bob can perform a PGM at a finite n, which approaches an optimal orthogonal measurement as $n \longrightarrow \infty$; therefore, we have the results of Eqs.(18) and (23). We assume that Alice can also choose highly distinguishable code words at a smaller n, and Bob can still perform a PGM for those n-states so that Eqs.(18) and (23) are still valid. We now consider the total random block ensemble, and calculate the average mutual information of the total random block ensemble.

Considering the total random block-message ensemble of Eqs.(2) and (3), whose corresponding message density operator is represented by σ, we can write the mutual information of the total random block-message ensemble as

$$\begin{aligned} & I_{\rm pure}(A:B) \\ =& \sum_{n=0}^{\infty}\left(\left[\sum_a p(\varphi_a)\sum_b \langle\varphi_a|F_b|\varphi_a\rangle \right.\right. \\ & \left.\left.\times \log\left[\frac{\langle\varphi_a|F_b|\varphi_a\rangle}{\sum_c (p(\varphi_c)/\lambda_n)\langle\varphi_c|F_b|\varphi_c\rangle}\right]\right]_{L(\varphi)=n}\right), \quad (24) \end{aligned}$$

$$\begin{aligned} I_{\rm mixed}(A:B) =& \sum_{n=0}^{n}\lambda_n \sum_{a,b} p_{n,a}{\rm Tr}(F_b\rho_{n,a}) \\ & \times \log\left[\frac{{\rm Tr}(F_b\rho_{n,a})}{\sum_c p_{n,c}{\rm Tr}(F_b\rho_{m,c})}\right], \quad (25) \end{aligned}$$

where $I_{\rm pure}(A:B)$ is the average mutual information for the pure-state random block-message ensemble and $I_{\rm mixed}(A:B)$ is that for the mixed-state random block-message ensemble.

The length $L(\varphi)$ of the message block $|\varphi\rangle$ is $\sum_{n=0}^{\infty} n\langle\varphi|\Pi_n|\varphi\rangle$, whereas the length $L(\sigma)$ of the message ensemble σ is given by

$$\begin{aligned} L(\sigma) =& <\langle\Phi|\hat{L}|\Phi\rangle> = \sum_{\varphi\in\Sigma} p(\varphi)L(\varphi) \\ =& \sum_{\varphi\in\Gamma}\sum_{n=0}^{\infty} np(\varphi)\langle\varphi|\Pi_n|\varphi\rangle = {\rm Tr}\{\sigma\hat{L}\}. \quad (26) \end{aligned}$$

Every block-message matrix σ_n has a definite length n, i.e. $\hat{L}\sigma_n = n\sigma_n$, hence it commutes with the length operator. Then the average length $L(\sigma)$ of the ensemble σ is $\sum_{n=0}^{\infty}\lambda_n n$. Therefore, the average mutual information for the random block-message ensemble per letter can be expressed as

$$\begin{aligned} I_1(A:B) =& I_{\rm pure,mixed}(A:B)/L(\sigma) \\ =& I_{\rm pure,mixed}(A:B)/\sum_{n=0}^{\infty} n\lambda_n. \quad (27) \end{aligned}$$

If we choose basis sets B_n of mutual orthogonal block messages $|e_{i_n}^n\rangle$ of the block-message matrix σ_n with length n, the block-message matrix becomes diagonal:

$$\sigma_n = \sum_{i_n=1}^{K^n}\left(\sum_{L(\varphi)=n}|\varphi_{i_n}^n|^2\right)|e_{i_n}^n\rangle\langle e_{i_n}^n|, \quad (28)$$

with the wave components $\varphi_{i_n}^n \equiv \langle e_{i_n}^n|\varphi\rangle$ and K^n is the dimensional number of the space $\mathcal{H}_Q^n$.

It is worthwhile noting that the block messages $|e_{i_n}^n\rangle$ are generally not product messages. It is clear that Eq.(28) looks like an ensemble $\varepsilon_n = \{|e_{i_n}^n\rangle,\ p(e_{i_n}^n)\}$ with *a priori* probability $p(e_{i_n}^n) = \sum_{L(\varphi)=n}|\varphi_{i_n}^n|^2$. To Bob, this ensemble $\varepsilon_n = \{|e_{i_n}^n\rangle, p(e_{i_n}^n)\}$ appears as if Alice would send him states $|e_{i_n}^n\rangle$ of well-defined length n with probability $p(e_{i_n}^n)$. If Bob takes a von Neumann measurement, then he obtains

$$S(\sigma_n) = -\sum_{i_n=1}^{K^n} p(e_{i_n}^n)\log p(e_{i_n}^n). \quad (29)$$

Because $|e_{i_n}^n\rangle$ are basis vectors with length $L(|e_{i_n}^n\rangle)$ equal to n, they are orthogonal normalized states. Inserting $p(e_{i_n}^n) = \sum_{L(\varphi)=n}|\varphi_{i_n}^n|^2$ into Eq.(29), we can obtain the mutual information of the n-letter ensemble $\varepsilon_n = \{|e_{i_n}^n\rangle, p(e_{i_n}^n)\}$:

$$\begin{aligned} I_n =& S(\sigma_n) \\ =& -\sum_{i_n=1}^{K^n}\sum_{L(\varphi)=n}|\varphi_{i_n}^n|^2\log\left(\sum_{L(\varphi)=n}|\varphi_{i_n}^n|^2\right). \quad (30) \end{aligned}$$

Using the random block-message ensemble (2) and the density matrix σ, and considering the orthogonal relationships between σ_n that have different length n, one can obtain the average mutual information of the random block-message ensemble under the von Neumann

measurement as follows

$$\begin{aligned} I_\sigma &= -\,\mathrm{Tr}\sigma\log\sigma \\ &= -\sum_{n=0}^{\infty}\mathrm{Tr}(\lambda_n\sigma_n)\log(\lambda_n\sigma_n) \\ &= -\sum_{n=0}^{\infty}\lambda_n\log\lambda_n-\sum_{n=0}^{\infty}\lambda_n\mathrm{Tr}\sigma_n\log\sigma_n \\ &= -\sum_{n=0}^{\infty}\lambda_n\log\lambda_n+\sum_{n=0}^{\infty}\lambda_n S(\sigma_n) \\ &= -\sum_{n=0}^{\infty}\lambda_n\log\lambda_n+\sum_{n=0}^{\infty}\lambda_n I_n. \end{aligned} \tag{31}$$

Hence, the mutual information per letter reads

$$I_1(\sigma)=I(\sigma)/L(\sigma)=I(\sigma)/\sum_{n=0}^{\infty}n\lambda_n. \tag{32}$$

The expressions (27) and (32) are the average mutual information per letter for the random block-message ensemble; they are, most likely, higher mutual information that can be achieved with any measurement over the many-letter random block messages.

4. Concluding remarks

By making use of the theoretical framework,[7] and allowing the treatment of messages with components of variable length n, we have constructed a many-letter space and generalized the standard quantum information theory of block messages of a fixed length to the varying length case. The PGM with highly distinguishable measured states can improve the mutual information. Firstly, Alice is able to convey more information to Bob by "increasing" the dimensional number of the Hilbert space of her code word. It is better for Alice to choose messages from the highly distinguishable signal states than from the less distinguishable ones. Secondly, Bob is able to read more information if he performs a collective measurement on many qubits instead of one at a time. The PGM[21] described in this paper can give better information gained from any known measurement.

In practice, Alice may convey varying-length block messages, and Bob has to read the average mutual information over the variable length random block messages. As we have demonstrated that the PGM can achieve a better mutual information for the block messages of a fixed length n, so the average mutual information is, most likely, the highest that can be achieved with any measurement over the many-letter random block messages.

The average mutual information of the random block messages for the many-letter systems is a very interesting problem. Moreover, the quantum channel of these ensembles is also interesting as well as useful, and it is worthwhile discussing in future works.

Acknowledgments

We are grateful to P Wang and K J Shi for helpful comments and the first author (S R Hao) wishes to thank L Y Wang for his encouragement and support.

References

[1] Holevo A S and Multivar J 1973 *Analysis* **3** 337
[2] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic) chap. 1
[3] Peres A 1993 *Quantum Theory: Concepts and Methods* (Dortrecht: Kluwer Academic) pp 279–289
[4] Peres A and Wootters W K 1991 *Phys. Rev. Lett.* **66** 1119
[5] Kato K, Osaki M, Sasaki M and Hirota O 1998 *IEEE Trans. Commun.* **47** 248
[6] Fuchs C A 1996 Distinguishability and accessible information in quantum theory *Preprint* quant-ph/9601020
[7] Bostroem K J 2000 Concepts of a quantum information theory of many letters *Preprint* quant-ph/0009052
[8] Hao S R and Wang L Y 2000 *Acta Phys. Sin.* **49** 610 (in Chinese)
[9] Liu D Y 1998 *Acta Phys. Sin.* **47** 1233 (in Chinese)
[10] Duan L M 1997 *Acta Phys. Sin.* (*Overseas Edition*) **6** 811
[11] Fang M F, Zhou Q P and Zhou P 1999 *Acta Phys. Sin.* (*Overseas Edition*) **8** 401
[12] Li L X and Guo G C 1998 *Acta Phys. Sin.* (*Overseas Edition*) **7** 727
[13] Eldar Y C and Forney G D 2000 On quantum detection and the square-root measurement *Preprint* quant-ph/0005132
[14] Peres A 1993 *Quantum Theory: Concepts and Methods* (Dortrecht: Kluwer Academic) chap. 5
[15] Yuen H P, Kennedy R S and Lax M 1975 *IEEE Trans. Inform. Theor.* **21** 125
[16] Ban M, Kurukowa K, Momose R and Hirota O 1997 *Int. J. Theor. Phys.* **36** 1269
[17] Levitin L B 1995 Optimal quantum measurements for two pure and mixed states *Quantum Communications and Measurement* ed V P Belavkin, O Hirota and R L Hudson (New York: Plenum) pp 439–448
[18] Sasaki M, Barnett S M, Jozsa R, Osaki M and Hirota O 1999 *Phys. Rev.* A **59** 3325 (*Preprint* quant-ph/9812062)
[19] Shor P W 2000 On the number of elements needed in a POVM attaining the accessible information *Preprint* quant-ph/0009077
[20] Cover T M and Thomas J A 1991 *Elements of Information Theory* (New York: Wiley) chap. 4
[21] Hausladen P and Wootters W K 1994 *J. Mod. Opt.* **41** 2385