Capacities of Quantum Channels and Quantum Coherent Information

Michael Westmoreland$^{(1)}$
Benjamin Schumacher,$^{(2)}$

$^{(1)}$*Department of Mathematics, Denison University, Granville, OH 43023*

$^{(2)}$*Department of Physics and Astronomy, Kenyon College, Gambier, OH 43022*

Abstract

We derive relation between a quantum channel's capacity to convey classical information and its ability to convey quantum information. We also show that these properties of a quantum channel are related to the channel's ability to convey quantum coherent information.

I. Introduction

A quantum communication channel can be used to perform a variety of tasks, including:

1. Conveying classical information from a sender to a receiver.
2. Conveying quantum information (including quantum entanglement) from a sender to a receiver.

Each of these tasks can be performed in the presence of noise. Indeed, in quantum cryptography the noise is of central importance in revealing the activity of an eavesdropper.

A central concern in the analysis of any noisy communications channel is the channel's *capacity*: the maximum rate at which information (classical or quantum) can be reliably transmitted through the channel. When we use quantum systems to convey information we confront concerns which are not present in classical channels. A chief concern is the measurement performed by the receiver in order to extract the information.

Measurement of the received message is not particular to quantum channels; a receiver using a classical channel must also measure the received message. Counting the numbers of dots and dashes in a Morse codeword is a measurement. What *is* a particular concern for the user of a quantum channel is that a measurement result may be ambiguous *even if the quantum channel is noiseless.* For example, assume that Alice, the sender and Bob, the receiver, agree to use vertically polarized photons to represents a "1" and horizontally polarized photons to represent a "0". We also assume that Bob measures for photons polarized along an axis that is 45 degrees from horizontal. Quantum mechanics tells us that, in this case, Bob could not tell if Alice sent a "1" or a "0". This is true even if Alice's photon reaches Bob without ant distortion in its angle of polarization. One might object (quite rightfully in this case) that Alice and Bob settled on a particularly silly measurement to use in reading the signals. If the letter states (photon polarizations here) are not orthogonal then there is no measurement Bob can perform that will perfectly distinguish them. This is a first difference between classical and quantum channels: measurements are, in general, ambiguous, even if no noise is present.

One might suggest that if Bob cannot use a single measurement to decode Alice's signal then Bob should perform several measurements on each photon. Quantum mechanics prevents implementation of such a scheme: a measurement of a quantum system fundamentally changes the properties of that system. We return to our example where Alice and Bob used vertical and horizontal polarizations and a measurement at 45 degrees. After Bob has measured a given photon, it is the in a pure state of 45 degree polarization. That is, if a second polarizer is set up immediately after the first with the same (45 degree) polarization then the photon will pass it with certainty. If the polarizer is set at any other angle, then there is a nonzero chance that the particle will not pass. Indeed, even if Alice dispatches the photon with vertical polarization and if it passes Bob's first polarizer, it will then have only a 50% chance of passing a subsequent vertical polarizer. This is a second difference between classical and quantum channels: measurements of the received signals fundamentally change the signal.

One might further suggest that if Bob's measurements change the state of the transmitted systems, then Bob should make several copies of each signal state. He could the perform a single measurement on each copy and thereby extract Alice's message. Once again, quantum mechanics prevents such a solution. The problem here is that general quantum states can not be perfectly copied. This is the content of the no-cloning theorem of Wootters and Zurek [3]. This is a third difference between quantum and classical channels: quantum states cannot be cloned.

It should be noted that the no-cloning theorem does not imply that Alice can not make multiple copies of the states she is sending. All that she needs to do is prepare multiple systems in the same way; this is not cloning. If Alice and Bob decide to do this to increase the reliability of their channel it would decrease the capacity of the channel. This is because capacity is the *rate* at which information is transmitted per letter state. If Alice send more letter states, the capacity of the channel is reduced.

Given what has been said to this point, one might conclude that even finding the classical capacity of a quantum channel is problematic at best. In fact, we do know the classical capacity, even for noisy quantum channels. This result is presented in Section II.

To this point, we have been concerned with the classical capacity of a quantum channel. One can also analyze the *quantum* capacity of a quantum communications channel. This is the ability of the channel to faithfully transmit a quantum state from one system to another. The capacity of a noiseless quantum channel is known [11] but the case of the noisy channel is still under analysis even though some progress has been made [20]. Section III presents the concept of coherent quantum information [18] and some of its properties. Section III concludes with an anlysis showing the relation between the classical capacity of a channel, the coherent information conveyed by that channel and the quantum capacity of the channel.

II. Classical capacity of a noisy quantum channel

Suppose Alice wishes to convey classical information to Bob by using a quantum system Q as a communication channel. Alice prepares the channel in one of various quantum states W_x with *a priori* probabilities p_x. Bob makes a measurement on the system Q, and from its result he tries to infer which state Alice prepared. A theorem stated by Gordon [1] and Levitin [5], first proved by Kholevo [6], gives an upper bound to the amount of information that Bob can obtain about Alice's signal. If $W = \sum_x p_x W_x$ is the density operator describing the ensemble of Alice's signals, then the mutual information $H(X:Y)$ between Alice's input X and Bob's output Y is bounded by

$$H(X:Y) \leq H(W) - \sum_x p_x H(W_x), \tag{1}$$

where $H(W) = -\mathrm{Tr}\, W \log W$, the von Neumann entropy of the density operator W. The upper bound in Equation 1 is in general a weak one, in that Bob may not be able to choose an observable that gives him an amount of information near to the upper bound [7].

Suppose that Alice employs signal states W_x that are *mixed* states. Then can Alice and Bob find a choice of code and decoding observable so that the general Levitin - Kholevo bound (Equation 1) can be approached arbitrarily closely? In this paper, we show that the answer to this question is "yes". That is, we prove the following result:

> **Theorem.** Suppose we have letter states W_x with *a priori* probabilities p_x, and let
>
> $$\chi = H(W) - \sum_x p_x H(W_x).$$

> Fix $\epsilon, \delta > 0$. Then for sufficiently large L, there exist a code (whose codewords are strings of L letters) and a decoding observable such that the information carried per letter is at least $\chi - \delta$ and the probability of error $P_E < \epsilon$.

The proof employs an average over randomly generated codes to establish the existence of a satisfactory code. (If the average probability of error is small for an ensemble of codes, the ensemble must contain specific codes with small probability of error.) We also use a similar prescription for Bob's decoding observable. The chief refinement in the proof presented here is the enforcement of stronger "typicality" conditions on various quantities associated with the channel.

The mixed states W_x may be thought of as the outputs of a *noisy* quantum channel. Thus, our main result will enable us to draw conclusions about the classical information capacity of a noisy quantum channel.

We have shown that it is possible to send information at any rate up to χ bits per letter with arbitrarily low probability of error. The capacity of a channel is defined as the maximum information per letter that may be sent through the channel with P_E arbitrarily small. Thus, χ provides a lower bound to the capacity of the quantum channel.

Classical information theory together with the Levitin - Kholevo Theorem also allows us to use χ to establish an *upper bound* for the capacity of the channel. Suppose X represents Alice's input and Y represents Bob's decoding measurement outcome. Then the Fano inequality [10] states that

$$-P_E \log P_E - (1-P_E)\log(1-P_E) + P_E \log(N_X - 1) \geq H(X|Y) \tag{2}$$

where P_E is the probability of error and N_X is the number of possible values of X. $H(X|Y)$ is the conditional Shannon entropy of X given Y—that is, the entropy of the conditional distribution $p(x|y)$, averaged over the various values of y. It is related to the mutual information $H(X:Y)$ by

$$H(X|Y) = H(X) - H(X:Y). \tag{3}$$

In the channel, Alice uses some signal states ρ_a with probabilities P_a. Levitin - Kholevo Theorem places an upper bound on the mutual information $H(X:Y)$:

$$H(X:Y) \leq H(\rho) - \sum_a P_a H(\rho_a).$$

(Note that, if the channel used by Alice and Bob consists of L letters used independently, then the Levitin - Kholevo bound is just $L\,\chi$, where χ is the Levitin - Kholevo bound for a single letter.) If the Alice's input X has an entropy $H(X)$ that exceeds $H(\rho) - \sum_a P_a H(\rho_a)$, then $H(X|Y) > 0$ and it will not be possible to make the probability of error P_E arbitrarily small.

Suppose we fix an alphabet $\Gamma = \{W_x\}$ of letter states W_x, and require that Alice use codewords a that are length-L strings of these letter states: $a = x_1 \ldots x_L$. Then the probability distribution P_a yields marginal probability distributions $p(x_1), \ldots, p(x_L)$ and average density operators $W_1, \ldots, W_L$ for the L different letters. It follows that

$$\begin{aligned} H(\rho) - \sum_a P_a H(\rho_a) \quad &\leq \quad \left(H(W_1) - \sum_{x_1} p(x_1) H(W_{x_1}) \right) \\ &\quad + \cdots + \left(H(W_L) - \sum_{x_L} p(x_L) H(W_{x_L}) \right) \qquad (4) \end{aligned}$$

where we have used the subadditivity of the entropy $H(\rho)$. We might write this as

$$\chi^{(L)} \leq \chi_1 + \cdots + \chi_L \qquad (5)$$

where $\chi^{(L)}$ represents the Levitin - Kholevo bound for the ensemble of codewords of length L, and $\chi_1, \ldots, \chi_L$ represent Levitin - Kholevo bounds for the individual letter ensembles.

We define the *fixed-alphabet capacity* C_Γ to be

$$C_\Gamma = \sup_{p(x)} \chi \qquad (6)$$

where $p(x)$ is the probability distribution over the letters states in Γ and χ is the single-letter Levitin - Kholevo bound. This quantity represents the maximum information rate per letter that Alice can send to Bob with arbitrarily low probability of error.

This claim follows directly from our results so far. Suppose Alice uses codewords of length L. Then $\chi^{(L)} \leq L\, C_\Gamma$; and by the above argument, if Alice attempts to send more than $L\, C_\Gamma$ bits using these codewords then the probability of error will not be arbitrarily small. Conversely, we can choose the letter probabilities so that χ is as close as required to C_Γ, and we have previously shown that a suitable choice of code and decoding observable can convey up to χ bits per letter with arbitrarily low P_E. Thus, the capacity C_Γ cannot be exceeded but can be approached arbitrarily closely.

The mixed states W_x used in our alphabet are the states available to Bob for decoding. They may in fact not be the original states of the channel Q chosen by Alice. In the interval between Alice's encoding and Bob's decoding, the system Q may have undergone unitary internal evolution (which Bob can correct by a suitable choice of "rotated" decoding observable) and interaction with the external environment (which Bob cannot in general correct).

The most general description of the evolution of a quantum system Q interacting with an environment is provided by a trace-preserving completely positive linear map on the set of density operators of Q [14]. Such a map is described by a superoperator $\mathcal{E}$:

$$\rho \longrightarrow \rho' = \mathcal{E}(\rho), \tag{7}$$

where ρ is the initial state of the system and ρ' is the final state. The superoperator $\mathcal{E}$ acts linearly, so that a convex combination of input states yields a convex combination of output states. This description clearly includes unitary evolution of Q as a special case, but it also can account for interaction with the environment.

A noisy quantum channel is defined by a superoperator $\mathcal{E}$ that describes the evolution of each letter as it is transmitted from Alice to Bob. We assume that the channel is memoryless—i.e., that the evolution of each letter is independent. This means, among other things, that a product state of several input letters will evolve into a product state output.

Alice's basic problem is to use input states w_x so that the output states $W_x = \mathcal{E}(w_x)$ can be distinguished by Bob. If Alice has a fixed alphabet $\{w_x\}$ of input states, then the maximum achievable information rate per letter is still given by our fixed-alphabet capacity C_Γ, where Γ is the alphabet of *output* states.

Now suppose that Alice is allowed to choose her input states in order to maximize the information conveyed to Bob over the noisy quantum channel, subject to the constraint that Alice must transmit codewords which are represented by product states of the letters. This *almost* reduces to the fixed-alphabet problem, where the fixed alphabet Γ now includes all of the possible output states of the channel. The maximum over probability distributions is now a maximum over all input ensembles of states chosen by Alice.

We say that this problem *almost* reduces to the fixed alphabet problem in that the argument that χ is an upper bound of the capacity must be modified in this case. Recall from the previous section that we applied the classical Fano inequality to show that if Alice attempts to send information at a rate exceeding χ then the probability of error cannot be made arbitrarily small. If we attempt to use the same argument in the present case then the Fano inequality does not help us for at least two reasons. First, the number of possible input states N_x is unbounded. Second, we do not have a characterization of $H(XlineY)$ that allows us to compare it with N_x. Thus we will modify the Fano inequality to understand the behavior of the probability of error in the present case.

We first note that the probability of "getting it right"

$$1 - P_E = \frac{1}{N} \sum_{\alpha k} p_{k|\alpha} |\, \langle \tilde{\mu}_{\alpha k} \mid \ s_{\alpha k} \rangle \,|^2 \tag{8}$$

is linear in the elements of the POM. Thus the probability of error, P_E is a convex function on the elements of the POM. We may modify the proof of a result of Davies

(Theorem 3 of [17]) to show that the convex function P_E is minimized by a POM having no more than d^2 elements, where d is the dimension of the support of the POM. Thus, the probability of error is minimized by a decision scheme in which at most d^2 of the inputs are identified by the decision scheme. Let us denote the output of such a scheme by Y_{min}. Fano's inequality gives us that

$$-P_E \log P_E - (1 - P_E)\log(1 - P_E) + P_E \log(d^2 - 1) \geq H(X|Y_{min}). \tag{9}$$

Note that

$$\begin{aligned} H(X|Y_m) &= H(X) - H(X : Y_m min) & (10) \\ &\geq H(X) - \chi, & (11) \end{aligned}$$

so that we conclude

$$-P_E \log P_E - (1 - P_E)\log(1 - P_E) + P_E \log(d^2 - 1) \geq H(X) - \chi. \tag{12}$$

Note that this is a relation between the minimum probability of error and a quantity $(H(X) - \chi)$ which does not depend on the particular decision scheme. We see that if Alice attempts to send information at a rate $H(X)$ in excess of χ then the probability of error can not be made arbitrarily small.

We now turn to a demonstration that this rate can be achieved. Alice wishes to choose a set of input states w_x (together with input probabilities p_x) so that χ is maximized for the output states W_x. We next show that Alice can do no better than choose the input states w_x to be pure. Let a set of (possibly mixed) input states w_x be given along with their a priori probabilities, and let

$$W = \sum_x p_x W_x = \sum_x p_x \mathcal{E}(w_x) \tag{13}$$

be the average output state. Then

$$\chi = H(W) - \sum_x p_x H(\mathcal{E}(w_x)). \tag{14}$$

Construct a new set of pure state inputs by resolving each mixed state input into a convex combination of pure states:

$$w_x = \lambda_{xk} \,|\, \psi_{xk}\rangle \,\langle \psi_{xk} \,| \,. \tag{15}$$

We will use the state $|\,\psi_{xk}\rangle$ with probability $p_{xk} = p_x\,\lambda_{xk}$. By linearity,

$$W_x = \mathcal{E}(w_x) = \sum_k \lambda_{xk} \mathcal{E}(|\,\psi_{xk}\rangle\,\langle\psi_{xk}\,|), \tag{16}$$

so that the average output state is still W, as before. By the convexity of the von Neumann entropy,

$$H(W_x) \geq \sum_k \lambda_{xk} H(\mathcal{E}(|\,\psi_{xk}\rangle\,\langle\psi_{xk}\,|). \tag{17}$$

It follows that

$$\begin{aligned} \chi' &= H(W) - \sum_{xk} p_{xk} H(\mathcal{E}(|\,\psi_{xk}\rangle\,\langle\psi_{xk}\,|) \\ &\geq H(W) - \sum_x p_x H(\mathcal{E}(w_x)) = \chi. \end{aligned} \tag{18}$$

In other words, for any ensemble of mixed input states, we can find an ensemble of pure input states whose output states have a χ at least as great. The optimal inputs for the noisy quantum channel are pure states.

To sum up, if Alice is required to use product states to represent her codewords, then the capacity $C^{(1)}$ of the noisy quantum channel is

$$C^{(1)} = \max \chi \tag{19}$$

where χ is the Levitin - Kholevo bound for the output states of the channel, and the maximum is taken over all ensembles of pure state inputs. Alice can reliably transmit information to Bob at any rate below $C^{(1)}$. We will refer to $C^{(1)}$ as the product state capacity. The superscript (1) reminds us that Alice is required to use the multiple available copies of the channel *one at a time*, coding her messages into product states.

III. Coherent quantum information and quantum capacity

The entropy exchange S_e measures the amount of information that is exchanged between the system Q and the environment E during their interaction. If the environment is initially in a pure state, the entropy exchange is just the environment's entropy after the interaction—i.e., $S_e = S(\rho^{E'})$, where $\rho^{E'}$ is the final state of E. (The entropy here is just the ordinary von Neumann entropy of a density operator, $S(\rho) = -\mathrm{Tr}\,\rho\log\rho$.) The entropy exchange is entirely determined by the initial state ρ^Q of Q and the channel dynamics superoperator $\mathcal{E}^Q$; that is, the entropy exchange is a property "intrinsic" to Q and its dynamics.

The coherent information I_e, introduced in [18], is given by

$$I_e = S(\rho^{Q'}) - S_e. \tag{20}$$

The coherent information has many properties that suggest it as the proper measure of the quantum information conveyed from Alice to Bob by the channel. For example, I_e can never be increased by quantum data processing performed by Bob on the channel output, and perfect quantum error correction of the channel output is possible for

Bob if and only if no coherent information is lost in the channel [18]. Finally, the coherent information seems to be related to the capacity of a quantum channel to convey quantum states with high fidelity [20].

Alice might be using the channel to send classical information to Bob. Alice prepares Q in one of a set of possible "signal states" ρ_k^Q, which are used by Alice with *a priori* probabilities p_k. The average state ρ^Q is given by

$$\rho^Q = \sum_k p_k \rho_k^Q . \tag{21}$$

Bob receives the kth signal as $\rho_k^{Q'} = \mathcal{E}^Q(\rho_k^Q)$. Because the superoperator is linear, the average received state is

$$\rho^{Q'} = \sum_k p_k \mathcal{E}^Q(\rho_k^Q) = \mathcal{E}^Q(\rho^Q). \tag{22}$$

Bob attempts to decode Alice's message (that is, to identify which signal state was chosen by Alice) by measuring some *decoding observable* on his received system Q'.

The amount of classical information conveyed from Alice to Bob, which we will denote H_{Bob}, is governed by the quantity $\chi^{Q'}$, defined by

$$\chi^{Q'} = S(\rho^{Q'}) - \sum_k p_k S(\rho_k^{Q'}). \tag{23}$$

This quantity is significant in two ways:

- $H_{Bob} \leq \chi^{Q'}$, regardless of the decoding observable chosen [21, 22].
- H_{Bob} can be made as close as desired to to $\chi^{Q'}$ by a suitable choice of code and decoding observable. To make H_{Bob} near $\chi^{Q'}$, Alice must in general use the channel many times and employ code words composed of many signals; Bob must perform his decoding measurement on entire code words. The net result is that the channel is used N times to send up to $N\chi^{Q'}$ bits of classical information reliably [2].

In short, $\chi^{Q'}$ represents an upper bound on the classical information conveyed from Alice to Bob, an upper bound that may be approached arbitrarily closely if Alice and Bob use the channel efficiently.

If this general picture is used to describe a noisy quantum channel, then we need to account for the information that is passed to the environment. Recall that the evolution superoperator $\mathcal{E}^Q$ describes all of the effects of the channel; or, to put it another way, all of properties of the link between Alice and Bob are contained in the interaction operator U^{QE}. The information passed to the environment H_E will be limited by

$$\chi^{E'} = S(\rho^{E'}) - \sum_k p_k S(\rho_k^{E'}). \tag{24}$$

Assume that the states of Q initially prepared by Alice are pure states $\left|\phi_k^Q\right\rangle$; also recall that the environment E can be presumed to begin in a pure state $\left|0^E\right\rangle$. After Q and E interact unitarily, the joint state $\left|\Psi_k^{QE'}\right\rangle = U^{QE}\left|\phi_k^Q\right\rangle \otimes \left|0^E\right\rangle$ will also be a pure state, generally an entangled one. The subsystem states, described by density operators

$$\begin{aligned} \rho_k^{Q'} &= \mathrm{Tr}_E \left|\Psi_k^{QE'}\right\rangle \left\langle \Psi_k^{QE'}\right| \\ \rho_k^{E'} &= \mathrm{Tr}_Q \left|\Psi_k^{QE'}\right\rangle \left\langle \Psi_k^{QE'}\right|, \end{aligned} \tag{25}$$

will have exactly the same non-zero eigenvalues, so that $S(\rho_k^{Q'}) = S(\rho_k^{E'})$. Therefore

$$\begin{aligned} I^Q &= S(\rho^{Q'}) - S_e \\ &= S(\rho^{Q'}) - S(\rho^{E'}) \\ &= S(\rho^{Q'}) - \sum_k p_k S(\rho_k^{Q'}) - S(\rho^{E'}) + \sum_k p_k S(\rho_k^{E'}) \\ I^Q &= \chi^{Q'} - \chi^{E'}. \end{aligned} \tag{26}$$

It is interesting to note that, although both $\chi^{Q'}$ and $\chi^{E'}$ depend on the choice of pure state inputs for the channel Q, the difference $\chi^{Q'} - \chi^{E'}$ depends only on the overall density operator ρ^Q for the inputs.

In is shown in [18] that perfect error correction is possible if and only if the coherent information of the channel equals the entropy of the input state. The quantity D_e is defined [24] as follows:

$$D_E = S(\rho^Q) - I^Q. \tag{27}$$

We can thus say that perfect error correction is possible if and only if $D_e = 0$.

Subtracting each side of equation (26) from the entropy of the input state yields:

$$S(\rho^Q) - I^Q = D_e = S(\rho^Q) - \chi^{Q'} + \chi^{E'}. \tag{28}$$

Recall that χ^Q is maximized when Alice uses pure state to encode her messages. In that case we have $\chi^Q = S(\rho^Q)$, so equation (28) takes the form

$$D_e = \chi^Q - \chi^{Q'} + \chi^{E'}. \tag{29}$$

This equation is quite informative. It implies that conveying quantum information perfectly depends on two tasks: Maximization of classical capacity and zero entropy loss to the environment. This implies a strong connection between the quantum capacity of a quantum channel and its classical capacity.

We would like to thank W. K. Wootters and M. A. Nielsen for helpful conversations and suggestions.

References

[1] J. P. Gordon, in *Quantum Electronics and Coherent Light, Proceedings of the International School of Physics "Enrico Fermi," Course XXXI*, edited by P. A. Miles (Academic, New York, 1964), pp. 156-181.

[2] P. Hausladen, R. Jozsa, B. Schumacher, M. D. Westmoreland and W. K. Wootters, Phys. Rev. A **54**, 1869 (1996). B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997). A. S. Holevo, IEEE Trans. Inf. Theory (to be published).

[3] W.K. Wootters and W.H. Zurek, *Nature***299**, 802 (1982)

[4] B. Schumacher and M. Westmoreland, "Coherent quantum information and quantum privacy", submitted to Phys. Rev. Lett.

[5] L. B. Levitin, "On the quantum measure of the amount of information," in *Proceedings of the IV National Conference on Information Theory*, Tashkent, 1969, pp. 111–115 (in Russian); "Information Theory for Quantum Systems," in *Information, Complexity, and Control in Quantum Physics*, edited by A. Blaquière, S. Diner, and G. Lochak (Springer, Vienna, 1987).

[6] A. S. Kholevo, *Probl. Inform. Transmission* **9**, 177 (1973) (translated from *Problemy Peredachi Informatsii*).

[7] C. A. Fuchs and C. M. Caves, *Phys. Rev. Lett.* **73**, 3047 (1994).

[8] P. Hausladen, R. Josza, B. Schumacher, M. Westmoreland, and W. K. Wootters, *Phys. Rev. A* **54**, 1869 (1996).

[9] A. S. Kholevo, to appear in *IEEE Transactions on Information Theory.*

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

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

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

[13] A. S. Kholevo *Probl. Inform. Transmission* **15**, 3 (1979) (translated from *Problemy Peredachi Informatsii*).

[14] K. Hellwig and K. Kraus, *Communications in Mathematical Physics* **16**, 142 (1970). M.-D. Choi, *Linear Algebra and its Applications* **10**, 285 (1975). K. Kraus, *States Effects and Operations: Fundamental Notions of Quantum Theory* (Springer-Verlag, Berlin, 1983).

[15] C. H. Bennett,C. A. Fuchs, and J. Smolin, to appear in *Proceedings of the 3rd International Workshop on Quantum Communication and Measurement*.

[16] C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379, 623 (1948).

[17] E. B. Davies, *IEEE Transactions on Information Theory* **IT-24**, 596 (1978).

[18] B. Schumacher and M. A. Nielsen, Phys. Rev. A **54**, 2629 (1996).

[19] B. Schumacher, Phys. Rev. A **54**, 2614 (1996).

[20] S. Lloyd, Phys. Rev. A **55**, 1613 (1997). H. Barnum, M. A. Nielsen and B. Schumacher, Report No. quant-ph/9702049.

[21] A. S. Kholevo, Probl. Peredachi Inf. **9**, 3 (1973) [Probl. Inf. Transm. (USSR) **9**, 110 (1973)].

[22] B. Schumacher, M. D. Westmoreland and W. K. Wootters, Phys. Rev. Lett. **76**, 3452 (1996).

[23] I. Csiszár and J. Körner, *IEEE Transactions on Information Theory* **24**, 339 (1978). U. M. Maurer, *IEEE Transactions on Information Theory* **39**, 733 (1993).

[24] B. Schumacher, “Entropy exchange and coherent quantum information” in *Proceedings of the Fourth Workshop on Physics and Computation*; edited by T. Toffoli, M. Biafore, and J. Leao (New England Complex Systems Institute, Boston, 1996).