Limits for compression of quantum information carried by ensembles of mixed states

Michał Horodecki[*]
Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80-952 Gdańsk, Poland

(Received 12 December 1997)

We consider the problem of compression of the quantum information carried by ensemble of mixed states. We prove that for arbitrary coding schemes the least number of quantum bits (qubits) per message needed to convey the signal states asymptotically faithfully is bounded from below by the Holevo function $S(\varrho) - \Sigma_i p_i S(\varrho_i)$. We also show that a compression protocol can be composed with another one, provided that the latter offers *perfect* transmission. Such a compound protocol is applied to the case of binary source. It is conjectured to reach the obtained bound. Finally, we point out that in the case of mixed signal states there could be a difference between the maximal compression rates at the coding schemes that are ‘‘blind’’ to the signal and the ones that assume the knowledge about the identities of the signal states. [S1050-2947(98)07105-4]

PACS number(s): 03.67.Hk

I. COMPRESSION OF QUANTUM INFORMATION

One of the important problems of information theory is the compression of information. A general limit for compression rate of classical information is placed by the so-called noiseless coding theorem [1]. Suppose that a source generates a message i with probability p_i and allow the subsequent messages to cumulate into long sequences and then represent them (encode) as sequences of bits as economically as it is possible (economically means here that we want to use the least possible average number of bits per message). The task of the receiver (Bob) is to convert (decode) the binary sequences into the original sequences of messages. Here we do not require perfect transmission but only an asymptotically faithful one. This means that Bob may be unable to recover correctly each sequence, but the probability of error tends to zero if the length of input blocks tends to infinity.

Now the noiseless coding theorem [1] says that the necessary and sufficient number of bits per message needed for asymptotically faithful transmission is equal to the Shannon entropy $S = -\Sigma_i p_i \log p_i$ (in this paper we use base-2 logarithms) of the probability distribution characterizing the source. Then this quantity says in fact how much information per message is actually produced by the source. Indeed, one can imagine, that after the most economical compression procedure, each piece of the compressed signal is equally essential as all redundancy was removed. Then the size of the maximally compressed signal can be interpreted as the quantity of information contained in the input (uncompressed) one.

Let us now turn to the problem of compression of quantum information, which was first considered by Schumacher [2]. The role of messages is here played by quantum states ϱ_i (signal states) and the bits are replaced by qubits, i.e., two-level systems. The probability of error is generalized to quantum case by means of chosen measures of fidelity or distortion [2–4] between two quantum states. Thus we will ask about the least number of two-level systems needed to carry the information asymptotically faithfully to Bob, i.e., so that the average distortion between the input and output states will tend to zero (or the average fidelity will tend to one) in the limit of input signal block of infinite length.

Before we review the results obtained so far, let us mention the fundamental difference between the quantum and classical case due to the no-cloning theorem for quantum states [5]. It was shown that the theorem is equivalent to the impossibility of measuring the state parameters of a single quantum system [6]. Then we can imagine two scenarios that, according to the above restriction for quantum information processing, could in principle produce different results [7]. Within the first scenario, we assume that Alice does not know the identities of the particular states produced by the source. Then, in accordance with the no-cloning theorem, Alice has no means to get this knowledge. Thus the most general of Alice’s coding protocol amounts to performing a quantum operation (trace preserving completely positive map—see the Appendix) [8] that depends only on the known characteristics of the source, i.e., the form of the generated ensemble $\{p_i, \varrho_i\}$. We will call it *blind* coding. If, however, we allow Alice to know each of the produced states, we deal with the second scenario (*arbitrary* or *nonblind* coding), where Alice’s coding amounts to replacing the sequences of signal states by completely arbitrary new states. It seems that in some cases it will produce more efficient compression than is possible within the previous scenario.

Let us now review the results obtained so far in the domain of compression of quantum information. For the ensemble of pure states, Schumacher showed [2] that, by means of blind coding, it is possible to reduce the needed number of qubits to the value of the von Neumann entropy of the total density matrix of ensemble $\varrho = \Sigma_i p_i \varrho_i$ (in short, the von Neumann entropy of the ensemble). The proposed coding-decoding protocol was then simplified by Jozsa and Schumacher [9] (we will refer to it as the SJ protocol). To obtain the converse statement saying that this quantity is also necessary for faithful recovery of the signal, Barnum *et al.* [7] considered an arbitrary coding scheme. It turned out that even in this case it is impossible to better compress the data, so that the obtained lower bound applied also for the first

[*] Electronic address: michalh@iftia.univ.gda.pl

scenario. Thus for the ensemble of pure states the problem of compression has been completely solved: the two scenarios give the same degree of compression, and the information per message contained in such an ensemble is equal to the von Neumann entropy $S(\varrho)$ of the ensemble. Note that this establishes a precise sense of the von Neumann entropy within the quantum information theory [10].

Now, the problem of ensemble of *mixed* states is still open. The SJ coding protocol allows one to compress such an ensemble down to the value of its von Neumann entropy [4,11] (see also [12] in this context), but one knows that in some cases the more efficient protocols are possible [4,11]. To illustrate it, let us consider the source producing with certainty some established mixed state. Then the ensemble has entropy greater than zero but of course it does not carry any information. This implies that the ‘‘information content’’ of the ensemble (for any of the two scenarios) cannot be, in general, merely a function of the density matrix of the ensemble. Instead, it must depend on the particular form of the ensemble. Moreover, it seems that for ensembles of mixed states the arbitrary coding could produce more efficient compression than the blind one. Under this consideration it is desirable to investigate the problem of compression of information carried by ensembles of mixed states. In particular, an important task is to provide some limits for the compression rates for the two types of coding.

In this paper we provide the lower bound for the necessary number of quantum bits (qubits) per message needed for faithful transmission of the quantum information carried by an ensemble of mixed states for arbitrary coding. The bound is equal to the function $S(\varrho)-\Sigma_i p_i S(\varrho_i)$ (we will call it Holevo information), which was shown by Holevo to be an upper bound for accessible information [13,14]. In particular it implies that for the ensembles of states of disjoint supports the two considered types of coding produce the same result. Further we investigate the problem of composing the compression protocols. We consider a class of nonblind coding protocols, which involve composition of two protocols: an ideal one, which amounts to replacing the input states by new states, which, partially traced, reproduce the former ones, and the SJ protocol (applied to mixed states). Finally, we conjecture that if the arbitrary coding schemes are allowed then the Holevo information is in fact equal to the minimal number of qubits needed for faithful transmission and the bound can be reached by means of the proposed class of protocols.

II. COMPRESSION PROTOCOLS

Suppose that Alice generates a signal state ϱ_i^0 (acting on a Hilbert space $\mathcal{H}_Q$) with probability p_i^0. The produced ensemble $\mathcal{E}_0=\{p_i^0,\varrho_i^0\}$ has the density matrix $\varrho^0=\Sigma_i p_i^0 \varrho_i^0$. Denote now the product $\varrho_{i_1}^0\otimes\cdots\otimes\varrho_{i_N}^0$ by ϱ_i, where i now stands for multi-index (to avoid complicated notation we do not write the index N explicitly unless necessary). The corresponding ensemble and state are denoted by $\mathcal{E}$ and ϱ, respectively. Now Alice performs a coding operation over the initial ensemble $\mathcal{E}$ ascribing to any input state ϱ_i a new state $\widetilde{\varrho}_i$. The map $\varrho_i\rightarrow\widetilde{\varrho}_i=\Lambda_A(\varrho_i)$ is supposed to be a quantum operation for blind coding or an arbitrary map for nonblind coding. In the latter case we allow Alice to know even which states are generated by the source, so that she can prepare separately each of the states $\widetilde{\varrho}_i$ for each i.

The new states $\widetilde{\varrho}_i$ represent the compressed signal, which is then flipped into the suitable number of qubits determined by the dimension of subspace occupied by the state $\widetilde{\varrho}$ of the ensemble and sent through the noiseless channel to Bob. Now the states $\widetilde{\varrho}_i$ are to be decoded to become close to the initial states ϱ_i. For this purpose Bob performs some established quantum operation Λ_B, which of course does not depend on i. Then the resulting states are $\varrho_i'=\Lambda_B(\widetilde{\varrho}_i)$ and the whole scheme is the following:

$$\varrho_i \xrightarrow[\Lambda_A]{\text{Alice's coding}} \widetilde{\varrho}_i \xrightarrow[I]{\text{noiseless channel}} \widetilde{\varrho}_i \xrightarrow[\Lambda_B]{\text{Bob's decoding}} \varrho_i', \tag{1}$$

where ϱ_i and ϱ_i' act on the Hilbert space $\otimes^N\mathcal{H}_Q$ while $\widetilde{\varrho}_i$ on the channel Hilbert space $\mathcal{H}_C$. Without loss of generality, we can assume (as in Ref. [7]) that $\mathcal{H}_C\subseteq\otimes^N\mathcal{H}_Q$. As a measure of distortion characterizing the quality of the transmission $\varrho_i\rightarrow\varrho_i'$ we choose the metric induced by the trace norm. The latter is defined as

$$\|A\|=\mathrm{Tr}|A| \tag{2}$$

with $|A|=\sqrt{A^\dagger A}$. Thus the trace norm of the Hermitian operator is simply the sum of absolute values of its eigenvalues. Consequently, the distortion is defined as

$$D(\varrho,\sigma)=\|\varrho-\sigma\|. \tag{3}$$

An important property of the proposed measure of distortion is the fact that it does not increase under the quantum operations (see Appendix). Then the average distortion $\overline{D}\equiv\Sigma_i p_i D(\varrho_i,\varrho_i')$ will indicate the quality of the process of recovery of quantum information by Bob after compression by Alice. Now, for a fixed source, determined by the ensemble $\mathcal{E}_0$, one considers the sequence of coding-decoding pairs (Λ_A,Λ_B) with the property that $\lim_{N\rightarrow\infty}\overline{D}=0$ (recall that the pair is implicitly indexed by N). Such sequences will be called protocols.

Define now the quantity R_P characterizing the asymptotic degree of compression of the initial quantum data at a given protocol P by

$$R_P=\lim_{N\rightarrow\infty}\frac{1}{N}\log\dim\widetilde{\varrho}. \tag{4}$$

Here $\dim\widetilde{\varrho}$ denotes the dimension of the support of the state $\widetilde{\varrho}$ given by the number of nonzero eigenvalues. The quantity $\log\dim\widetilde{\varrho}$ has the interpretation of the number of qubits needed to carry the state $\widetilde{\varrho}$ undisturbed ($\widetilde{\varrho}$ is to be transferred by a noiseless channel).

Now, given a class $\mathcal{P}$ of protocols, we define the quantity

$$I_{\mathcal{P}}=\inf_{P\in\mathcal{P}} R_P, \tag{5}$$

which is equal to the least number of qubits per message needed for asymptotically faithful transmission of the initial signal states from Alice to Bob within the considered class of protocols (to be strict one needs $I_{\mathcal{P}}+\delta$ qubits per message, where δ can be chosen arbitrarily small). As discussed in Sec. I, we are interested in two classes of protocols—the ones with blind and arbitrary coding schemes. Accordingly we will consider two kinds of information—the *passive information* $I_p=I_{\mathcal{P}}$ where $\mathcal{P}$ is the class of protocols with blind coding and the *effective information* I_e with the infimum taken over protocols with arbitrary coding. The effective information represents the amount of information that seems to be *actually* carried out by the ensemble while the passive information I_p represents the information that is ''seen'' by the quantum apparatus, which is ''blind'' to the signal. Although the actual information contents of the ensemble could be in fact lower, the apparatus cannot benefit it, as it cannot in general read the identities of the signal states without disturbance. As a result the compression rate is restricted by the value of passive information. Finally, it is convenient to introduce the *information defect* $I_d=I_p-I_e$. This quantity tells us how the ensemble is ''unkind'' to us: while carrying little information the ensemble requires to be processed as if it contained a large amount of information. Let us recall here that for an ensemble of pure states the impossibility of reading the input states does not decrease the compression efficiency and the defect is equal to zero in spite of nonorthogonality of the signal states [7].

III. THE BOUND FOR EFFECTIVE INFORMATION

In this section we will prove the main result of this paper.

Theorem. The Holevo information $I(\mathcal{E}_0)=S(\varrho^0)-\Sigma_i p_i^0 S(\varrho_i^0)$ of the ensemble is the lower bound for its effective information:

$$I_e(\mathcal{E}_0)\geqslant I(\mathcal{E}_0). \tag{6}$$

Note that, since by definition we have $I_e\leqslant I_p$ the theorem provides automatically the lower bound for passive information I_p. Note also that for ensembles of pure states the Holevo information is simply equal to the entropy of the ensemble so that the theorem is compatible with the result of Ref. [7] (up to the measure of the quality of transmission).

To prove the theorem we need the lemma saying that if the average distortion between the two ensembles is small, then the difference between their Holevo informations per message is also small.

Lemma. Let $\Sigma_i p_i\|\varrho_i-\varrho_i'\|=\epsilon\leqslant\frac{1}{2}$. Then the following inequality is valid

$$|I(\mathcal{E})-I(\mathcal{E}')|\leqslant 2[\epsilon N\log d+\eta(\epsilon)], \tag{7}$$

where $\eta(x)=-x\log x$ with $\eta(0)=0$, $d=\dim\mathcal{H}_Q$.

Proof. We will use the following estimate [16]

$$|S(\varrho)-S(\sigma)|\leqslant\|\varrho-\sigma\|\log\dim\mathcal{H}+\eta(\|\varrho-\sigma\|) \tag{8}$$

which is valid for states ϱ, σ acting on the Hilbert space $\mathcal{H}$, with $\|\varrho-\sigma\|\leqslant\frac{1}{2}$. Based on the above inequality, we obtain

$$\begin{aligned}|S(\varrho)-S(\varrho')|&\leqslant N\log d\|\varrho-\varrho'\|+\eta(\|\varrho-\varrho'\|)\\ &\leqslant N\log d\sum_i p_i\|\varrho_i-\varrho_i'\|\\ &\quad+\eta\left(\sum_i p_i\|\varrho_i-\varrho_i{\prime}\|\right)\\ &=\epsilon N\log d+\eta(\epsilon),\end{aligned} \tag{9}$$

where we used the fact that the trace norm is convex and that the function η is increasing on the interval $(0,\frac{1}{2})$.

We also have

$$\begin{aligned}&\sum_i p_i|S(\varrho_i)-S(\varrho_i')|\\ &\quad\leqslant\sum_i p_i[N\log d\|\varrho_i-\varrho_i'\|+\eta(\|\varrho_i-\varrho_i'\|)]\\ &\quad\leqslant\epsilon N\log d+\eta(\epsilon),\end{aligned} \tag{10}$$

where the concavity of the function η was used. Now adding the two above inequalities we obtain the desired result.

Now we can start to prove the theorem. For this purpose let us estimate the quantity $\log\dim\widetilde{\varrho}$. First, it is bounded from below by $I(\widetilde{\mathcal{E}})$. This follows from the obvious fact that the von Neumann entropy of a state is less than or equal to the logarithm of the dimension of the Hilbert space the state acts on. Now let us note [14,15] that the function $I(\mathcal{E})$ can be written as the mean relative entropy between the components ϱ_i of the ensemble and the density matrix ϱ of the latter,

$$I(\mathcal{E})=\sum_i p_i S(\varrho_i|\varrho), \tag{11}$$

where the relative entropy [17] is given by

$$S(\varrho|\sigma)=\mathrm{Tr}(\varrho\log\varrho-\varrho\log\sigma). \tag{12}$$

Then we can benefit from the Uhlmann monotonicity theorem [18], which states that the relative entropy does not increase under the action of completely positive trace preserving map (quantum operation). Thus we obtain the inequality

$$I(\widetilde{\mathcal{E}})\geqslant I(\mathcal{E}') \tag{13}$$

as the ensemble $\mathcal{E}'$ is produced by Bob's quantum operation from the ensemble $\widetilde{\mathcal{E}}$. Using the inequality (13) and applying the lemma we get

$$\log\dim\widetilde{\varrho}\geqslant I(\mathcal{E})-2[\overline{D}N\log\dim\mathcal{H}_Q+\eta(\overline{D})]. \tag{14}$$

Noting that $I(\mathcal{E})=NI(\mathcal{E}_0)$, dividing both sides of the obtained inequality by N, and taking the limit $N\to\infty$ we obtain the desired result.

Let us now summarize the idea of the proof. First, the number of needed qubits per message is bounded from below by $I(\widetilde{\mathcal{E}})/N$. Now Bob obtains the final ensemble $\mathcal{E}'$ from the ensemble $\widetilde{\mathcal{E}}$ by means of quantum operation, which by the

Uhlmann theorem can only decrease the Holevo information per message. Hence we have $I(\tilde{\mathcal{E}})/N \geqslant I(\mathcal{E}')/N$. But from the lemma it follows that the initial and final ensembles have asymptotically equal Holevo information per message $I(\mathcal{E})/N \approx I(\mathcal{E}')/N$ hence we obtain $I(\tilde{\mathcal{E}})/N \geqslant I(\mathcal{E})/N$ in the limit of large N. Note here that if the bound is to be reached, then the asymptotic mean entropy of the ensemble $\tilde{\mathcal{E}}$ per message must vanish. This follows from the fact that the estimate of the $\log \dim \tilde{\varrho}$ by the Holevo information is not too rough only if the latter amounts to the von Neumann entropy.

Finally, note that for the case of blind coding the Holevo information per message must be equal for all three ensembles $\mathcal{E}$, $\tilde{\mathcal{E}}$, $\mathcal{E}'$. In other words we can say that the Holevo information is *invariant* under the asymptotically reversible quantum operations. The same *cannot* be stated for von Neumann entropy. Indeed, otherwise we would not be able to compress the signal more than indicated by the von Neumann entropy. However, we know that it is possible, e.g., for a particular ensemble considered in Ref. [4], which consists of states of disjoint supports. Here the signal states can be measured and replaced by pure ones. Then the entropy of the ensemble decreases to the value of its Holevo information. The reversal is done again by measuring the pure states and replacing them by the initial, possibly mixed ones. Applying the theorem we find that the passive and effective information are equal and take the value of the Holevo information of the ensemble. Then the information defect vanishes not only for the ensemble of pure states but also for the ensemble of mixed states with disjoint supports.

IV. COMPOSING PROTOCOLS

From the discussion of the previous section it follows that the entropy of the density matrix of the ‘‘intermediate’’ ensemble $\tilde{\mathcal{E}}$ should be as low as possible. In this section we will present a particular class of nonblind protocols that aim at decreasing the entropy. Namely, Alice can replace the input states ϱ_i with such new ones $\tilde{\varrho}_i$, acting on larger Hilbert space $\mathcal{H} = (\otimes^N \mathcal{H}_Q) \otimes \mathcal{H}'$, that $\mathrm{Tr}_{\mathcal{H}'} \tilde{\varrho}_i = \varrho_i$. Then Bob's decoding amounts to performing a partial trace, i.e., discarding the systems described by the Hilbert space $\mathcal{H}'$. Then the states $\tilde{\varrho}_i$ can produce the density matrix $\tilde{\varrho}$ of lower entropy than the initial one. Clearly, the above scheme provides perfect transmission. However, the matrix $\tilde{\varrho}$, although of perhaps small entropy, will usually occupy larger Hilbert space than the source space. To avoid it one could compose the present (ideal) protocol with the SJ protocol. Then the overall scheme is the following

$$\varrho_{i_1} \otimes \cdots \otimes \varrho_{i_k} \xrightarrow{\text{Alice's coding}} \tilde{\varrho}_{i_1} \otimes \cdots \otimes \tilde{\varrho}_{i_k} \xrightarrow{\text{SJ protocol}} \tilde{\varrho}_{i_1 \cdots i_k} \xrightarrow{\text{Bob's partial trace}} \varrho'_{i_1 \cdots i_k}. \tag{15}$$

Here i_j's are multi-indices of length N; $\varrho_{i_1} \otimes \cdots \otimes \varrho_{i_k}$ and $\varrho'_{i_1 \cdots i_k}$ act on the Hilbert space $\otimes^k (\otimes^N \mathcal{H})$ while $\tilde{\varrho}_{i_1} \otimes \cdots \otimes \tilde{\varrho}_{i_k}$ and $\tilde{\varrho}_{i_1 \cdots i_k}$ act on $\otimes^k [(\otimes^N \mathcal{H}) \otimes \mathcal{H}']$. The latter two states can be obtained from the former ones by tracing over the space $\otimes^k \mathcal{H}'$. As the used distortion measure does not increase under the partial trace operation (see the Appendix), the average distortion produced by the composed protocol is less than or equal to the one within the ‘‘intermediate’’ SJ protocol. The latter distortion tends to zero if N is kept fixed and k tends to infinity (of course N, although fixed, can be chosen arbitrarily large). Then composing the two protocols we have obtained again a compression protocol. The result can be immediately generalized as follows. *Any protocol providing perfect transmission can be composed with some other protocol, so that the full one is again a protocol, i.e., it offers asymptotically faithful transmission.*

Turning back to the considered case, we see that since the SJ protocol compresses the signal down to the value of entropy of the source ensemble per message [4,11], the following inequality holds:

$$I_p \leqslant \lim_{N \to \infty} \frac{1}{N} S(\tilde{\varrho}), \tag{16}$$

where the infimum is taken over the states $\tilde{\varrho}$ of ensembles, which partially traced produce the input ensemble $\mathcal{E}$.

Let us illustrate the above result by means of an example of binary source, i.e., the one that generates two kinds of messages ϱ_1^0 and ϱ_2^0 with probabilities p_1^0 and p_2^0, respectively (for convenience we will further omit the indices 0). Suppose that Alice replaces the single signal states by their purifications $P_i = |\psi_i\rangle\langle\psi_i|$ acting on the Hilbert space $\mathcal{H}_Q \otimes \mathcal{H}'$ [20]. As the source produces only two kinds of states, the entropy of the ensemble of purifications can be calculated explicitly

$$S(\tilde{\varrho}) = H\left[\frac{1}{2}\left[1 + \sqrt{(p_1 - p_2)^2 + 4 p_1 p_2 |\langle \psi_1 | \psi_2 \rangle|^2}\right]\right], \tag{17}$$

where $H(x) = -x \log x - (1-x) \log(1-x)$ is the binary entropy function. The minimal entropy is obtained if the overlap of ψ_1 and ψ_2 is the largest. The supremum of the overlaps of purifications of the two states ϱ_1 and ϱ_2 is given by the fidelity of the states [19,3]

$$\max |\langle \psi_1 | \psi_2 \rangle|^2 \equiv F(\varrho_1, \varrho_2) = (\mathrm{Tr} \sqrt{\sqrt{\varrho_1} \varrho_2 \sqrt{\varrho_1}})^2, \tag{18}$$

so that we obtain

$$I_e \leqslant S_{\min}(\tilde{\varrho}) = H\left[\frac{1}{2}\left[1 + \sqrt{(p_1 - p_2)^2 + 4 p_1 p_2 F(\varrho_1, \varrho_2)}\right]\right]. \tag{19}$$

Now if ϱ_1 and ϱ_2 have disjoint supports, then $F(\varrho_1, \varrho_2) = 0$ and $S_{\min}$ is equal to the Holevo information of the ensemble, which is compatible with discussion in Sec. III (and discussion in Ref. [4]). Note that the presented protocol is performed *separately* on the single messages. It seems reasonable to conjecture that if the protocol was applied to the blocks of messages then one could reach the bound of Holevo information for general ensembles. In other words it is very probable that in fact $I_e = I(\mathcal{E}_0) = \lim_{N \to \infty} (1/N)$

$\min_{\tilde{\varrho}} S(\tilde{\varrho})$. However, it is difficult to calculate the minimal asymptotic entropy per message even for the case of binary source.

V. CONCLUSION

In conclusion, we have considered the problem of compression of quantum information carried by an ensemble of mixed states. We have proved that the minimal number of qubits per message needed for asymptotically faithful transmission is greater than the Holevo information of the initial ensemble. We have also showed that any protocol providing perfect transmission can by successfully composed with another protocol. We proposed a nonblind protocol involving composition of a perfect protocol with the Schumacher-Jozsa one. The first stage is based on replacing the signal states by new states, which, partially traced, reproduce the initial ones. The proposed scheme, if applied to blocks of messages, is conjectured to reach the bound. Then the Holevo information would acquire the physical sense within the quantum information theory, being a proper generalization of von Neumann entropy to the case of ensembles of mixed states and representing the actual quantity of quantum information produced by a source [21]. The problem whether the passive information (equal to the number of needed qubits if the blind coding schemes are considered) could be sometimes strictly greater than the effective information associated with arbitrary coding schemes, remains open. Finally, we believe that the presented results will be useful in further investigations of the information content of ensemble of mixed states.

Note added: Recently the author received information that similar results were also obtained by H. Barnum, C. Caves, Ch. Fuchs, R. Jozsa, and B. Schumacher (unpublished).

ACKNOWLEDGMENTS

The author is grateful to R. Horodecki and P. Horodecki for many helpful comments, discussions, and technical remarks. He also thanks T. Matsuoka and N. Watanabe for an interesting discussion on quantum information theory. This work was supported in part by Polish Committee for Scientific Research Contract No. 2 P03B 024 12 and by the Foundation for Polish Science.

APPENDIX

Let $\Lambda : S(\mathcal{H}) \rightarrow S(\mathcal{H})$ be a trace preserving completely positive map, i.e., let it be of the following form [8]:

$$\Lambda(\varrho) = \sum_i V_i \varrho V_i^\dagger . \tag{A1}$$

Here $S(\mathcal{H})$ is the set of density matrices acting on the finite dimensional Hilbert space $\mathcal{H}$, V_i's are operators satisfying $\Sigma_i V_i^\dagger V_i = I$. It is known that Λ is of the form (A1) if and only if it can be implemented by means of a unitary transformation over a larger system [8]

$$\Lambda(\varrho) = \mathrm{Tr}_{\mathcal{H}'} U(\varrho \otimes P) U^\dagger . \tag{A2}$$

Here P is a pure state acting on the additional Hilbert space $\mathcal{H}'$; U is unitary transformation over the whole space $\mathcal{H} \otimes \mathcal{H}'$. The form (A2) justifies the fact that the completely positive trace preserving maps are identified with quantum operations.

Here we will prove the following proposition.

Proposition. The distortion $D(\varrho, \sigma)$ does not increase under quantum operations, i.e., we have

$$D(\Lambda(\varrho), \Lambda(\sigma)) \leqslant D(\varrho, \sigma). \tag{A3}$$

Proof. In view of the form (A2) it suffices to check whether D does not increase under the three components of the quantum operation: unitary transformation, partial trace, and the operation $\varrho \rightarrow \varrho \otimes P$. As $D(\varrho, \sigma) = \|\varrho - \sigma\|$ depends only on the eigenvalues of the operator $A \equiv \varrho - \sigma$, then it is unitarily invariant. Subsequently, the operators A and $A \otimes P$ have the same positive eigenvalues, so that $D(\varrho \otimes P, \sigma \otimes P) = D(\varrho, \sigma)$. Finally, suppose that A acts on the Hilbert space $\mathcal{H} \otimes \mathcal{H}'$ and has the spectral decomposition $A = \Sigma_i \lambda_i P_i$. Let us estimate the trace norm of its partial trace

$$\|\mathrm{Tr}_{\mathcal{H}'} A\| = \Big\|\sum_i \lambda_i \varrho_i\Big\| \leqslant \sum_i |\lambda_i| \, \|\varrho_i\| = \sum_i |\lambda_i| = \|A\|, \tag{A4}$$

where $\varrho_i = \mathrm{Tr}_{\mathcal{H}'} P_i$. Here we used triangle inequality for the norm and the fact that $\|\varrho_i\| = 1$. This completes the proof. The proposition holds also in the case where the operation Λ maps $S(\mathcal{H}_1)$ into $S(\mathcal{H}_2)$ with different Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$.

[1] E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).

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

[3] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).

[4] Hoi-Kwong Lo, Opt. Commun. **119**, 552 (1995).

[5] W. K. Wooters and W. H. Zurek, Nature (London) **299**, 802 (1982).

[6] G. M. D'Adriano and H. P. Yuen, Phys. Rev. Lett. **76**, 2832 (1996).

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

[8] K. Kraus, *States, Effects and Operations: Fundamental Notions of Quantum Theory* (Wiley, New York, 1991).

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

[10] There are troubles with thermodynamical justification of the von Neumann entropy formula, see, e.g., D. Dieks and V. van Dijk, Am. J. Phys. **56**, 430 (1988).

[11] R. Jozsa (unpublished).

[12] A. E. Allahverdyan and D. B. Saakian, e-print quant-ph/9702034.

[13] A. S. Holevo, Probl. Peredachi Inf. **8**, 63 (1973).

[14] H. Yuen and M. Ozawa, Phys. Rev. Lett. **70**, 363 (1993).

[15] H. Scutaru, Phys. Rev. Lett. **75**, 773 (1995).

[16] M. Fannes, Commun. Math. Phys. **31**, 291 (1973); see also M.

Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer-Verlag, Berlin, 1993), p. 22.

[17] H. Umegaki, Kodai. Math. Sem. Rep. **14**, 59 (1962).

[18] G. Lindblad, Commun. Math. Phys. **39**, 111 (1974); A. Uhlmann, Commun. Math. Phys. **54**, 21 (1977).

[19] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).

[20] Recall that the purification of a state ϱ acting on $\mathcal{H}$ is a *pure* state P acting on $\mathcal{H}\otimes\mathcal{H}'$ such that ϱ is its partial trace over the space $\mathcal{H}'$.

[21] Note that the Holevo information is a proper generalization of von Neumann entropy for the problem of sending classical information via quantum signal states, see P. Hauslanden, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wooters, Phys. Rev. A **54**, 1869 (1996); A. S. Holevo, e-print quant-ph/9708046; B. Schumacher and M. Westmoreland, Phys. Rev. A **56**, 131 (1997).