Information transmission through a noisy quantum channel

Howard Barnum,[1,2,*] M. A. Nielsen,[1,2,†] and Benjamin Schumacher[3,‡]

[1] *Institute for Theoretical Physics, University of California, Santa Barbara, California 93106-4030*

[2] *Center for Advanced Studies, Department of Physics and Astronomy, University of New Mexico,
Albuquerque, New Mexico 87131-1156*

[3] *Department of Physics, Kenyon College, Gambier, Ohio 43022*

Noisy quantum channels may be used in many information-carrying applications. We show that different applications may result in different channel capacities. Upper bounds on several of these capacities are proved. These bounds are based on the *coherent information*, which plays a role in quantum information theory analogous to that played by the mutual information in classical information theory. Many new properties of the coherent information and entanglement fidelity are proved. Two nonclassical features of the coherent information are demonstrated: the failure of subadditivity, and the failure of the pipelining inequality. Both properties arise as a consequence of quantum entanglement, and give quantum information new features not found in classical information theory. The problem of a noisy quantum channel with a classical observer measuring the environment is introduced, and bounds on the corresponding channel capacity proved. These bounds are always greater than for the unobserved channel. We conclude with a summary of open problems. [S1050-2947(98)04005-0]

I. INTRODUCTION

A central result of Shannon's classical theory of information [1–3] is the *noisy-channel coding theorem*. This result provides an *effective procedure* for determining the *capacity* of a noisy channel—the maximum rate at which classical information can be reliably transmitted through the channel. There has been much recent work on quantum analogues of this result [4–8].

This paper has two central purposes. The first purpose is to develop general techniques for proving upper bounds on the capacity of a noisy quantum channel, which are applied to several different classes of quantum noisy-channel problems. Second, we point out some essentially new features that quantum mechanics introduces into the noisy-channel problem.

The paper is organized as follows. In Sec. II we give a basic introduction to the problem of the noisy quantum channel, and explain the key concepts. Section III reviews the *quantum operations* formalism that is used throughout the paper to describe a noisy quantum channel, and Sec. IV reviews the concept of the *entropy exchange* associated with a quantum operation. Section V shows how the classical noisy-channel coding theorem can be put into the quantum language, and explains why the capacities that arise in this context are not useful for applications such as quantum computing and teleportation. Section VI discusses the *entanglement fidelity*, which is the measure we use to quantify how well a state and its entanglement are transmitted through a noisy quantum channel. Section VII discusses the *coherent information* introduced in [5] as an analogue to the concept

of *mutual information* in classical information theory. Many new results about the coherent information are proved, and we show that quantum entanglement allows the coherent information to have properties that have no classical analogue. These properties are critical to understanding what is essentially quantum about the quantum noisy-channel coding problem. Section VIII brings us back to noisy-channel coding, and formally sets up the class of noisy-channel coding problems we consider. Section IX proves a variety of upper bounds on the capacity of a noisy quantum channel, depending on what class of coding schemes one is willing to allow. This is followed in Sec. X by a discussion of the achievability of these upper bounds and of earlier work on channel capacity. Section XI formulates the new problem of a noisy quantum channel with measurement, allowing classical information about the environment to be obtained by measurement, and then used during the decoding process. Upper bounds on the corresponding channel capacity are proved. Finally, Sec. XII concludes with a summary of our results, a discussion of the new features that quantum mechanics adds to the problem of the noisy channel, and suggestions for further research.

II. NOISY-CHANNEL CODING

The problem of noisy-channel coding will be outlined in this section. Precise definitions of the concepts used will be given in later sections. The procedure is illustrated in Fig. 1.
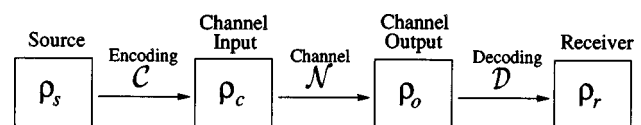

FIG. 1. The noisy quantum channel, together with encodings and decodings.

*Electronic address: hbarnum@tangelo.phys.unm.edu

†Electronic address: mnielsen@tangelo.phys.unm.edu

‡Electronic address: schumacb@kenyon.edu

There is a *quantum source* emitting unknown quantum states, which we wish to transmit through the channel to some receiver. Unfortunately, the channel is usually subject to noise, which prevents it from transmitting states with high fidelity. For example, an optical fiber suffers losses during transmission. Another important example of a noisy quantum channel is the memory of a quantum computer. There the idea is to transmit quantum states *in time*. The effect of transmitting a state from time t_1 to t_2 can be described as a noisy quantum channel. Quantum teleportation [9] can also be described as a noisy quantum channel whenever there are imperfections in the teleportation process [6,10].

The idea of noisy-channel coding is to encode the quantum state emitted by the source, ρ_s, which one wishes to transmit, using some *encoding operation*, which we denote $\mathcal{C}$. The encoded state is then sent through the channel, whose operation we denote by $\mathcal{N}$. The output state of the channel is then *decoded* using some *decoding operation* $\mathcal{D}$. The objective is for the decoded state to match with high fidelity the state emitted by the source. As in the classical theory, we consider the fidelity of large blocks of material produced by repeated emission from the source, and allow the encoding and decoding to operate on these blocks. A channel is said to transmit a source reliably if a sequence of block-coding and block-decoding procedures can be found that approaches perfect fidelity in the limit of large block size.

What then is the *capacity* of such a channel—the highest rate at which information can be reliably transmitted through the channel? The goal of a *channel capacity theorem* is to provide a procedure to answer this question. This procedure must be an *effective procedure*, that is, an explicit algorithm to evaluate the channel capacity. Such a theorem comes in two parts. One part proves an upper bound on the rate at which information can be reliably transmitted through the channel. The other part demonstrates that there are coding and decoding schemes that attain this bound, which is therefore the channel capacity. We do not prove such a channel capacity theorem in this paper. We do, however, derive bounds on the rate at which information can be sent through a noisy quantum channel.

III. QUANTUM OPERATIONS

What is a quantum noisy channel, and how can it be described mathematically? This section reviews the formalism of quantum operations, which is used to describe noisy channels. Previous papers on the noisy-channel problem [4–8] have used apparently different formalisms to describe the noisy channel. In fact, all the formalisms can be shown to be equivalent, as we shall see in this section. Historically, quantum operations have also sometimes been known as *completely positive maps* or *superscattering operators*. The motivation in all cases has been to describe general state changes in quantum mechanics.

A simple example of a state change in quantum mechanics is the unitary evolution experienced by a closed quantum system. The final state of the system is related to the initial state by a unitary transformation U,

$$\rho \rightarrow \mathcal{E}(\rho) = U\rho U^{\dagger}. \tag{3.1}$$

Although all closed quantum systems are described by unitary evolutions, in accordance with Schrödinger's equation, more general state changes are possible for open quantum systems, such as noisy quantum channels.

How does one describe a general state change in quantum mechanics? The answer to this question is provided by the quantum operations formalism. This formalism is described in detail by Kraus [11] (see also Hellwig and Kraus [12]) and is given short but detailed reviews in Choi [13] and in the Appendix to [4]. In this formalism there is an *input state* and an *output state*, which are connected by a map,

$$\rho \rightarrow \frac{\mathcal{E}(\rho)}{\mathrm{tr}[\mathcal{E}(\rho)]}. \tag{3.2}$$

This map is a *quantum operation* $\mathcal{E}$, a linear, trace-decreasing map that preserves positivity. The trace in the denominator is included in order to preserve the trace condition, $\mathrm{tr}(\rho) = 1$.

The most general form for $\mathcal{E}$ that is physically reasonable (in addition to being linear and trace decreasing and preserving positivity, a physically reasonable $\mathcal{E}$ must satisfy an additional property called complete positivity), can be shown to be [11]

$$\mathcal{E}(\rho) = \sum_i A_i \rho A_i^{\dagger}. \tag{3.3}$$

The system operators A_i, which must satisfy $\Sigma_i A_i^{\dagger} A_i \leq I$, completely specify the quantum operation. In the particular case of a unitary transformation, there is only one term in the sum $A_1 = U$, leaving us with the transformation (3.1).

A class of operations that is of particular interest is the *trace-preserving* or *nonselective* operations. Physically, these arise in situations where the system is coupled to some environment that is not under observation; the effect of the evolution is averaged over all possible outcomes of the interaction with the environment. Trace-preserving operations are defined by the requirement that

$$\sum_i A_i^{\dagger} A_i = I. \tag{3.4}$$

This is equivalent to requiring that for all density operators ρ,

$$\mathrm{tr}[\mathcal{E}(\rho)] = 1, \tag{3.5}$$

explaining the nomenclature ''trace preserving.'' Notice that this means the evolution equation (3.2) reduces to the simpler form

$$\rho \rightarrow \mathcal{E}(\rho), \tag{3.6}$$

when $\mathcal{E}$ is trace preserving.

The following *representation theorem* is proved in [11], [13], and [4]. It shows the connection between trace-preserving quantum operations and systems interacting unitarily with an environment, and thus provides part of the justification for the physical interpretation of trace-preserving quantum operations described above.

Theorem (representation theorem for trace-preserving quantum operations). Suppose $\mathcal{E}$ is a trace-preserving quan-
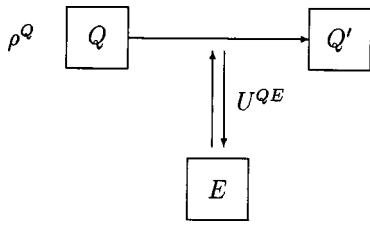
FIG. 2. Quantum operations arise when a system Q interacts with an environment E.

FIG. 3. Quantum operations in the presence of a reference system R.

tum operation on a system with a d-dimensional state space. Then it is possible to construct an ''environment'' E of at most d^2 dimensions, such that the system and environment are initially uncorrelated, the environment is initially in a pure state $\sigma = |s\rangle\langle s|$, and there exists a unitary evolution U on system and environment such that

$$\mathcal{E}(\rho) = \mathrm{tr}_E[U(\rho \otimes \sigma)U^\dagger]. \tag{3.7}$$

Here and elsewhere in the paper a subscript on a trace indicates a partial trace over the corresponding system (E in this case).

Conversely, given any initially uncorrelated environment σ (possibly of more than d^2 dimensions, and initially impure), a unitary interaction U between the system and the environment gives rise to a trace-preserving quantum operation,

$$\mathcal{E}(\rho) = \mathrm{tr}_E[U(\rho \otimes \sigma)U^\dagger]. \tag{3.8}$$

This theorem tells us that any trace-preserving quantum operation can always be *mocked up* as a unitary evolution by adding an environment with which the system can interact unitarily. Conversely, it tells us that any such unitary interaction with an initially uncorrelated environment gives rise to a trace-preserving quantum operation. Both of these facts are useful in what follows. The picture we have of a quantum operation is neatly summarized in Fig. 2.

Here, Q denotes the state of the system before the interaction with the environment, and Q' the state of the system after the interaction. Unless stated otherwise we follow the convention that Q and Q' are d dimensional. The environment system E and the operator U^{QE} might be chosen to be the actual physical environment and its interaction with Q, but this is not necessary. The only thing that matters for the description of noisy channels is the dynamics of Q. For any given quantum operation $\mathcal{E}$ there are many possible representations of $\mathcal{E}$ in terms of environments E and interactions U^{QE}. We always assume that the initial state of E is a *pure state*, and regard E as a mathematical artifice. Of course, the actual physical environment E_A may be initially impure, but the above representation theorem shows that for the purposes of describing the dynamics of Q, it can be replaced by an ''environment'' E that is initially pure and gives rise to exactly the same dynamics. In what follows it is this latter E that is most useful.

Shannon's classical noisy coding theorem is proved for *discrete memoryless channels*. Discrete means that the channel only has a finite number of input and output states. By analogy we define a discrete quantum channel to be one that has a finite number of Hilbert space dimensions. In the clas-
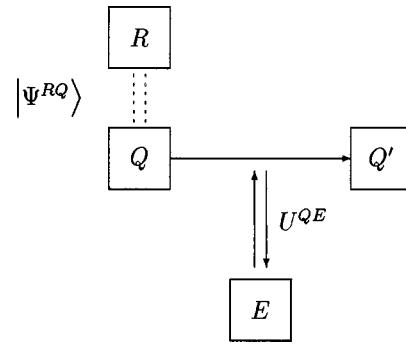
sical case, memoryless means that the output of the channel is independent of the past, conditioned on knowing the state of the source. Quantum mechanically we take this to mean that the output of the channel is completely determined by the encoded state of the source, and is not affected by the previous history of the source.

Phrased in the language of quantum operations, we assume that there is a quantum operation $\mathcal{N}$ describing the dynamics of the channel. The input ρ_i of the channel is related to the output ρ_o by the equation

$$\rho_i \rightarrow \rho_o = \mathcal{N}(\rho_i). \tag{3.9}$$

For the majority of this paper we assume, as in the previous equation, that the operation describing the action of the channel is trace preserving. This corresponds to the physical assumption that no classical information about the state of the system or its environment is obtained by an external classical observer. All previous work on noisy-channel coding with the exception of [14] has assumed that this is the case, and we do so for the majority of the paper. In Sec. XI we consider the case of a noisy channel that is being observed by some classical observer.

In addition to the environment E it is also extremely useful to introduce a *reference system* R in the following way. One might imagine that the system Q is initially part of a larger system RQ and that the total is in a pure state $|\psi^{RQ}\rangle$ satisfying

$$\rho^Q = \mathrm{tr}_R(|\psi^{RQ}\rangle\langle\psi^{RQ}|). \tag{3.10}$$

Such a state $|\psi^{RQ}\rangle$ is called a *purification* of ρ^Q, and it can be shown [15] that such a system R and purifications $|\psi^{RQ}\rangle$ always exist. From our point of view R is introduced simply as a mathematical device to purify the initial state. The joint system RQ evolves according to the dynamics $\mathcal{I}_R \otimes \mathcal{E}$ given by

$$\rho^{R'Q'} = (\mathcal{I}_R \otimes \mathcal{E})(\rho^{RQ}), \tag{3.11}$$

where $\mathcal{I}_R$ is the identity dynamics for the reference system R.

The overall picture we have of a trace-preserving quantum operation is shown in Fig. 3.

The picture we have described thus far applies only to *trace-preserving* quantum operations. Later in the paper we will also be interested in quantum operations that are not trace preserving. That is, they do not satisfy the relation
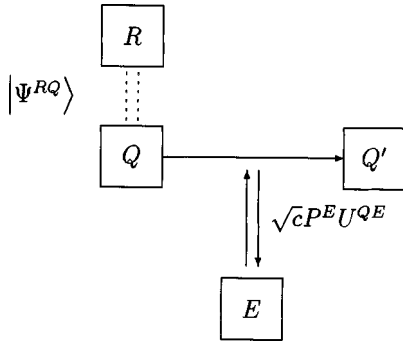
FIG. 4. Trace decreasing quantum operations.

$\Sigma_i A_i^\dagger A_i = I$, and thus $\text{tr}[\mathcal{E}(\rho)] \neq 1$ in general. Such quantum operations arise in the theory of *generalized measurements*. To each outcome m of a measurement there is an associated quantum operation $\mathcal{E}_m$ with an operator-sum representation,

$$\mathcal{E}_m(\rho) = \sum_i A_{mi} \rho A_{mi}^\dagger. \qquad (3.12)$$

The probability of obtaining outcome m is postulated to be

$$\Pr(m) = \text{tr}[\mathcal{E}_m(\rho)] = \text{tr}\left(\sum_i A_{mi}^\dagger A_{mi} \rho\right). \qquad (3.13)$$

The completeness relation for probabilities $\Sigma_m \Pr(m) = 1$ is equivalent to the completeness relation for the operators appearing in the operator-sum representations

$$\sum_{mi} A_{mi}^\dagger A_{mi} = I. \qquad (3.14)$$

Thus for each m,

$$\sum_i A_{mi}^\dagger A_{mi} \leq I. \qquad (3.15)$$

As an aside, it is interesting to note that the formulation of quantum measurement based on the projection postulate [16–18], taught in most classes on quantum mechanics, is a special case of the quantum operations formalism, obtainable by using a single projector $A_m = P_m$ in the operator-sum representation for $\mathcal{E}_m$. The formalism of positive operator valued measures (POVM's) [15] is also related to the generalized measurements formalism: $E_m \equiv \Sigma_i A_{mi}^\dagger A_{mi}$ are the elements of the POVM that is measured.

A result analogous to the earlier representation theorem for trace-preserving quantum operations can be proved for general operations.

Theorem (general representation theorem for operations). Suppose $\mathcal{E}$ is a general quantum operation. Then it is possible to find an environment E initially in a pure state $\sigma = |s\rangle\langle s|$ uncorrelated with the system, a unitary U^{QE}, a projector P^E onto the environment alone, and a constant $c > 0$, such that

$$\mathcal{E}(\rho) = c \, \text{tr}_E[P^E U^{QE}(\rho \otimes \sigma) U^{QE\dagger} P^E]. \qquad (3.16)$$

Furthermore, in the case of a generalized measurement described by operations $\mathcal{E}_m$ it is possible to do so in such a way

that for each m the corresponding constant $c_m = 1$, and the projectors P_m^E form a complete orthogonal set, $\Sigma_m P_m^E = I$, $P_m^E P_{m'}^E = \delta_{m,m'} P_m^E$.

Conversely, any map of the form (3.16) is a quantum operation.

Once again, introducing a reference system R that purifies ρ^Q we are left with a picture of the dynamics that looks like that shown in Fig. 4.

A few miscellaneous remarks will be useful later on.

(1) A prime always denotes a *normalized* state. For instance,

$$\rho^{R'Q'} = \frac{(\mathcal{I}_R \otimes \mathcal{E})(\rho^{RQ})}{\text{tr}[(\mathcal{I}_R \otimes \mathcal{E})(\rho^{RQ})]}. \qquad (3.17)$$

(2) The total state of the system RQE starts and remains pure. That is, $\rho^{R'Q'E'}$ is a pure state. Purity gives very useful relations among Von Neumann entropies $S(\rho) \equiv -\text{tr}(\rho \log_2 \rho)$, such as $S(\rho^{R'Q'}) = S(\rho^{E'})$ and all other permutations among R, Q and E. These are used frequently in what follows.

(3) Generically we denote quantum operations by $\mathcal{E}$ and the dimension of the quantum system Q by d.

(4) *Trace-preserving* quantum operations arise when a system interacts with an environment, and *no measurement* is performed on the system plus environment. Non-trace-preserving operations arise when classical information about the state of the system is made available by such a measurement. For most of this paper the noisy quantum channel is described by a trace-preserving quantum operation.

(5) Sometimes we consider the composition of two (or more) quantum operations. Generically we use the notation $\mathcal{E}_1, \mathcal{E}_2, \ldots$ for the different operations, and the notation $\mathcal{E}_2 \circ \mathcal{E}_1$ to denote composition of operations,

$$(\mathcal{E}_2 \circ \mathcal{E}_1)(\rho) \equiv \mathcal{E}_2(\mathcal{E}_1(\rho)). \qquad (3.18)$$

Furthermore it is sometimes useful to use the RQE picture of quantum operations to discuss compositions. We denote the environment corresponding to operation $\mathcal{E}_i$ by E_i, and assume environments corresponding to different values of i are independent and initially pure. So, for example, the initial state for a two-stage composition would be

$$\rho^{RQE_1E_2} = |\psi^{RQ}\rangle\langle\psi^{RQ}| \otimes |s_1\rangle\langle s_1| \otimes |s_2\rangle\langle s_2|. \qquad (3.19)$$

A single prime denotes the state of the system after the application of $\mathcal{E}_1$, and a double prime denotes the state of the system after the application of $\mathcal{E}_2 \circ \mathcal{E}_1$, and so on.

IV. ENTROPY EXCHANGE

This section briefly reviews the definition and some basic results about the *entropy exchange*, which was independently introduced by Schumacher [4] and Lloyd [7]. The entropy exchange turns out to be central to understanding the noisy quantum channel.

The *entropy exchange* of a quantum operation $\mathcal{E}$ with input ρ is defined to be

$$S_e(\rho, \mathcal{E}) \equiv S(\rho^{E'}), \qquad (4.1)$$

where $\rho^{E'}$ is the state of an initially pure environment (the "mock" environment of the previous section) after the operation, and $S(\rho) \equiv - \text{tr}(\rho \log_2 \rho)$ is the Von Neumann entropy. If $\mathcal{E}(\rho) = \Sigma_i A_i \rho A_i^\dagger$ then a convenient form for the entropy exchange is found by defining a matrix W with elements

$$W_{ij} \equiv \frac{\text{tr}(A_i \rho A_j^\dagger)}{\text{tr}[\mathcal{E}(\rho)]}. \tag{4.2}$$

It can be shown [4,14] that

$$S_e(\rho, \mathcal{E}) = S(W) \equiv - \text{tr}(W \log_2 W). \tag{4.3}$$

The last equation is frequently useful when performing calculations.

V. CLASSICAL NOISY CHANNELS IN A QUANTUM SETTING

In this section we show how classical noisy channels can be formulated in terms of quantum mechanics. We begin by reviewing the formulation in terms of classical information theory.

A classical noisy channel is described in terms of distinguishable channel states, which we label by x. If the input to the channel is symbol x then the output is symbol y with probability $p_{y|x}$. The channel is assumed to act independently on each input. For each x, the probability sum rule $\Sigma_y p_{y|x} = 1$ is satisfied. These *conditional probabilities* $p_{y|x}$ completely describe the classical noisy channel.

Suppose the input to the channel x is represented by some classical random variable X and the output by a random variable Y. The mutual information between X and Y is defined by

$$H(X:Y) \equiv H(X) + H(Y) - H(X,Y), \tag{5.1}$$

where $H(X)$ is the Shannon information of the random variable X defined by

$$H(X) \equiv - \sum_x p(x) \log_2 p(x), \tag{5.2}$$

with $0 \log_2 0 \equiv \lim_{p \to 0} p \log_2 p = 0$.

Shannon showed that the capacity of a noisy classical channel is given by the expression

$$C_S = \max_{p(x)} H(X:Y), \tag{5.3}$$

where the maximum is taken over all possible distributions $p(x)$ for the channel input X. Notice that although this is not an explicit expression for the channel capacity in terms of the conditional probabilities $p_{x|y}$, the maximization can easily be performed using well-known techniques from numerical mathematics. That is, Shannon's result provides an effective procedure for computing the capacity of a noisy classical channel.

All these results may be reexpressed in terms of quantum mechanics. We suppose the channel has some preferred orthonormal basis $|x\rangle$ of signal states. For convenience we assume the set of input states $|x\rangle$ is the same as the set of output states $|y\rangle$ of the channel, although more general schemes are possible. For the purpose of illustration the present level of generality suffices. A classical input random variable X corresponds to an input density operator for the quantum channel,

$$\rho_X \equiv \sum_x p(x) |x\rangle\langle x|. \tag{5.4}$$

The statistics of X are recoverable by measuring ρ_X in the $|x\rangle$ basis. Defining operators E_{xy} by

$$E_{xy} \equiv |y\rangle\langle x|, \tag{5.5}$$

we find that the channel operation defined by

$$\mathcal{N}(\rho) \equiv \sum_{xy} p_{y|x} E_{xy} \rho E_{xy}^\dagger \tag{5.6}$$

is a trace-preserving quantum operation, and that

$$\mathcal{N}(\rho_X) = \rho_Y = \sum_y p(y) |y\rangle\langle y|, \tag{5.7}$$

where ρ_Y is the density operator corresponding to the random variable Y that would have been obtained from X given a classical channel with probabilities $p_{y|x}$. This gives a quantum mechanical formalism for describing classical sources and channels. It is interesting to see what form the mutual information and channel capacity take in the quantum formalism.

Notice that

$$H(X) = S(\rho_X), \tag{5.8}$$

$$H(Y) = S(\rho_Y) = S(\mathcal{N}(\rho_X)). \tag{5.9}$$

Next we compute the entropy exchange associated with the channel operating on input ρ_X, by computing the W matrix given by Eq. (4.2). The W matrix corresponding to the channel with input ρ_X has entries

$$W_{(xy)(x'y')} = \delta_{x,x'} \delta_{y,y'} p(x) p(y|x), \tag{5.10}$$

but the joint distribution of (X,Y) satisfies $p(x)p(y|x) = p(x,y)$. Thus W is diagonal with eigenvalues $p(x,y)$, so the entropy exchange is given by

$$S_e(\rho_X, \mathcal{N}) = H(X,Y). \tag{5.11}$$

It follows that

$$H(X:Y) = S(\rho_X) + S(\mathcal{N}(\rho_X)) - S_e(\rho_X, \mathcal{N}), \tag{5.12}$$

and thus the Shannon capacity C_S of the classical channel is given in the quantum formalism by

$$C_S = \max_{\rho_X} [S(\rho_X) + S(\mathcal{N}(\rho_X)) - S_e(\rho_X, \mathcal{N})], \tag{5.13}$$

where the maximization is over all input states for the channel ρ_X that are diagonal in the $|x\rangle$ basis.

The problem we have been considering is that of transmitting a discrete set of orthogonal states (the states $|x\rangle$)

through the channel. In many quantum applications one is not only interested in transmitting a discrete set of states, but also arbitrary superpositions of those states. That is, one wants to transmit entire *subspaces* of states. In this case, the capacity of interest is the maximum rate of transmission of subspace dimensions. This may occur in quantum computing, cryptography, and teleportation. It is also interesting in these applications to transmit the *entanglement* of states. This cannot be done by considering the transmission of a set of orthogonal pure states alone.

It is not difficult to see that C_S is not correct as a measure of how many subspace dimensions may be reliably transmitted through a quantum channel. For example, consider the classical noiseless channel,

$$\mathcal{N}(\rho) = \sum_x |x\rangle\langle x|\rho|x\rangle\langle x|, \qquad (5.14)$$

where $|x\rangle$ is an orthonormal set of basis states for the channel. It is easily seen that

$$C_S = \log_2 d, \qquad (5.15)$$

where d is the number of channel dimensions. Yet it is intuitively clear, and is later proved in a more rigorous fashion, that such a channel cannot be used to transmit any nontrivial subspace of state space, nor can it be used to transmit any entanglement, and thus its capacity for transmitting these types of quantum resources is zero.

VI. ENTANGLEMENT FIDELITY

In this section we review a quantity known as the *entanglement fidelity* [4]. It is this quantity that we use to study the effectiveness of schemes for sending information through a noisy quantum channel.

The entanglement fidelity is defined for a *process*, specified by a quantum operation $\mathcal{E}$ acting on some initial state ρ. We denote it by $F_e(\rho,\mathcal{E})$. The concerns motivating the definition of the entanglement fidelity are twofold: (1) $F_e(\rho,\mathcal{E})$ measures how well the *state* ρ is preserved by the operation $\mathcal{E}$. An entanglement fidelity close to one indicates that the process preserves the state well. (2) $F_e(\rho,\mathcal{E})$ measures how well the *entanglement* of ρ with other systems is preserved by the operation $\mathcal{E}$. An entanglement fidelity close to one indicates the process preserves the entanglement well.

Conversely, an entanglement fidelity close to zero indicates that the state or its entanglement were not well preserved by the operation $\mathcal{E}$.

Formally, the entanglement fidelity is defined by

$$F_e(\rho,\mathcal{E}) \equiv \langle\psi^{RQ}|(\mathcal{I}_R\otimes\mathcal{E})(|\psi^{RQ}\rangle\langle\psi^{RQ}|)|\psi^{RQ}\rangle. \quad (6.1)$$

That is, the entanglement fidelity is the overlap between the initial purification $|\psi^{RQ}\rangle$ of the state *before* it is sent through the channel with the state of the joint system RQ *after* it has been sent through the channel. The entanglement fidelity depends only on ρ and $\mathcal{E}$, not on the particular purification $|\psi^{RQ}\rangle$ of ρ that is used [4]. If $\mathcal{E}$ has operation elements $\{A_i\}$ then the entanglement fidelity has the expression [4,14]

$$F_e(\rho,\mathcal{E}) = \frac{\sum_i |\operatorname{tr}(A_i\rho)|^2}{\operatorname{tr}[\mathcal{E}(\rho)]}. \qquad (6.2)$$

This expression simplifies for trace-preserving quantum operations since the denominator is 1. The entanglement fidelity has the following properties [4,5,14]: (1) $0 \leq F_e(\rho,\mathcal{E}) \leq 1$. (2) $F_e(\rho,\mathcal{E})=1$ if and only if for all pure states $|\psi\rangle$ lying in the support of ρ,

$$\mathcal{E}(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi|. \qquad (6.3)$$

(3) The entanglement fidelity is a lower bound on the fidelity defined by Jozsa [19] in the following sense:

$$F_e(\rho,\mathcal{E}) \leq F(\rho,\mathcal{E}(\rho)). \qquad (6.4)$$

(4) Suppose $\{|\psi_i\rangle, p_i\}$ is an ensemble realizing ρ,

$$\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|. \qquad (6.5)$$

Then the entanglement fidelity is a lower bound on the average fidelity for the pure states $|\psi_i\rangle$,

$$F_e(\rho,\mathcal{E}) \leq \sum_i p_i\langle\psi_i|\mathcal{E}(|\psi_i\rangle\langle\psi_i|)|\psi_i\rangle. \qquad (6.6)$$

(5) Again suppose $\{|\psi_i\rangle, p_i\}$ is an ensemble realizing ρ. Then if the pure-state fidelity $\langle\psi|\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle \geq 1-\eta$ for all $|\psi\rangle$ in the support of ρ, $F_e(\rho,\mathcal{E}) \geq 1-(3/2)\eta$ (Knill and Laflamme [20]).

There are several reasons for using the entanglement fidelity as our measure of success in transmitting quantum states. If we succeed in sending a source ρ_s with high entanglement fidelity, we can send *any* ensemble for ρ_s with high average pure-state fidelity, by item (4) above. Entanglement fidelity is thus a more severe requirement of quantum coherence than average pure-state fidelity. Moreover, the ability to preserve entanglement is of great importance in applications of quantum coding to, say, quantum computation, where one would like to be able to apply error correction in a modular fashion to small portions of a quantum computer despite the fact that they may, in the course of quantum computation, become entangled with other parts of the computer [21]. (Of course, the general problem of finding the capacity of a noisy quantum channel for a *given* ensemble with average pure-state fidelity as the reliability measure is also worth investigating.)

An appropriate measure of how well a *subspace* of quantum states is transmitted is the *subspace fidelity*,

$$F_s(P,\mathcal{E}) \equiv \min_{|\psi\rangle}\langle\psi|\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle, \qquad (6.7)$$

where the minimization is over all pure states $|\psi\rangle$ in the subspace whose projector is P. Item (5) above implies that if the subspace fidelity is close to one, the entanglement fidelity is also close to one. The converse is not in general true. That is, reliable transmission of subspaces is a more stringent requirement than transmission of entanglement. Therefore using entanglement fidelity as our criterion for reliable transmission yields capacities at least as great as those obtained

when subspace fidelity is used as the criterion. We conjecture that these two capacities are identical.

As an alternative measure of subspace fidelity, one might consider the average pure-state fidelity,

$$\int d|\psi\rangle\langle\psi|\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle, \tag{6.8}$$

where the integration is done using the unitarily invariant measure on the subspace of interest. By item (4) above, the capacity resulting from this measure of reliability is at least as great as that which results when entanglement fidelity is used as the measure of reliability. We do not know whether these two capacities are equal.

The lesson to be learned from this discussion is that there are many different measures that may be used to quantify how reliably quantum states are transmitted, and different measures may result in different capacities. Which measure is used depends on what resource is most important for the application of interest. For the remainder of this paper, we use the entanglement fidelity as our measure of reliability.

There is a very useful inequality, the *quantum Fano inequality*, which relates the entropy exchange and the entanglement fidelity. It is [4]

$$S_e(\rho,\mathcal{E}) \leq h(F_e(\rho,\mathcal{E})) + [1 - F_e(\rho,\mathcal{E})]\log_2(d^2 - 1), \tag{6.9}$$

where $h(p) \equiv -p\log_2 p - (1-p)\log_2(1-p)$ is the dyadic Shannon information associated with p. It is useful to note for our later work that $0 \leq h(p) \leq 1$ and $\log_2(d^2 - 1) \leq 2\log_2 d$, so from the quantum Fano inequality,

$$S_e(\rho,\mathcal{E}) \leq 1 + 2[1 - F_e(\rho,\mathcal{E})]\log_2 d. \tag{6.10}$$

The proof of the quantum Fano inequality (6.9) is simple enough that for convenience we repeat it here. Consider an orthonormal set of d^2 basis states $|\psi_i\rangle$ for the system RQ. This basis set is chosen so that $|\psi_1\rangle = |\psi^{RQ}\rangle$. If we form the quantities $p_i \equiv \langle\psi_i|\rho^{R'Q'}|\psi_i\rangle$, then it is possible to show (see, for example, [22], p. 240)

$$S(\rho^{R'Q'}) \leq H(p_1, \ldots, p_{d^2}), \tag{6.11}$$

where $H(p_i)$ is the Shannon information of the set p_i. But by easily verified grouping properties of the Shannon entropy,

$$H(p_1, \ldots, p_{d^2}) = h(p_1) + (1 - p_1)H\left(\frac{p_2}{1 - p_1}, \ldots, \frac{p_{d^2}}{1 - p_1}\right), \tag{6.12}$$

and it is easy to show that

$$H\left(\frac{p_2}{1 - p_1}, \ldots, \frac{p_{d^2}}{1 - p_1}\right) \leq \log(d^2 - 1).$$

Combining these results and noting that $p_1 = F_e(\rho,\mathcal{E})$ by definition of the entanglement fidelity,

$$S_e(\rho,\mathcal{E}) \leq h(F_e(\rho,\mathcal{E})) + [1 - F_e(\rho,\mathcal{E})]\log_2(d^2 - 1), \tag{6.13}$$

which is the quantum Fano inequality.

For applications it is useful to understand the continuity properties of the entanglement fidelity. To that end we prove the following lemma.

Lemma (continuity lemma for entanglement fidelity). Suppose $\mathcal{E}$ is a trace-preserving quantum operation, ρ is a density operator, and Δ is a Hermitian operator with trace zero. Then

$$|F_e(\rho + \Delta,\mathcal{E}) - F_e(\rho,\mathcal{E})| \leq 2\,\mathrm{tr}(|\Delta|) + \mathrm{tr}(|\Delta|)^2. \tag{6.14}$$

To prove the lemma we apply Eq. (6.2) to obtain

$$|F_e(\rho + \Delta,\mathcal{E}) - F_e(\rho,\mathcal{E})| \leq 2\sum_i |\,\mathrm{tr}(A_i\rho)|\,|\,\mathrm{tr}(A_i^\dagger\Delta)|$$

$$+ \sum_i |\,\mathrm{tr}(A_i\Delta)|^2. \tag{6.15}$$

Applying a Cauchy-Schwarz inequality to each sum, the first with respect to the complex inner product $\Sigma_i x_i^* y_i$, the second with respect to the Hilbert-Schmidt inner product $\mathrm{tr}(X^\dagger Y)$, we obtain

$$|F_e(\rho + \Delta,\mathcal{E}) - F_e(\rho,\mathcal{E})|$$

$$\leq 2\left(\sum_i |\,\mathrm{tr}(A_i\rho)|^2 \sum_j |\,\mathrm{tr}(A_j^\dagger\Delta)|^2\right)^{1/2}$$

$$+ \sum_i |\,\mathrm{tr}(A_i|\Delta|A_i^\dagger)|\,|\,\mathrm{tr}(|\Delta|)|, \tag{6.16}$$

where $|\Delta| \equiv \sqrt{\Delta^\dagger\Delta}$. Applying Eq. (6.2) and $F_e(\rho,\mathcal{E}) \leq 1$ to the first sum and the trace-preserving property of $\mathcal{E}$ to the final sum gives

$$|F_e(\rho + \Delta,\mathcal{E}) - F_e(\rho,\mathcal{E})| \leq 2\sqrt{\sum_j |\,\mathrm{tr}(A_j^\dagger\Delta)|^2} + \mathrm{tr}(|\Delta|)^2. \tag{6.17}$$

One final application of the Cauchy-Schwarz inequality and the trace-preserving property of $\mathcal{E}$ gives

$$|F_e(\rho + \Delta,\mathcal{E}) - F_e(\rho,\mathcal{E})| \leq 2\,\mathrm{tr}(|\Delta|) + \mathrm{tr}(|\Delta|)^2, \tag{6.18}$$

as required.

This result gives bounds on the change in the entanglement fidelity when the input state is perturbed. Note, incidentally, that during the proof a coefficient $\sqrt{F_e(\rho,\mathcal{E})}$ was dropped from the first term on the right-hand side of the inequality. For some applications it may be useful to apply the inequality with this coefficient in place.

VII. COHERENT INFORMATION

In this section we investigate the *coherent information*. The coherent information was defined in [5], where it was suggested that the coherent information plays a role in quantum information theory analogous to the role played by mutual information in classical information theory in the following sense. Consider a classical random process,

$$\overset{\mathcal{M}}{X \rightarrow Y}, \qquad (7.1)$$

in which the random variable X is used as the input to some process that produces as output the random variable Y. The distributions of X and Y are related by a linear map $\mathcal{M}$ determined by the conditional probabilities of the process. An example of such a process is a noisy classical channel with input X and output Y. As discussed earlier, an important quantity in information theory is the mutual information $H(X:Y)$ between the input X and the output Y of the process. Note that $H(X:Y)$ can be regarded as a function of the input X and the map $\mathcal{M}$ only, since the joint distribution of X and Y is determined by these.

Quantum mechanically we can consider a process defined by an input ρ, and output ρ', with the process described by a quantum operation $\mathcal{E}$,

$$\overset{\mathcal{E}}{\rho \rightarrow \rho' = \mathcal{E}(\rho)}. \qquad (7.2)$$

We assert that the coherent information, defined by

$$I(\rho, \mathcal{E}) \equiv S\left(\frac{\mathcal{E}(\rho)}{\text{tr}[\mathcal{E}(\rho)]}\right) - S_e(\rho, \mathcal{E}), \qquad (7.3)$$

plays a role in quantum information theory analogous to that played by the mutual information $H(X:Y)$ in classical information theory. This is not obvious from the definition, and one goal of this section is to make it appear plausible that this is the case. Of course, the true justification for regarding the coherent information as the quantum analogue of the mutual information is its success as the quantity appearing in results on channel capacity, as discussed in later sections. This is the true motivation for all definitions in information theory, whether classical or quantum: their success at quantifying the resources needed to perform some interesting physical task, not some abstract mathematical motivation.

In Sec. VII A we review the data-processing inequality that provides motivation for regarding the coherent information as a quantum analogue of the mutual information, and whose application is crucial to later reasoning. Section VII B studies in detail the properties of the coherent information. In particular, we prove several results related to convexity that are useful both as calculational aids, and also for proving later results. Section VII C proves a lemma about the entanglement fidelity that glues together many of our later proofs of upper bounds on the channel capacity. Finally, Secs. VII D and VII E describe two important ways that the behavior of the coherent information differs from the behavior of the mutual information when quantum entanglement is allowed.

A. Quantum data-processing inequality

The role of coherent information in quantum information theory is intended to be similar to that of mutual information in classical information theory. This is not obvious from the definition, but can be given an operational motivation in terms of a procedure known as *data processing*. The classical data-processing inequality [3] states that any three variable Markov process,

$$X \rightarrow Y \rightarrow Z, \qquad (7.4)$$

satisfies a data-processing inequality,

$$H(X) \geqslant H(X:Y) \geqslant H(X:Z). \qquad (7.5)$$

The idea is that the operation $Y \rightarrow Z$ represents some kind of ''data processing'' of Y to obtain Z, and the mutual information after processing, $H(X:Z)$, can be no higher than the mutual information before processing, $H(X:Y)$. Furthermore, suppose we have a Markov process,

$$X \rightarrow Y, \qquad (7.6)$$

such that $H(X) = H(X:Y)$. Intuitively, one might expect that it should be possible to do data processing on Y to recover X. It is not difficult to show that it is possible, using Y alone, to construct a third variable Z forming a third stage in the Markov process,

$$X \rightarrow Y \rightarrow Z, \qquad (7.7)$$

such that $X = Z$ with probability 1, if and only if $H(X) = H(X:Y)$.

An analogous quantum result has been proved by Schumacher and Nielsen [5]. It states that given trace-preserving quantum operations $\mathcal{E}_1$ and $\mathcal{E}_2$ defining a quantum process,

$$\rho \rightarrow \mathcal{E}_1(\rho) \rightarrow (\mathcal{E}_2 \circ \mathcal{E}_1)(\rho), \qquad (7.8)$$

then

$$S(\rho) \geqslant I(\rho, \mathcal{E}_1) \geqslant I(\rho, \mathcal{E}_2 \circ \mathcal{E}_1). \qquad (7.9)$$

Furthermore, it was shown in [5] that given a process

$$\rho \rightarrow \mathcal{E}_1(\rho), \qquad (7.10)$$

it is possible to find an operation $\mathcal{E}_2$ that reverses $\mathcal{E}_1$ if and only if

$$S(\rho) = I(\rho, \mathcal{E}_1). \qquad (7.11)$$

The close analogy between the classical and quantum data-processing inequalities provides a strong operational motivation for considering the coherent information to be the quantum analogue of the classical mutual information.

The proof of the quantum data-processing inequality is repeated here in order to address the issue of what happens when $\mathcal{E}_1$ and $\mathcal{E}_2$ are not trace preserving. The proof of the first inequality is to apply the subadditivity inequality [22] $S(\rho^{R'E'}) \leqslant S(\rho^{R'}) + S(\rho^{E'})$ in the RQE picture of operations to obtain

$$I(\rho, \mathcal{E}_1) = S[\mathcal{E}_1(\rho)] - S_e(\rho, \mathcal{E}_1) \qquad (7.12)$$

$$= S(\rho^{Q'}) - S(\rho^{E'}) \qquad (7.13)$$

$$= S(\rho^{R'E'}) - S(\rho^{E'}) \qquad (7.14)$$

$$\leqslant S(\rho^{R'}) = S(\rho^R) = S(\rho). \qquad (7.15)$$

It is clear that this part of the inequality need not hold when $\mathcal{E}_1$ is not trace preserving. The reason for this is that it is no longer necessarily the case that $\rho^{R'} = \rho^R$, and thus it may not be possible to make the identification $S(\rho^{R'}) = S(\rho^R)$. For example, suppose we have a three-dimensional state space with orthonormal states $|1\rangle$, $|2\rangle$, and $|3\rangle$. Let P_{12} be the projector onto the two-dimensional subspace spanned by $|1\rangle$ and $|2\rangle$, and P_3 the projector onto the subspace spanned by $|3\rangle$. Let $\rho = (p/2)P_{12} + (1-p)P_3$, where $0 < p < 1$, and $\mathcal{E}(\rho) = P_{12}\rho P_{12}$. Then by choosing p small enough we can make $S(\rho) \approx 0$, but $I(\rho,\mathcal{E}) = 1$, so we have an example of a non-trace-preserving operation that does not obey the data-processing inequality.

The proof of the second part of the data-processing inequality is to apply the strong subadditivity inequality [22],

$$S(\rho^{R''E_1''E_2''}) + S(\rho^{E_1''}) \leq S(\rho^{R''E_1''}) + S(\rho^{E_1''E_2''}), \quad (7.16)$$

where we are now using an RQE_1E_2 picture of the operations. From purity of the total state of RQE_1E_2 it follows that

$$S(\rho^{R''E_1''E_2''}) = S(\rho^{Q''}). \quad (7.17)$$

Neither of the systems R or E_1 are involved in the second stage of the dynamics in which Q and E_2 interact unitarily. Thus, their state does not change during this stage: $\rho^{R''E_1''} = \rho^{R'E_1'}$. But from the purity of RQE_1 after the first stage of the dynamics,

$$S(\rho^{R''E_1''}) = S(\rho^{R'E_1'}) = S(\rho^{Q'}). \quad (7.18)$$

The remaining two terms in the subadditivity inequality are now recognized as entropy exchanges,

$$S(\rho^{E_1''}) = S(\rho^{E_1'}) = S_e(\rho,\mathcal{E}_1), \quad (7.19)$$

$$S(\rho^{E_1''E_2''}) = S_e(\rho,\mathcal{E}_2 \circ \mathcal{E}_1). \quad (7.20)$$

Making these substitutions into the inequality obtained from strong subadditivity (7.16) yields

$$S(\rho^{Q''}) + S_e(\rho,\mathcal{E}_1) \leq S(\rho^{Q'}) + S_e(\rho,\mathcal{E}_2 \circ \mathcal{E}_1), \quad (7.21)$$

which can be rewritten as the second stage of the data-processing inequality,

$$I(\rho,\mathcal{E}_1) \geq I(\rho,\mathcal{E}_2 \circ \mathcal{E}_1). \quad (7.22)$$

Notice that this inequality holds provided $\mathcal{E}_2$ is trace preserving, and does not require any assumption that $\mathcal{E}_1$ is trace preserving. This is very useful in our later work.

B. Properties of coherent information

The set of completely positive maps forms a positive cone, that is, if $\mathcal{E}_i$ is a collection of completely positive maps and λ_i is a set of non-negative numbers then $\Sigma_i \lambda_i \mathcal{E}_i$ is also a completely positive map. In this section we prove two very useful properties of the coherent information. First, it is easy to see that for any quantum operation $\mathcal{E}$ and non-negative λ,

$$I(\rho,\lambda\mathcal{E}) = I(\rho,\mathcal{E}). \quad (7.23)$$

This follows immediately from the definition of the coherent information. A slightly more difficult property to prove is the following.

Theorem (generalized convexity theorem for coherent information). Suppose $\mathcal{E}_i$ are quantum operations. Then

$$I\left(\rho,\sum_i \mathcal{E}_i\right) \leq \frac{\Sigma_i \, \mathrm{tr}[\mathcal{E}_i(\rho)]I(\rho,\mathcal{E}_i)}{\mathrm{tr}[\Sigma_i \mathcal{E}_i(\rho)]}. \quad (7.24)$$

This result is extremely useful in our later work. An important and immediate corollary is the following.

Corollary (convexity theorem for coherent information). If a trace-preserving operation $\mathcal{E} = \Sigma_i p_i \mathcal{E}_i$ is a convex sum $(p_i \geq 0, \Sigma_i p_i = 1)$ of trace-preserving operations $\mathcal{E}_i$, then the coherent information is convex,

$$I\left(\rho,\sum_i p_i\mathcal{E}_i\right) \leq \sum_i p_i I(\rho,\mathcal{E}_i). \quad (7.25)$$

The proof of the corollary is immediate from the theorem. The theorem follows from the concavity of the *conditional* entropy (see references cited in [22], pages 249–250), which for two systems 1 and 2 is defined by

$$S(2|1) \equiv S(\rho_{12}) - S(\mathrm{tr}_2(\rho_{12})). \quad (7.26)$$

This expression is concave in ρ_{12}. Now notice that

$$I(\rho,\mathcal{E}) = S(\rho^{Q'}) - S(\rho^{R'Q'}) = -S(R'|Q'). \quad (7.27)$$

The theorem now follows from the concavity of the conditional entropy.

A further useful result concerns the additivity of coherent information.

Theorem (additivity for independent channels). Suppose $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are quantum operations and $\rho_1, \ldots, \rho_n$ are density operators. Then

$$I(\rho_1 \otimes \cdots \otimes \rho_n, \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n) = \sum_i I(\rho_i,\mathcal{E}_i). \quad (7.28)$$

The proof is immediate from the additivity property of entropies for product states.

C. A lemma about entanglement fidelity

The following lemma is the glue that holds together much of our later work on proving upper bounds to channel capacities. In this section we prove the lemma only for the special case of trace-preserving operations. A similar but more complicated result is true for general operations, and is given in Sec. XI.

We begin by repeating the proof of a simple inequality that was first proved in [4], which states that the decrease (if any) in system entropy must be bounded above by the increase in the entropy of a pure environment. This applies only for trace-preserving operations $\mathcal{E}$. Applying the subadditivity inequality [22] $S(\rho^{Q'E'}) \leq S(\rho^{Q'}) + S(\rho^{E'})$ and the relationship $S(\rho^{R'}) = S(\rho^{Q'E'})$, that follows from purity we obtain

$$S(\rho) = S(\rho^R), \tag{7.29}$$

$$= S(\rho^{R'}), \tag{7.30}$$

$$= S(\rho^{Q'E'}), \tag{7.31}$$

$$\leq S(\rho^{Q'}) + S(\rho^{E'}). \tag{7.32}$$

Rewriting this slightly gives

$$S(\rho) - S(\mathcal{E}(\rho)) \leq S_e(\rho, \mathcal{E}), \tag{7.33}$$

for any trace-preserving quantum operation $\mathcal{E}$.

Lemma (entanglement fidelity lemma for operations). Suppose $\mathcal{E}$ is a trace-preserving quantum operation, and ρ is some quantum state. Then for all trace-preserving quantum operations $\mathcal{D}$,

$$S(\rho) \leq I(\rho, \mathcal{E}) + 2 + 4[1 - F_e(\rho, \mathcal{D} \circ \mathcal{E})] \log_2 d. \tag{7.34}$$

This lemma is extremely useful in obtaining proofs of bounds on the channel capacity. In order for the entanglement fidelity to be close to 1, the quantity appearing on the right-hand side must be close to zero. This shows that the entropy of ρ cannot greatly exceed the coherent information $I(\rho, \mathcal{E})$ if the entanglement fidelity is to be close to 1.

To prove the lemma, notice that by the second part of the data-processing inequality (7.9),

$$S(\rho) - I(\rho, \mathcal{E}) \leq S(\rho) - S((\mathcal{D} \circ \mathcal{E})(\rho)) + S_e(\rho, \mathcal{D} \circ \mathcal{E}). \tag{7.35}$$

Applying inequality (7.33) gives

$$S(\rho) - S((\mathcal{D} \circ \mathcal{E})(\rho)) \leq S_e(\rho, \mathcal{D} \circ \mathcal{E}), \tag{7.36}$$

and combining the last two inequalities gives

$$S(\rho) - I(\rho, \mathcal{E}) \leq 2 S_e(\rho, \mathcal{D} \circ \mathcal{E}) \tag{7.37}$$

$$\leq 2h[F_e(\rho, \mathcal{D} \circ \mathcal{E})] + 2(1 - F_e(\rho, \mathcal{D} \circ \mathcal{E})) \log_2 (d^2 - 1), \tag{7.38}$$

where the second step follows from the quantum Fano inequality (6.9). But the dyadic Shannon entropy h is bounded above by 1 and $\log_2 (d^2 - 1) \leq 2 \log_2 d$, so

$$S(\rho) \leq I(\rho, \mathcal{E}) + 2 + 4[1 - F_e(\rho, \mathcal{D} \circ \mathcal{E})] \log_2 d. \tag{7.39}$$

This completes the proof.

This inequality is strong enough to prove the asymptotic bounds that are of most interest for our later work. The somewhat stronger inequality (7.38) is also useful when proving one-shot results, that is, when no block coding is being used.

D. Quantum characteristics of the coherent information I

There are at least two important respects in which the coherent information behaves differently from the classical mutual information. In this subsection and the next we explain what these differences are.

Classically, suppose we have a Markov process,

$$X \to Y \to Z. \tag{7.40}$$

Intuitively we expect that

$$H(X:Z) \leq H(Y:Z), \tag{7.41}$$

and, indeed, it is not difficult to prove such a "pipelining inequality," based on the definition of the mutual information. The idea is that any information about X that reaches Z must go through Y, and therefore is also information that Z has about Y. However, the quantum mechanical analogue of this result fails to hold. We shall see that the reason it fails is due to quantum entanglement.

Example 1. Suppose we have a two-part quantum process described by quantum operations $\mathcal{E}_1$ and $\mathcal{E}2$:

$$\rho \to \mathcal{E}_1(\rho) \to (\mathcal{E}_2 \circ \mathcal{E}_1)(\rho). \tag{7.42}$$

Then, in general

$$I(\rho, \mathcal{E}_2 \circ \mathcal{E}_1) \leq I(\mathcal{E}_1(\rho), \mathcal{E}_2). \tag{7.43}$$

An explicit example showing that this is the case is given below. It is not possible to prove a general inequality of this sort for the coherent information—examples may be found where a $<, >$ or $=$ sign could occur in the last equation. We now show how the purely quantum mechanical effect of entanglement is responsible for this property of coherent information.

Notice that the truth of the equation

$$I(\rho, \mathcal{E}_2 \circ \mathcal{E}_1) \leq I(\mathcal{E}_1(\rho), \mathcal{E}_2) \tag{7.44}$$

is equivalent to

$$S_e(\mathcal{E}_1(\rho), \mathcal{E}_2) \leq S_e(\rho, \mathcal{E}_2 \circ \mathcal{E}_1). \tag{7.45}$$

This last equation makes it easy to see why Eq. (7.44) may fail. It is because the entropy of the joint environment for processes $\mathcal{E}_1$ and $\mathcal{E}_2$ (the quantity on the right-hand side) may be less than the entropy of the environment for process $\mathcal{E}_2$ alone (the quantity on the left). This is a property peculiar to quantum mechanics, which is caused by entanglement; there is no classical analogue. An example of this type of phenomenon is provided by an EPR pair, where the entropy of either system alone (one bit) is greater than that of the entire system, which is pure and thus has zero bits of entropy.

An example of Eq. (7.43) is as follows. For convenience we use the language of coding and channel operations, since that language is most convenient later. $\mathcal{E}_1$ is to be identified with the coding operation, $\mathcal{C}$, and $\mathcal{E}_2$ is to be identified with the channel operation, $\mathcal{N}$.

Suppose we have a four-dimensional state space. We suppose that we have an orthonormal basis $|1\rangle, |2\rangle, |3\rangle, |4\rangle$, and that P_{12} is the projector onto the space spanned by $|1\rangle$ and $|2\rangle$, and P_{34} is the projector onto the space spanned by $|3\rangle$ and $|4\rangle$. Let U be a unitary operator defined by

$$U \equiv |3\rangle\langle 1| + |4\rangle\langle 2| + |1\rangle\langle 3| + |2\rangle\langle 4|. \tag{7.46}$$

The channel operation is

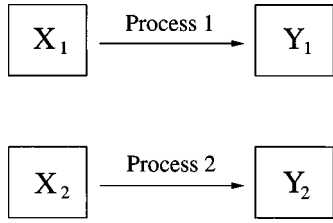$$\mathcal{N}(\rho) = P_{12} \rho P_{12} + U^\dagger P_{34} \rho P_{34} U, \tag{7.47}$$

FIG. 5. Dual classical channels operating on inputs X_1 and X_2 produce outputs Y_1 and Y_2.

and we use an encoding defined by

$$\mathcal{C}(\rho) = \tfrac{1}{2} P_{12}\rho P_{12} + \tfrac{1}{2} U P_{12}\rho P_{12} U^{\dagger} + P_{34}\rho P_{34}. \quad (7.48)$$

It is easily checked that for any state ρ whose support lies wholly in the space spanned by $|1\rangle$ and $|2\rangle$,

$$(\mathcal{N} \circ \mathcal{C})(\rho) = \rho. \quad (7.49)$$

It follows that

$$I(\rho, \mathcal{N} \circ \mathcal{C}) = S(\rho). \quad (7.50)$$

It is also easy to verify that

$$I(\mathcal{C}(\rho), \mathcal{N}) = 2S(\rho) - 1. \quad (7.51)$$

Thus there exist states ρ such that

$$I(\rho, \mathcal{N} \circ \mathcal{C}) > I(\mathcal{C}(\rho), \mathcal{N}), \quad (7.52)$$

providing an example of Eq. (7.43).

E. Quantum characteristics of the coherent information II

The second important difference between coherent information and classical mutual information is related to the property known classically as *subadditivity of mutual information*. Suppose we have several independent channels operating. Figure 5 shows the case of two channels.

These channels are numbered $1, \ldots, n$ and take as inputs random variables $X_1, \ldots, X_n$. The channels might be separated spatially, as shown in the figure, or in time. The channels are assumed to act independently on their respective inputs, and produce outputs $Y_1, \ldots, Y_n$. It is not difficult to show that

$$H(X_1, \ldots, X_n : Y_1, \ldots, Y_n) \le \sum_i H(X_i : Y_i). \quad (7.53)$$

This property is known as the *subadditivity* of mutual information. It is used, for example, in proofs of the weak converse to Shannon's noisy-channel coding theorem. We now show that the corresponding quantum statement about coherent information fails to hold.

Example 2. There exists a quantum operation $\mathcal{E}$ and a density operator ρ_{12} such that

$$I(\rho_{12}, \mathcal{E} \otimes \mathcal{E}) \nleq I(\rho_1, \mathcal{E}) + I(\rho_2, \mathcal{E}), \quad (7.54)$$

where $\rho_1 \equiv \mathrm{tr}_2(\rho_{12})$ and $\rho_2 \equiv \mathrm{tr}_1(\rho_{12})$ are the usual reduced density operators for systems 1 and 2.
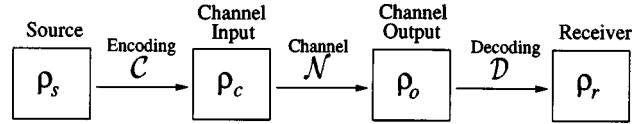

FIG. 6. The noisy quantum channel, together with encodings and decodings.

An example of Eq. (7.54) is the following. Suppose system 1 consists of two qubits, A and B. System 2 consists of two more qubits, C and D. As the initial state we choose

$$\rho_{12} = \frac{I_A}{2} \otimes |\psi^{BD}\rangle\langle\psi^{BD}| \otimes \frac{I_C}{2}, \quad (7.55)$$

where $|\psi^{BD}\rangle$ is a Bell state shared between systems B and D.

The action of the channel on A and B is as follows: it sets bit B to some standard state, $|0\rangle$, and allows A through unchanged. This is achieved by swapping the state of B out into the environment. Formally,

$$\mathcal{E}(\rho_{AB}) = \rho_A \otimes |0\rangle\langle 0|. \quad (7.56)$$

The same channel is now set to act on systems C and D:

$$\mathcal{E}(\rho_{CD}) = \rho_C \otimes |0\rangle\langle 0|. \quad (7.57)$$

A straightforward though slightly tedious calculation shows that with this channel setup

$$I(\rho_1, \mathcal{E}) = I(\rho_2, \mathcal{E}) = 0, \quad (7.58)$$

and

$$I(\rho_{12}, \mathcal{E} \otimes \mathcal{E}) = 2. \quad (7.59)$$

Thus this setup provides an example of Eq. (7.54).

VIII. NOISY-CHANNEL CODING REVISITED

In this section we return to noisy-channel coding. Recall the basic procedure for noisy channel coding, as illustrated in Fig. 6.

Suppose a quantum source has output ρ_s. A quantum operation, which we shall denote $\mathcal{C}$, is used to *encode* the source, giving the input state to the channel, $\rho_c \equiv \mathcal{C}(\rho_s)$. The encoded state is used as input to the noisy channel, giving a channel output $\rho_o \equiv \mathcal{N}(\rho_c)$. Finally, a decoding quantum operation $\mathcal{D}$ is used to decode the output of the channel, giving a *received state*, $\rho_r \equiv \mathcal{D}(\rho_o)$. The goal of noisy-channel coding is to find out what source states can be sent with high entanglement fidelity. That is, we want to know for what states ρ_s encoding and decoding operations can be found such that

$$F_e(\rho_s, \mathcal{D} \circ \mathcal{N} \circ \mathcal{C}) \approx 1. \quad (8.1)$$

If large blocks of source states with entropy R per use of the channel can be sent through the channel with high fidelity, we say the channel is transmitting at the rate R.

Shannon's noisy-channel coding theorem is an example of a *channel capacity* theorem. Such theorems come in two parts: (1) An *upper bound* is placed on the rate at which information can be sent reliably through the channel. There

should be an effective procedure for calculating this upper bound. (2) It is proved that a reliable scheme for encoding and decoding exists that comes arbitrarily close to *attaining* the upper bound found in (1).

This maximum rate at which information can be reliably sent through the channel is known as the *channel capacity*.

Channel capacity results may be understood in the language of error correcting codes (see [23] for a review of the classical theory of error correcting codes, or [24] for a review and many references for the quantum theory). In order to protect information against the effects of noise, it is encoded using an error correcting code, with the encoding operation represented by $\mathcal{C}$, then subjected to the noise, represented by $\mathcal{N}$, and finally the encoding is undone using the decoding operation $\mathcal{D}$. Finding a good error correcting code means finding a pair $\mathcal{C}$ and $\mathcal{D}$ that preserves the information being encoded. A channel capacity theorem places an ultimate achievable limit on the effectiveness of these error correcting codes, for a given noise model $\mathcal{N}$.

In this paper we consider only the first of these two tasks, the placing of upper bounds on the rate at which quantum information can be reliably sent through a noisy quantum channel. The results we prove are analogous to the weak converse of the classical noisy coding theorem, but cannot be considered true converses until attainability of our bounds is demonstrated. We do consider it likely that our bounds are equal to the true quantum channel capacity.

A. Mathematical formulation of noisy-channel coding

Up to this point the procedure for doing noisy-channel coding has been discussed in broad outline but we have not made all of our definitions mathematically precise. This subsection gives a precise formulation for the most important concepts appearing in our work on noisy-channel coding.

Define a *quantum source* $\Sigma = (H_s, Y)$ to consist of a Hilbert space H_s and a sequence $Y = \{\rho_s^1, \rho_s^2, \ldots, \rho_s^n, \ldots\}$ where ρ_s^1 is a density operator on H_s, ρ_s^2 a density operator on $H_s \otimes H_s$, and ρ_s^n a density operator on $H_s^{\otimes n}$, etc . . . . Using, for example, ''tr_{34}'' to denote the partial trace over the third and fourth copies of H_s, we require as part of our definition of a quantum source that for all j and all $n > j$,

$$\mathrm{tr}_{j+1,\ldots,n}(\rho_s^n) = \rho_s^j, \qquad (8.2)$$

i.e., that density operators in the sequence be consistent with each other in the sense that earlier ones be derivable from later ones by an appropriate partial trace. The nth density operator is meant to represent the state of n emissions from the source, normally thought of as taking n units of time. (We could have used a single density operator on a countably infinite tensor product of spaces H_s, but we wish to avoid the technical issues associated with such products.) We define the *entropy rate* of a general source Σ as

$$S(\Sigma) \equiv \lim_{n \to \infty} \sup \frac{S(\rho_s^n)}{n}. \qquad (8.3)$$

A special case of this general definition of a quantum source is the i.i.d. source $(H_s, \{\rho_s, \rho_s \otimes \rho_s, \ldots, \rho_s^{\otimes n}, \ldots\})$, for some fixed ρ_s. Such a source corresponds to the classical

notion of an *independent, identically distributed* classical source, thus the term i.i.d. The entropy rate of this source is simply $S(\rho_s)$.

A *discrete memoryless channel* $(H_c, \mathcal{N})$ consists of a finite-dimensional Hilbert space H_c and a trace-preserving quantum operation $\mathcal{N}$. The *nth extension* of that channel is given by the pair $(H_s^{\otimes n}, \mathcal{N}^{\otimes n})$, where $\otimes n$ is used to denote n-fold tensor products. The memoryless nature of the channel is reflected in the fact that the operation performed on the n copies of the channel system is a tensor product of independent single-system operations.

Define an *n code* $(\mathcal{C}, \mathcal{D})$ from H_s into H_c to consist of a trace-preserving quantum operation $\mathcal{C}$ from $H_s^{\otimes n}$ to $H_c^{\otimes n}$, and a trace-preserving quantum operation $\mathcal{D}$ from $H_c^{\otimes n}$ to $H_s^{\otimes n}$. We refer to $\mathcal{C}$ as the *encoding* and $\mathcal{D}$ as the *decoding*.

The *total coding operation* $\mathcal{T}$ is given by

$$\mathcal{T} \equiv \mathcal{D} \circ \mathcal{N}^{\otimes n} \circ \mathcal{C}. \qquad (8.4)$$

The measure of success we use for the total procedure is the *total entanglement fidelity*,

$$F_e(\rho_s^n, \mathcal{T}). \qquad (8.5)$$

In practice we frequently abuse notation, usually by omitting explicit mention of the Hilbert spaces H_s and H_c. Note also that, in principle, the channel could have different input and output Hilbert spaces. To ease notational clutter we do not consider that case here, but all the results we prove go through without change.

Given a source state ρ_s and a channel $\mathcal{N}$, the goal of noisy-channel coding is to find an encoding $\mathcal{C}$ and a decoding $\mathcal{D}$ such that $F_e(\rho_s, \mathcal{T})$ is close to 1; that is, ρ_s and its entanglement is transmitted almost perfectly. In general this is not possible to do. However, Shannon showed in the classical context that by considering blocks of output from the source and performing block encoding and decoding it is possible to considerably expand the class of source states ρ_s for which this is possible. The quantum mechanical version of this procedure is to find a sequence of n codes $(\mathcal{C}^n, \mathcal{D}^n)$ such that as $n \to \infty$, the measure of success $F_e(\rho_s^n, \mathcal{T}^n)$ approaches 1, where $\mathcal{T}^n = \mathcal{D}^n \circ \mathcal{N}^{\otimes n} \circ \mathcal{C}^n$ (we will sometimes refer to such a sequence as a *coding scheme*).

Suppose such a sequence of codes exists for a given source Σ. In this case the channel is said to transmit Σ reliably. We also say that the channel can transmit reliably at a *rate* $R = S(\Sigma)$. (Note that this definition does not require that the channel be able to transmit reliably *any* source with entropy rate less than or equal to R; that is a different potential definition of what it means for a channel to transmit reliably at rate R. We conjecture that the two definitions are equivalent in the contexts considered in this paper.)

A noisy-channel coding theorem enables one to determine, for any source and channel, whether or not the source can be transmitted reliably on that channel. Classically, this is determined by comparing the entropy rate of the source to the capacity of the channel. If the entropy rate of the source is greater than the capacity, the source cannot be transmitted reliably. If the entropy rate is less than the capacity, it can. The conjunction of these two statements is precisely the noisy-channel coding theorem. (The case when the entropy

rate of the source equals the capacity requires separate consideration; sometimes reliable transmission is achievable, and sometimes not.) We expect that in quantum mechanics, the entropy rate $S(\Sigma)$ of the source will play the role of the classical entropy rate. A channel will be able to transmit reliably any source with entropy rate less than the capacity; furthermore, *no* source with entropy rate greater than the capacity will be reliably transmissible (i.e., the channel will be unable to transmit reliably at a rate greater than the capacity). The first part of this would constitute a quantum noisy-channel coding theorem; the second, a "weak converse" of the theorem. (A "strong converse" would require not just that no source with entropy rate greater than the capacity can be reliably transmitted, i.e., transmitted with asymptotic fidelity approaching unity, but would require that all such sources have asymptotic fidelity of transmission approaching zero.)

IX. UPPER BOUNDS ON THE CHANNEL CAPACITY

In this section we investigate a variety of upper bounds on the capacity of a noisy quantum channel.

A. Unitary encodings

This subsection is concerned with the case where the encoding $\mathcal{C}$ is unitary.

For this subsection only we define

$$C_n \equiv \max_{\rho} I(\rho, \mathcal{N}^{\otimes n}), \qquad (9.1)$$

where the maximization is over all inputs ρ to n copies of the channel. The bound on the channel capacity proved in this section is defined by

$$C(\mathcal{N}) \equiv \lim_{n \to \infty} \frac{C_n}{n}. \qquad (9.2)$$

It is not immediately obvious that this limit exists. To see that it does, notice that $C_n \leq n \log_2 d$ and $C_m + C_n \leq C_{m+n}$ and apply the lemma proved in Appendix A. Notice that $C(\mathcal{N})$ is a function of the channel operation only.

We begin with a theorem that places a limit on the entropy rate of a source that can be sent through a quantum channel.

Theorem. Suppose we consider a source $\Sigma = (H_s, \{\cdots \rho_s^n \cdots\})$ and a sequence of unitary encodings $\mathcal{U}^n$ for the source. Suppose further that there exists a sequence of decodings $\mathcal{D}^n$ such that

$$\lim_{n \to \infty} F_e(\rho_s^n, \mathcal{D}^n \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}^n) = 1. \qquad (9.3)$$

Then

$$S(\Sigma) \equiv \lim_{n \to \infty} \sup \frac{S(\rho_s^n)}{n} \leq C(\mathcal{N}). \qquad (9.4)$$

This theorem tells us that we cannot reliably transmit more than $C(\mathcal{N})$ qubits of information per use of the channel.

For unitary $\mathcal{U}^n$ we have

$$I(\rho_s, \mathcal{N}^{\otimes n} \circ \mathcal{U}^n) = I(\mathcal{U}^n(\rho_s), \mathcal{N}^{\otimes n}), \qquad (9.5)$$

and thus

$$I(\rho_s, \mathcal{N}^{\otimes n} \circ \mathcal{U}^n) \leq C_n. \qquad (9.6)$$

By Eq. (7.34) with $\mathcal{E} \equiv \mathcal{N}^{\otimes n} \circ \mathcal{U}^n$, and the fact that $I(\mathcal{U}^n(\rho_s), \mathcal{N}^{\otimes n}) \leq \max_\rho I(\rho, \mathcal{N}^{\otimes n}) \equiv C_n$, it now follows that

$$\frac{S(\rho_s^n)}{n} \leq \frac{C_n}{n} + \frac{2}{n} + 4[1 - F_e(\rho_s^n, \mathcal{D}^n \circ \mathcal{N}^{\otimes n} \circ \mathcal{U}^n)] \log_2 d. \qquad (9.7)$$

[Note that d here is the dimension of a single copy of the source Hilbert space, so that we have inserted d^n for the overall dimension d of Eq. (7.34).] Taking lim sups on both sides of the equation completes the proof of the theorem.

It is extremely useful to study this result at length, since the basic techniques employed to prove the bound are the same as those that appear in a more elaborate guise later in the paper. In particular, what features of quantum mechanics necessitate a change in the proof methods used to obtain the classical bound?

Suppose the quantum analogue of the classical subadditivity of mutual information were true, namely,

$$I(\rho^n, \mathcal{N}^{\otimes n}) \leq \sum_{i=1}^{n} I(\rho_i^n, \mathcal{N}), \qquad (9.8)$$

where ρ^n is any density operator that can be used as input to n copies of the channel, and ρ_i^n is the density operator obtained by tracing out all but the ith channel. Then it would follow easily from the definition that $C_n = C_1$ for all n, and thus

$$C(\mathcal{N}) = C_1 = \max_{\rho} I(\rho, \mathcal{N}). \qquad (9.9)$$

This expression is exactly analogous to the classical expression for channel capacity as a maximum over input distributions of the mutual information between channel input and output. If this were truly a bound on the quantum channel capacity then it would allow easy numerical evaluations of bounds on the channel capacity, as the maximization involved is easy to do numerically, and the coherent information is not difficult to evaluate.

Unfortunately, it is not possible to assume that the quantum mechanical coherent information is subadditive, as shown by example (7.54), and thus in general it is possible that

$$C(\mathcal{N}) > C_1. \qquad (9.10)$$

We will later discuss results of Shor and Smolin [25] (see also DiVincenzo, Shor, and Smolin [26]) that demonstrate that the channel capacity can exceed C_1.

Notice that to evaluate the bound $C(\mathcal{N})$ involves taking the limit in Eq. (9.2). To numerically evaluate this limit directly is certainly not a trivial task, in general. The result we have presented, that Eq. (9.2) is an upper bound on channel capacity, is an important theoretical result that may aid in the

development of effective numerical procedures for obtaining general bounds. But it does not yet constitute an effective procedure.

B. General encodings

We now consider the case where something more general than a unitary encoding is allowed. In principle, it is always possible to perform a nonunitary encoding $\mathcal{C}$ by introducing an extra ancilla system, performing a joint unitary on the source plus ancilla, and then discarding the ancilla.

We define

$$C_n \equiv \max_{\rho,\mathcal{C}} I(\rho, \mathcal{N}^{\otimes n} \circ \mathcal{C}), \tag{9.11}$$

where the maximization is over all inputs ρ to the encoding operation $\mathcal{C}$, which in turn maps to n copies of the channel. The bound on the channel capacity proved in this section is defined by

$$C(\mathcal{N}) \equiv \lim_{n \to \infty} \frac{C_n}{n}. \tag{9.12}$$

Once again, to prove that this limit exists one applies the lemma proved in Appendix A.

To prove that this quantity is a bound on the channel capacity, one applies almost exactly the same reasoning as in the preceding subsection. The result is the following. *Theorem.* Suppose we consider a source $\Sigma = (H_s, \{\cdots \rho_s^n \cdots\})$ and a sequence of encodings $\mathcal{C}^n$ for the source. Suppose further that there exists a sequence of decodings $\mathcal{D}^n$ such that

$$\lim_{n \to \infty} F_e(\rho_s^n, \mathcal{D}^n \circ \mathcal{N}^{\otimes n} \circ \mathcal{C}^n) = 1. \tag{9.13}$$

Then

$$S(\Sigma) \equiv \lim_{n \to \infty} \sup \frac{S(\rho_s^n)}{n} \leqslant C(\mathcal{N}). \tag{9.14}$$

Again, this result places an upper bound on the rate at which information can be reliably transmitted through a noisy quantum channel. The proof is very similar to the earlier proof of a bound for unitary encodings. One simply applies Eq. (7.34) with $\mathcal{E} = \mathcal{N}^{\otimes n} \circ \mathcal{C}^n$ and $\mathcal{D} = \mathcal{D}^n$, to give

$$\frac{S(\rho_s^n)}{n} \leqslant \frac{C_n}{n} + \frac{2}{n} + 4[1 - F_e(\rho_s^n, \mathcal{D}^n \circ \mathcal{N}^{\otimes n} \circ \mathcal{C}^n)] \log_2 d. \tag{9.15}$$

Taking lim sups on both sides of the equation completes the proof.

It is instructive to see why the proof fails when the maximization is done over channel input states alone, rather than over all source states and encoding schemes. The basic idea is that there may exist source states ρ_s and encoding schemes $\mathcal{C}$, for which

$$I(\rho, \mathcal{N} \circ \mathcal{C}) > I(\mathcal{C}(\rho), \mathcal{N}). \tag{9.16}$$

This possibility stems from the failure of the quantum pipelining inequality (7.43). It is clear that the existence of such a scheme would cause the line of proof suggested above to fail. Classically the pipelining inequality holds, and therefore the complication of having to maximize over encodings does not arise.

Having proved that $C(\mathcal{N})$ is an upper bound on the channel capacity, let us now investigate some of the properties of this bound. First we examine the range over which $C(\mathcal{N})$ can vary. Note that

$$0 \leqslant C_n \leqslant n \log_2 d, \tag{9.17}$$

since if ρ is pure then $I(\rho, \mathcal{N}^{\otimes n} \circ \mathcal{C}) = 0$ for any encoding $\mathcal{C}$, and for all ρ and $\mathcal{C}$, $I(\rho, \mathcal{N}^{\otimes n} \circ \mathcal{C}) \leqslant \log d^n = n \log_2 d$, since the channel output has d^n dimensions. It follows that

$$0 \leqslant C(\mathcal{N}) \leqslant \log_2 d. \tag{9.18}$$

This parallels the classical result, which states that the channel capacity varies between 0 and $\log_2 s$, where s is the number of channel symbols. The upper bound on the classical capacity is attained if and only if the classical channel is noiseless.

In the case when $\mathcal{N}$ takes a constant value,

$$\mathcal{N}(\rho) = \sigma, \tag{9.19}$$

for all channel inputs ρ it is not difficult to verify that $C(\mathcal{N}) = 0$. This is consistent with the obvious fact that the capacity for quantum information of such a channel is zero.

The ''completely decohering channel'' is defined by

$$\mathcal{N}(\rho) = \sum_i P_i \rho P_i, \tag{9.20}$$

with $P_i \equiv |i\rangle\langle i|$ a complete orthonormal set of one-dimensional projectors. This channel is classically noiseless, yet a straightforward application of Eq. (7.24) yields $C(\mathcal{N}) = 0$, and therefore this channel has zero capacity for the transmission of entanglement.

More generally, if $\mathcal{N}(\rho) = \Sigma_i A_i \rho A_i^\dagger$, where $A_i = \lambda_i |a_i\rangle\langle b_i|$, then $C(\mathcal{N}) = 0$ by the same argument, and thus the channel capacity for such a channel is zero. As a special case of this result, it follows that the capacity of *any* classical channel as defined in Sec. V to transmit entanglement is zero.

Provided the input and output dimensions of the channel are the same, it is not difficult to show that $C(\mathcal{N}) = \log_2 d$ if and only if $\mathcal{N}$ is unitary.

It is also of interest to consider what happens when channels $\mathcal{N}_1$ and $\mathcal{N}_2$ are composed, forming a concatenated channel, $\mathcal{N} = \mathcal{N}_2 \circ \mathcal{N}_1$. From the data-processing inequality it follows that

$$C(\mathcal{N}_1) \geqslant C(\mathcal{N}). \tag{9.21}$$

It is clear by repeated application of the data-processing inequality that this result also holds if we compose more than two channels together, and even holds if we allow intermediate decoding and reencoding stages. Classical channel capacities also behave in this way: the capacity of a channel

made by composing two (or more) channels together is no greater than the capacity of the first part of the channel alone.

Although Eq. (7.43) might seem to suggest otherwise, in fact,

$$C(\mathcal{N}_2) \geqslant C(\mathcal{N}). \tag{9.22}$$

For let us suppose that $\mathcal{C}$ is the encoding that achieves $C(\mathcal{N})$, so that the total operation is $\mathcal{D} \circ \mathcal{N} \circ \mathcal{C} \equiv \mathcal{D} \circ \mathcal{N}_2 \circ \mathcal{N}_1 \circ \mathcal{C}$. As our encoding for the channel $\mathcal{N}_2$, we may use $\mathcal{N}_1 \circ \mathcal{C}$ and decode with $\mathcal{D}$, hence achieving precisely the same total operation.

Inequalities analogous to Eqs. (9.21) and (9.22) may also be stated for the actual channel capacity. Clearly these statements are true as well.

C. Other encoding protocols

So far we have considered two allowed classes of encodings: encodings where a general unitary operation can be performed on a block of quantum systems, and encodings where a general trace-preserving quantum operation can be performed on a block of quantum systems. If large-scale quantum computation ever becomes feasible it may be realistic to consider encoding protocols of this sort. However, for present-day applications of quantum communication such as quantum cryptography and teleportation, it is realistic to consider much more restricted classes of encodings. In this section we describe several such classes.

We begin by considering the class involving local unitary operations only. We refer to this class as U-L. It consists of the set of operations $\mathcal{C}$ that can be written in the form

$$\mathcal{C}(\rho) = (U_1 \otimes \cdots \otimes U_n) \rho (U_1^\dagger \otimes \cdots \otimes U_n^\dagger), \tag{9.23}$$

where $U_1, \ldots, U_n$ are local unitary operations on systems 1 through n. Another possibility is the class L of encodings involving local operations only, i.e., operations of the form

$$\sum_{i_1, \ldots, i_n} (A_{i_1} \otimes B_{i_2} \otimes \cdots \otimes Z_{i_n}) \rho$$
$$\times (A_{i_1}^\dagger \otimes B_{i_2}^\dagger \otimes \cdots \otimes Z_{i_n}^\dagger). \tag{9.24}$$

In other words, the overall operation has a tensor product form $\mathcal{A} \otimes \mathcal{B} \otimes \cdots \otimes \mathcal{Z}$.

A more realistic class is 1-L—encoding by local operations with one way classical communication. The idea is that the encoder is allowed to do encoding by performing arbitrary quantum operations on individual members (typically, a single qubit) of the strings of quantum systems emitted by a source. This is not unrealistic with present-day technology for manipulating single qubits. Such operations could include arbitrary unitary rotations, and also generalized measurements. After the qubit is encoded, the results of any measurements done during the encoding may be used to assist in the encoding of later qubits. This is what we mean by one way communication—the results of the measurement can only be used to assist in the encoding of later qubits, not earlier qubits.

Another possible class is 2-L—encoding by local operations with two-way classical communication. This may arise in a situation where there are many identical channels operating side by side in space. Once again it is assumed that the encoder can perform arbitrary local operations, only this time two-way classical communication is allowed when performing the encoding.

For any class of encodings Λ arguments analogous to those used above for general and for unitary block coding, ensure that the expression

$$C_\Lambda(\mathcal{N}) \equiv \lim_{n \to \infty} \frac{C_{\Lambda,n}}{n}, \tag{9.25}$$

where

$$C_{\Lambda,n} \equiv \max_{\rho, \mathcal{C} \in \Lambda} I(\rho, \mathcal{N}^{\otimes n} \circ \mathcal{C}), \tag{9.26}$$

is an upper bound to the rate at which quantum information can be reliably transmitted using encodings in Λ. Thus, in addition to the bounds for general and unitary encodings, there are bounds $C_{U\text{-}L}, C_L, C_{1\text{-}L}$, and $C_{2\text{-}L}$, which provide upper bounds on the rate of quantum information transmission for these types of encodings. *A priori* it is not clear what the exact relationships are among these bounds, although various inequalities may easily be proved,

$$C_{U\text{-}L} \leqslant C_L \leqslant C_{1\text{-}L} \leqslant C_{2\text{-}L} \leqslant C_{\text{general}} \tag{9.27}$$

$$C_{U\text{-}L} \leqslant C_{\text{unitary}} \tag{9.28}$$

$$C_{\text{unitary}} \leqslant C_{\text{general}} . \tag{9.29}$$

Furthermore, note that these bounds allow general decoding schemes. It is possible that much tighter bounds may result if we restrict the decoding schemes in the same way we have restricted the encoding schemes.

An interesting and important question is whether there are closed-form characterizations of the sets of quantum operations corresponding to particular types of encoding schemes such as 1-L and 2-L. For example, in the cases of U-L and L there are explicit forms [(9.23),(9.24)] for the classes of encodings allowed. For 1-L the operations take the form

$$\sum_{i_1, \ldots, i_n} (A_{i_1} \otimes B_{i_1,i_2} \otimes \cdots \otimes Z_{i_1,i_2,\ldots i_n}) \rho$$
$$\times (A_{i_1}^\dagger \otimes B_{i_1,i_2}^\dagger \otimes \cdots \otimes Z_{i_1,i_2,\ldots i_n}^\dagger). \tag{9.30}$$

A drawback to this expression is that it is not written in a closed form, making it difficult to perform optimizations over 1-L. It would be extremely valuable to obtain a closed form for the set of operations in 1-L. One possible approach to doing this is to limit the range of the indices in the previous expression. This is related to the number of rounds of classical communication that are involved in the operation. Similar remarks to these also apply to the class 2-L. Indeed, it is not yet clear to us if there is an expression analogous to Eq. (9.30) for 2-L encodings. One possibility is

$$\sum_i (A_i \otimes B_i \otimes \cdots \otimes Z_i) \rho (A_i^\dagger \otimes B_i^\dagger \otimes \cdots \otimes Z_i^\dagger). \tag{9.31}$$

However, although all 2-*L* operations involving a finite number of rounds of communication can certainly be put in this form, we do not presently see why all operations expressible in this form should be realizable with local operations and two-way classical communication.

The classes we have described in this subsection are certainly not the only realistic classes of encodings. Many more classes may be considered, and in specific applications this may well be of great interest. What we have done is illustrated a general technique for obtaining *bounds* on the channel capacity for different classes of encodings. A major difference between classical information theory and quantum information theory is the greater interest in the quantum case in studying different classes of encodings. Classically it is, in principle, possible to perform an arbitrary encoding and decoding operation using a look-up table. However, quantum mechanically this is far from being the case, so there is correspondingly more interest in studying the channel capacities that may result from considering different classes of encodings and decodings.

X. DISCUSSION

What then can be said about the status of the quantum noisy-channel coding theorem in the light of comments made in the preceding sections? While we have established upper bounds, we have not proved achievability of these bounds. How might one prove that these bounds are achievable?

Lloyd [7] has also proposed an expression involving a maximum of the coherent information as the channel capacity,

$$\max_{\rho} I(\rho, \mathcal{N}), \qquad (10.1)$$

and outlines a technique involving random coding for achieving rates up to this quantity. The criterion for reliable transmission used by Lloyd appears to be the subspace fidelity criterion of Eq. (6.7). As noted earlier, this criterion is at least as strong as the criterion based on entanglement fidelity that we have been using, that is, asymptotically good coding schemes with respect to subspace fidelity are also asymptotically good with respect to the entanglement fidelity.

Suppose one applies coding schemes to achieve rates up to Eq. (10.1), but with the basic system used in blocking taken to be *n* of the old systems. Then it is clear that rates up to

$$\max_{\rho} \frac{I(\rho, \mathcal{N}^{\otimes n})}{n} \qquad (10.2)$$

may be achieved using such coding schemes, where the maximization is done over density operators for *n* copies of the source. It follows that rates up to

$$\lim_{n \to \infty} \max_{\rho} \frac{I(\rho, \mathcal{N}^{\otimes n})}{n} \qquad (10.3)$$

may be achieved. This quantity is simply the bound (9.2) that we found earlier for noisy channels with the class of encodings restricted to be unitary. As remarked in the last section, it is in general not possible to identify the quantity appearing
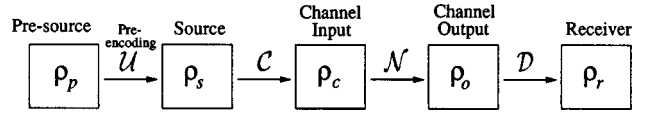

FIG. 7. Noisy quantum channel with an extra stage, a restricted *preencoding* $\mathcal{U}$.

in the previous equation with the quantity (10.1), because the coherent information is not, in general, subadditive, cf. Eq. (7.54).

The coding schemes considered by Lloyd appear to be restricted to be projections followed by unitaries. We call such encodings *restricted encodings*, since they do not cover the full class of encodings possible. For the purposes of proving upper bounds it is not sufficient to consider a restricted class of encodings, since it is possible that other coding schemes may do better, and therefore that the capacity is somewhat larger than Eq. (10.2). We suspect that this is not the case, but have been unable to provide a rigorous proof. A heuristic argument is provided in Sec. X A.

In the light of these remarks it is interesting that the coding scheme of Shor and Smolin [25] (see also DiVincenzo, Shor, and Smolin [26]) provides an example where the nonsubadditivity of the coherent information is exploited to achieve rates of transmission exceeding Eq. (10.1). Nevertheless, the coding schemes considered by DiVincenzo, Shor, and Smolin cannot beat the general bound (9.12), which takes nonsubadditivity into account. A full exposition of this topic will appear elsewhere.

However, one can still make progress towards a proof that the expression (9.12), which bounds the channel capacity, is the correct capacity. If we accept that it is possible to attain rates up to Eq. (10.2), then the four-stage construction illustrated in Fig. 7 shows that Eq. (9.12) is a correct expression for the capacity; i.e., that in addition to being an upper bound as shown in Sec. IX, it is also achievable.

For a fixed block size *n* one finds an encoding $\mathcal{C}^n$ for which the maximum in

$$C_n \equiv \max_{\mathcal{C}^n, \rho_s} TI(\rho_s, \mathcal{C}^n) \qquad (10.4)$$

is achieved. One then regards the composition $\mathcal{N}^{\otimes n} \circ \mathcal{C}^n$ as a single noisy quantum channel, and applies the achievability result on restricted encodings to the joint channel $\mathcal{N}^{\otimes n} \circ \mathcal{C}^n$ to achieve an even longer *mn* block coding scheme with high entanglement fidelity.

This gives a joint coding scheme $\mathcal{U}^{mn} \circ (\mathcal{C}^n)^{\otimes m}$, which for sufficiently large blocks *m* and *n* can come arbitrarily close to achieving the channel capacity (9.12). An important open question is whether Eq. (9.12) is equal to Eq. (9.2). It is clear that the former expression is at least as large as the latter; we give a heuristic argument for equality in the next subsection, but rigorous results are needed.

Thus, we think it likely that the expression (9.2) will turn out to be the maximum achievable rate of reliable transmission through a quantum channel. But this is still not satisfactory as an expression for the capacity, because of the difficulty of evaluating the limit involved. At a minimum, we would like to know enough about the rate of convergence of C_n to its limit to be able to accurately estimate the error in a

numerical calculation of capacity, thus providing an effective procedure for calculating the capacity to any desired degree of accuracy.

A. Unitary versus nonunitary encoding

For the purposes of obtaining a capacity theorem for general encodings and decodings, a restriction on the class of encodings is clearly unacceptable. For example, given a source density operator whose eigenvalues are not all equal, we may not even be able to send it reliably through a noiseless channel whose capacity is just greater than the source entropy rate without doing nonunitary compression as described in Refs. [27—29]. This compression, which is essentially projection onto the *typical subspace* [27] of the source, is not a unitary operation, and thus we expect that nonunitary operations will be essential to showing achievability of the noisy-channel capacity.

We conjecture that once the projection onto the typical source subspace is accomplished, nonunitary operations are of no further use in achieving reliable transmission through a noisy channel. Although we have not yet rigorously shown this, we give a heuristic argument below. If the conjecture is true then it can be used to show that expressions (9.2) and (9.12) are equal.

Our heuristic argument applies only to sources for which a *typical subspace* [27] exists. This includes all i.i.d. sources, for which the output is of the form $\rho_s^{\otimes n}$. Let Λ be the projector onto the typical subspace after n uses of the source, and $\bar{\Lambda}$ the projector onto the orthogonal subspace. Given any positive δ it is true that for sufficiently large n,

$$\mathrm{tr}(\bar{\Lambda}\rho_s^{\otimes n}\bar{\Lambda}) \leq \delta. \quad (10.5)$$

Defining the restriction of the source to the typical subspace,

$$\rho_T^n \equiv \frac{\Lambda\rho_s^{\otimes n}\Lambda}{\mathrm{tr}(\Lambda\rho_s^{\otimes n}\Lambda)}, \quad (10.6)$$

and applying the continuity lemma for entanglement fidelity (6.14), we see that

$$|F_e(\rho_T^n, \mathcal{E}) - F_e(\rho_s^{\otimes n}, \mathcal{E})| \leq \frac{4\delta}{(1-\delta)^2}, \quad (10.7)$$

for any trace-preserving operation $\mathcal{E}$. By choosing n sufficiently large δ can be made arbitrarily small, and thus we see that for the entanglement fidelity for the source to be high asymptotically, it is necessary and sufficient that the entanglement fidelity be high asymptotically for the restriction of the source to the typical subspace.

We now come to the heuristic argument. In order that the entanglement fidelity for the total channel be high, it is asymptotically necessary and sufficient that the composite operation $\mathcal{D}^n \circ \mathcal{N}^{\otimes n} \circ \mathcal{C}^n$ have high entanglement fidelity when the source is restricted to the typical subspace τ. Hence, if an encoding $\mathcal{C}^n$ is nonunitary on τ, it must be "close to reversible" on τ, and $\mathcal{D}^n \circ \mathcal{N}^{\otimes n}$ must be close to reversing it. In [14] it is shown that perfect reversibility of an operation on a subspace M is equivalent to the statement that the operation, restricted to that subspace, may be represented by operators
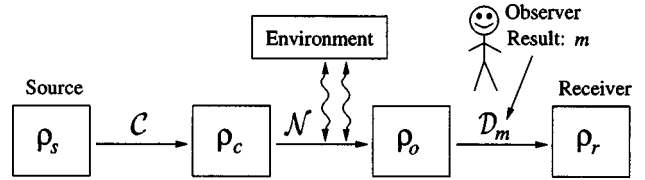

FIG. 8. Noisy quantum channel with a classical observer.

$\{\sqrt{p_i}U_iP_M\}$, where $P_MU_j^\dagger U_iP_M = \delta_{ij}P_M$ and P_M is the projector onto M. That is, the operation randomly (with probabilities p_i) chooses a unitary that moves the state into one of a mutually orthogonal set of subspaces. Hence $\mathcal{C}^n$, *in its action on the source's typical subspace,* is close to some perfectly reversible operation $\mathcal{C}_*^n$ consisting of "randomly picking a unitary into an orthogonal subspace." Hence the entanglement fidelity of the total operation $\mathcal{T}$ is close to that of $\mathcal{T}_*^n$, in which the encoding $\mathcal{C}^n$ is replaced with $\mathcal{C}_*^n$. The linearity of the entanglement fidelity in the operation implies that for at least one of the unitaries U_i in the random-unitaries representation of the perfectly reversible operation $\mathcal{C}_*^n$, the entanglement fidelity is at least as good if the unitary is substituted for $\mathcal{C}_*^n$. Therefore, arbitrary encodings $\mathcal{C}^n$ are close to unitary encodings of τ into a subspace of the channel's Hilbert space. Thus the only nonunitarity that it is necessary to consider is the restriction to the source's typical subspace.

XI. CHANNELS WITH A CLASSICAL OBSERVER

In this section we consider a generalized version of the quantum noisy-channel coding problem. Suppose that in addition to a noisy interaction with the environment there is also a classical observer who is able to perform a measurement. This measurement may be on the channel or the environment of the channel, or possibly on both.

The result of the measurement is then sent to the decoder, who may use the result to assist in decoding. We assume that this transmission of classical information is done noiselessly, although it is also interesting to consider what happens when the classical transmission also involves noise. It can be shown [11] that the state received by the decoder is again related to the state ρ used as input to the channel by a quantum operation $\mathcal{N}_m$, where m is the measurement result recorded by the classical observer,

$$\rho \rightarrow \frac{\mathcal{N}_m(\rho)}{\mathrm{tr}[\mathcal{N}_m(\rho)]}. \quad (11.1)$$

The basic situation is illustrated in Fig. 8. The idea is that by giving the decoder access to classical information about the environment responsible for noise in the channel it may be possible to improve the capacity of that channel, by allowing the decoder to choose different decodings $\mathcal{D}_m$ depending on the measurement result m.

We now give a simple example that illustrates that this can be the case. Suppose we have a two-level system in a state ρ and an initially uncorrelated four-level environment initially in the maximally mixed state $I/4$, so the total state of the joint system is

$$\rho \otimes \frac{I}{4}. \tag{11.2}$$

We fix an orthonormal basis $|1\rangle, |2\rangle, |3\rangle, |4\rangle$ for the environment. We assume that a unitary interaction between the system and environment takes place, given by the unitary operator

$$U = I \otimes |1\rangle\langle 1| + \sigma_x \otimes |2\rangle\langle 2| + \sigma_y \otimes |3\rangle\langle 3| + \sigma_z \otimes |4\rangle\langle 4|. \tag{11.3}$$

The output of the channel is thus

$$\rho \to \mathcal{N}(\rho) \equiv \mathrm{tr}_E \left[U\left(\rho \otimes \frac{I}{4}\right) U^\dagger \right]. \tag{11.4}$$

The quantum operation $\mathcal{N}$ can be given two particularly useful forms,

$$\mathcal{N}(\rho) = \tfrac{1}{4}(I\rho I + \sigma_x \rho \sigma_x + \sigma_y \rho \sigma_y + \sigma_z \rho \sigma_z) \tag{11.5}$$

$$= \frac{I}{2}. \tag{11.6}$$

It is not difficult to show from the second form that

$$C(\mathcal{N}) = 0, \tag{11.7}$$

and thus the channel capacity for the channel $\mathcal{N}$ is equal to zero. Suppose now that an observer is introduced, who is allowed to perform a measurement on the environment. This measurement is a Von Neumann measurement in the $|1\rangle, |2\rangle, |3\rangle, |4\rangle$ basis, and yields a corresponding measurement result, $m = 1, 2, 3, 4$. Then the quantum operations corresponding to these four measurement outcomes are

$$\mathcal{N}_1(\rho) = \tfrac{1}{4}\rho \tag{11.8}$$

$$\mathcal{N}_2(\rho) = \tfrac{1}{4}\sigma_x \rho \sigma_x \tag{11.9}$$

$$\mathcal{N}_3(\rho) = \tfrac{1}{4}\sigma_y \rho \sigma_y \tag{11.10}$$

$$\mathcal{N}_4(\rho) = \tfrac{1}{4}\sigma_z \rho \sigma_z. \tag{11.11}$$

Each of these is unitary, up to a constant multiplying factor, so conditioned on knowing the measurement result m, the corresponding channel capacity C_m is perfect. That is,

$$C_m = 1 \tag{11.12}$$

for all measurement outcomes m. This is an example where the capacity of the observed channel is strictly greater than for the unobserved channel.

This result is particularly clear in the context of teleportation. Nielsen and Caves [10] showed that the problem of teleportation can be understood as the problem of a quantum noisy channel with an auxiliary classical channel. In the single qubit teleportation scheme of Bennett *et al.* [9] there are four quantum operations relating the state Alice wishes to teleport to the state Bob receives, corresponding to each of the four measurement results. In that scheme it happens that those four operations are the $\mathcal{N}_m$ we have described above. Furthermore in the absence of the classical channel, that is,

when Alice does not send the result of her measurement to Bob, the channel is described by the single operation $\mathcal{N}$. Clearly, in order that causality be preserved we expect that the channel capacity be zero. On the other hand, in order that teleportation be able to occur we expect that the channel capacity C_m is equal to 1, as was shown above. Teleportation understood in this way as a noisy channel with a classical side channel offers a particularly elegant way of seeing that the transmission of quantum information may sometimes be greatly improved by making use of classical information.

The remainder of this section is organized into three subsections. Section XI A proves bounds on the capacity of an observed channel. This requires nontrivial extensions of the techniques developed earlier for proving bounds on the capacity of an unobserved channel. Section XI B relates work done on the observed channel to the work done on the unobserved channel. Section XI C discusses possible extensions to this work on observed channels.

A. Upper bounds on channel capacity

We now prove several results bounding the channel capacity of an observed channel, just as we did earlier for the unobserved channel. The following lemma generalizes the earlier entanglement fidelity lemma for quantum operations, which was the foundation of our earlier proofs of upper bounds on the channel capacity.

Lemma (generalized entanglement fidelity lemma for operations). Suppose $\mathcal{E}_m$ are a set of quantum operations such that $\sum_m \mathcal{E}_m$ is a trace-preserving quantum operation. Suppose further that $\mathcal{D}_m$ is a trace-preserving quantum operation for each m. Then

$$S(\rho) \leq \sum_m \mathrm{tr}[\mathcal{E}_m(\rho)]I(\rho,\mathcal{E}_m) + 2 + 4[1 - F_e(\rho,\mathcal{T})] \log_2 d, \tag{11.13}$$

where

$$\mathcal{T} \equiv \sum_m \mathcal{D}_m \circ \mathcal{E}_m. \tag{11.14}$$

By the second step of the data processing inequality (7.9), $I(\rho,\mathcal{E}_m) \geq I(\rho,\mathcal{D}_m \circ \mathcal{E}_m)$ for each m, and noting also that by the trace-preserving property of $\mathcal{D}_m$, $\mathrm{tr}[\mathcal{E}_m(\rho)] = \mathrm{tr}[(\mathcal{D}_m \circ \mathcal{E}_m)(\rho)]$, we obtain

$$S(\rho) \leq S(\rho) + \sum_m \{\mathrm{tr}[\mathcal{E}_m(\rho)]I(\rho,\mathcal{E}_m)$$
$$- \mathrm{tr}[(\mathcal{D}_m \circ \mathcal{E}_m)(\rho)]I(\rho,\mathcal{D}_m \circ \mathcal{E}_m)\}. \tag{11.15}$$

Applying the generalized convexity theorem for coherent information (7.24) gives

$$-\sum_m \mathrm{tr}[(\mathcal{D}_m \circ \mathcal{E}_m)(\rho)]I(\rho,\mathcal{D}_m \circ \mathcal{E}_m) \leq -I(\rho,\mathcal{T}). \tag{11.16}$$

We obtain

$$S(\rho) \leq \sum_m \ \mathrm{tr}[\mathcal{E}_m(\rho)]I(\rho,\mathcal{E}_m)+S(\rho)-I(\rho,\mathcal{T}).$$
(11.17)

But $\mathcal{T}=\Sigma_m \mathcal{D}_m \circ \mathcal{E}_m$ is trace preserving since $\mathcal{D}_m$ is trace preserving and $\Sigma_m \mathcal{E}_m$ is trace preserving, and thus by Eq. (7.33),

$$S(\rho)-I(\rho,\mathcal{T})=S(\rho)-S(\mathcal{T}(\rho))+S_e(\rho,\mathcal{T}) \quad (11.18)$$

$$\leq 2S_e(\rho,\mathcal{T}). \quad (11.19)$$

Finally, an application of the quantum Fano inequality (6.9) along with the observations that the entropy function h appearing in that inequality is bounded above by 1, and $\log_2 (d^2-1)\leq 2\log_2 d$, gives

$$S(\rho)\leq \sum_m \ \mathrm{tr}[(\mathcal{D}_m \circ \mathcal{E}_m)(\rho)]I(\rho,\mathcal{D}_m \circ \mathcal{E}_m)+2$$
$$+4[1-F_e(\rho,\mathcal{T})]\log_2 d, \quad (11.20)$$

as we set out to prove.

If we define

$$C(\{\mathcal{N}_m\})\equiv \lim_{n\to\infty}\max_{\mathcal{C}^n,\rho}\sum_{m_1,\dots,m_n}\mathrm{tr}[(\mathcal{N}_{m_1}\otimes\cdots\otimes\mathcal{N}_{m_n}\circ\mathcal{C}^n)(\rho)]$$
$$\times\frac{I(\rho,\mathcal{N}_{m_1}\otimes\cdots\otimes\mathcal{N}_{m_n}\circ\mathcal{C}^n)}{n}, \quad (11.21)$$

we may use Eq. (11.13) to easily prove that $C(\{\mathcal{N}_m\})$ is an upper bound on the rate of reliable transmission through an observed channel, in precisely the same way we earlier used (7.34) to prove bounds for unobserved channels.

We may derive the same bound in another fashion if we associate observed channels with trace-preserving unobserved channels in the following fashion suggested by examples in [8]. To an observed channel $\{\mathcal{N}_m\}$ we associate a single trace-preserving operation $\mathcal{M}$ from H_c to a larger Hilbert space $H_c\otimes M$, where M is a "register" Hilbert space. Each dimension of M corresponds to a different measurement result m. The operation is specified by

$$\mathcal{M}(\rho)=\sum_m \ \mathcal{N}_m(\rho)\otimes|m\rangle\langle m|, \quad (11.22)$$

where $|m\rangle$ is some set of orthogonal states corresponding to the measurement results that may occur. This map is an "all-quantum" version of the observed channel.

Since our upper bounds to the capacity of an unobserved channel apply also to channels with output Hilbert spaces of different dimensionality than the input space, they apply to this map as well. It is easily verified that the coherent information for the map $\mathcal{M}$ acting on ρ is the same as the average coherent information for the observed channel $\mathcal{N}_m$ acting on ρ, which appears in Eq. (11.13) and in the bound (11.21). To show this, define

$$p_m\equiv\mathrm{tr}[\mathcal{N}_m(\rho^Q)], \quad (11.23)$$

where we are again working in the RQ picture of operations. Then $\rho^{Q'}=\mathcal{M}(\rho^Q)$ is given by Eq. (11.22), so that

$$S(\rho^{Q'})=H(p_m)+\sum_m \ p_mS\!\left(\frac{\mathcal{N}_m(\rho^Q)}{p_m}\right), \quad (11.24)$$

since the density matrices $\mathcal{N}_m(\rho^Q)\otimes|m\rangle\langle m|$ are mutually orthogonal. Also,

$$\rho^{R'Q'}=\left(\mathcal{I}\otimes\sum_m \ \mathcal{N}_m^*\right)(\rho^{RQ}), \quad (11.25)$$

where by definition $\mathcal{N}_m^*(\rho)=\mathcal{N}_m(\rho)\otimes|m\rangle\langle m|$. Thus

$$S(\rho^{R'Q'})=H(p_m)+\sum_m \ p_mS\!\left(\frac{(\mathcal{I}\otimes\mathcal{N}_m)(\rho^{RQ})}{p_m}\right). \quad (11.26)$$

Hence the coherent information is

$$I(\rho^Q,\mathcal{M})=\sum_m \ p_m\!\left[S\!\left(\frac{\mathcal{N}_m(\rho^Q)}{p_m}\right)-S\!\left(\frac{(\mathcal{I}\otimes\mathcal{N}_m)(\rho^{RQ})}{p_m}\right)\right], \quad (11.27)$$

which can be rewritten as the average coherent information for $\{\mathcal{N}_m\}$,

$$I(\rho^Q,\mathcal{M})=\sum_m \ p_mI(\rho^Q,\mathcal{N}_m). \quad (11.28)$$

So an application of the bound (9.12) on the rate of transmission through the unobserved channel $\mathcal{M}$ shows that the expression on the right-hand side of Eq. (11.21) that bounds the capacity of the observed channel $\{\mathcal{N}_m\}$ also bounds the capacity of $\mathcal{M}$. This result provides some evidence for the intuitively reasonable proposition that $\mathcal{M}$ and $\{\mathcal{N}_m\}$ are equivalent with respect to transmission of quantum information.

Bennett, DiVincenzo, and Smolin [8] derive capacities for three simple channels that may be viewed as taking the form (11.22). The *quantum erasure channel* takes the input state to a fixed state orthogonal to the input state with probability ϵ; otherwise, it transmits the state undisturbed. An equivalent observed channel would with probability ϵ replace the input state with a standard pure state $|0\rangle\langle 0|$ within the input subspace, and also provide classical information as to whether this replacement has occurred or not. The *phase erasure channel* randomizes the phase of a qubit with probability δ, and otherwise transmits the state undisturbed; it also supplies classical information as to which of these alternatives occurred. The *mixed erasure or phase-erasure channel* may either erase or phase erase, with exclusive probabilities ϵ and δ. Bennett, DiVincenzo, and Smolin note that the capacity $\max(0,1-2\epsilon)$ of the erasure channel is in fact the one-shot maximal coherent information. We have verified that the capacities they derive for the phase-erasure channel $(1-\delta)$ and the mixed erasure or phase-erasure channel $\max(0,1-2\epsilon-\delta)$ are the same as the one-shot maximal average coherent information for the corresponding observed channels, lending some additional support to the view that the bounds we have derived here are in fact the capacities.

B. Relationship to unobserved channel

Suppose a quantum system passes through a channel, interacts with an environment, and then measurements are performed on the *environment alone*. How is the capacity of this observed channel related to the capacity of the channel that results if *no measurement* had been performed on the environment? Physically, it is clear that the capacity when measurements are performed must be at least as great as when no measurements on the environment are performed, since the decoder can always ignore the result of the measurement. In this subsection we show that bounds we have derived on channel capacity have this same property: observation of the environment can never decrease the bounds we have obtained.

Suppose $\{\mathcal{N}_m\}$ are the operations describing the different possible measurement outcomes. Then the operation describing the same channel, but without any observation of the environment, is

$$\mathcal{N} = \sum_m \mathcal{N}_M. \qquad (11.29)$$

Recall the expressions for the bound on the capacity of the unobserved channel,

$$C(\mathcal{N}) = \lim_{n \to \infty} \max_{\mathcal{C}^n, \rho} \frac{I(\rho, \mathcal{N}^{\otimes n} \circ \mathcal{C}^n)}{n}, \qquad (11.30)$$

and the observed channel,

$$C(\{\mathcal{N}_m\}) = \lim_{n \to \infty} \max_{\mathcal{C}^n, \rho} \sum_{m_1, \ldots, m_n} \text{tr}[(\mathcal{N}_{m_1} \otimes \cdots \otimes \mathcal{N}_{m_n} \circ \mathcal{C}^n)(\rho)]$$

$$\times \frac{I(\rho, \mathcal{N}_{m_1} \otimes \cdots \otimes \mathcal{N}_{m_n} \circ \mathcal{C}^n)}{n}, \qquad (11.31)$$

but the generalized convexity theorem (7.24) for coherent information implies that

$$\sum_{m_1, \ldots, m_n} \text{tr}[(\mathcal{N}_{m_1} \otimes \cdots \otimes \mathcal{N}_{m_n} \circ \mathcal{C}^n)(\rho)]$$

$$\times \frac{I(\rho, \mathcal{N}_{m_1} \otimes \cdots \otimes \mathcal{N}_{m_n} \circ \mathcal{C}^n)}{n} \leq \frac{I(\rho, \mathcal{N}^{\otimes n} \circ \mathcal{C}^n)}{n}, \qquad (11.32)$$

and thus

$$C(\mathcal{N}) \leq C(\{\mathcal{N}_m\}). \qquad (11.33)$$

To see that this inequality may sometimes be strict, return to the example considered earlier in this section in the context of teleportation. In that case it is not difficult to verify that

$$0 = C(\mathcal{N}) < C(\{\mathcal{N}_m\}) = 1. \qquad (11.34)$$

What these results show is that our bounds on the channel capacity are never made any worse by observing the environment, but sometimes they can be made considerably better. This is a property that we certainly expect the quantum channel capacity to have, and we take as an encouraging sign that the bounds we have proved in this paper are in fact achievable, that is, the true capacities.

C. Discussion

All the questions asked about the bounds on channel capacity for an unobserved channel can be asked again for the observed channel: questions about achievability of bounds, the differences in power achievable by different classes of encodings and decodings, and so on. We do not address these problems here, beyond noting that they are important problems that need to be addressed by future research.

Many new twists on the problem of the quantum noisy channel arise when an observer of the environment is allowed. For example, one might consider the situation where the classical channel connecting the observer to the decoder is noisy. What then are the resources required to transmit coherent quantum information?

It may also be interesting to prove results relating the classical and quantum resources that are required to perform a certain task. For example, in teleportation it can be shown that one requires not only the quantum channel, but also two bits of classical information, in order to transmit quantum information with perfect reliability.

XII. CONCLUSION

In this paper we have shown that different information transmission problems may result in different channel capacities for the same noisy quantum channel. We have developed some general techniques for proving upper bounds on the amount of information that may be transmitted reliably through a noisy quantum channel.

Perhaps the most interesting thing about the quantum noisy-channel problem is to discover what is new and essentially *quantum* about the problem. The following list summarizes what we believe are the essentially new features.

(1) The insight that there are many essentially different information transmission problems in quantum mechanics, all of them of interest depending on the application. These span a spectrum between two extremes: (i) The transmission of a discrete set of mutually orthogonal quantum states through the channel. Such problems are problems of transmitting classical information through a noisy quantum channel. (ii) The transmission of entire subspaces of quantum states through the channel, which necessarily keeps all other quantum resources, including entanglement, intact. This is likely to be of interest in applications such as quantum computation, cryptography, and teleportation where superpositions of quantum states are crucial. Such problems are problems of transmitting coherent quantum information through a noisy quantum channel. Both of these cases and a variety of intermediate cases are important for specific applications. For each case, there is great interest in considering different classes of allowed encodings and decodings. For example, it may be that encoding and decoding can only be done using local operations and one-way classical communication. This may give rise to a different channel capacity than occurs if we allow nonlocal encoding and decoding. Thus there are different noisy-channel problems depending on what classes of encodings and decodings are allowed.

(2) The use of quantum entanglement to construct examples where the quantum analogue of the classical pipelining inequality $H(X:Z) \leqslant H(Y:Z)$ for a Markov process $X \rightarrow Y \rightarrow Z$, fails to hold [cf. Eq. (7.43)].

(3) The use of quantum entanglement to construct examples where the subadditivity property of mutual information,

$$H(X_1, \ldots, X_n : Y_1, \ldots, Y_n) \leqslant \sum_i H(X_i : Y_i), \quad (12.1)$$

fails to hold [cf. Eq. (7.54)].

There are many more interesting open problems associated with the noisy-channel problem than have been addressed here. The following is a sample of those problems that we believe to be particularly important:

(1) The development of an effective procedure for determining channel capacities. We believe that this is the most important problem remaining to be addressed. Assuming our upper bound

$$C(\mathcal{N}) = \lim_{n \to \infty} \max_{\rho, \mathcal{C}} \frac{I(\rho, \mathcal{N}^{\otimes n} \circ \mathcal{C})}{n} \quad (12.2)$$

is, in fact, the channel capacity for general encodings, it still remains to find an effective procedure for evaluating this quantity. Both maximizations can be done relatively easily, since they are of a continuous function over a compact set. However, we do not yet understand the convergence of the limit well enough to have an effective procedure for evaluating this quantity.

(2) Estimation of channel capacities for realistic channels. This work could certainly be done theoretically and perhaps also experimentally. Recent work on *quantum process tomography* [30,31] points the way toward experimental determination of the quantum channel capacity. A related problem is to analyze how stable the determination of channel capacities is with respect to experimental error.

(3) As suggested in Sec. IX C it would be interesting to see what channel capacities are attainable for different classes of allowable encodings and/or decodings, for example, encodings where the encoder is only allowed to do local operations and one-way classical communication, or encodings where the encoder is allowed to do local operations and two-way classical communication. We have shown how to prove bounds on the channel capacity in these cases; whether these bounds are attainable is unknown.

(4) The development of rigorous general techniques for proving attainability of channel capacities, which may be applied to different classes of allowed encodings and decodings.

(5) Finding the capacity of a noisy quantum channel for classical information. A related problem arises in the context of *superdense coding*, where one half of an EPR pair can be used to send two bits of classical information. It would be interesting to know to what extent this performance is degraded if the pair of qubits shared between sender and receiver is not an EPR pair, but rather the sharing is done using a noisy quantum channel, leading to a decrease in the number of classical bits that can be sent. Given a noisy quantum channel, what is the maximum amount of classical information that can be sent in this way?

(6) All work done thus far has been for discrete channels, that is, channels with finite dimensional state spaces. It is an important and nontrivial problem to extend these results to channels with infinite dimensional state spaces.

(7) A more thorough study of noisy channels that have a classical side channel. Can the classical information obtained by an observer be related to changes in the channel capacity? What if the classical side channel is noisy? Many other fascinating problems, too many to enumerate here, suggest themselves in this context.

There are many other ways the classical results on noisy channels have been extended—considering channels with *feedback*, developing *rate-distortion* theory, understanding *networks* consisting of more than one channel, and so on. Each of these could give rise to highly interesting work on noisy quantum channels. It is also to be expected that interesting new questions will arise as experimental efforts in the field of quantum information develop further. Perhaps of chief interest to us is to develop a still clearer understanding of the essential differences between the quantum noisy-channel and the classical noisy-channel problems.

APPENDIX A: EXISTENCE OF LIMITS

This appendix contains a lemma that can be used to prove the existence of several limits that appear in this paper.

Lemma. Suppose $c_1, c_2, \ldots$ is a nonnegative sequence such that $c_n \leqslant kn$ for some $k \geqslant 0$, and

$$c_m + c_n \leqslant c_{m+n}, \quad (A1)$$

for all m and n. Then

$$\lim_{n \to \infty} \frac{c_n}{n} \quad (A2)$$

exists and is finite.

Proof. Define

$$c \equiv \lim_n \sup \frac{c_n}{n}. \quad (A3)$$

This always exists and is finite, since $c_n \leqslant kn$ for some $k \geqslant 0$. Fix $\epsilon > 0$ and choose n sufficiently large that

$$\frac{c_n}{n} > c - \epsilon. \quad (A4)$$

Suppose m is any integer strictly greater than $\max(n, n/\epsilon)$. Then by Eq. (A1),

$$\frac{c_m}{m} \geq \frac{c_n}{n} \frac{n}{m}\left(1 + \frac{c_{m-n}}{c_n}\right). \qquad \text{(A5)}$$

Using the fact that $lc_n \leq c_{ln}$ [an immediate consequence of Eq. (A1)] with $l = \lfloor m/n \rfloor - 1$ gives

$$\frac{c_{m-n}}{c_n} \geq \left\lfloor \frac{m}{n} \right\rfloor - 1 \qquad \text{(A6)}$$

$$\geq \frac{m}{n} - 2, \qquad \text{(A7)}$$

where $\lfloor x \rfloor$ is the integer immediately below x. Plugging the last inequality into Eq. (A5) gives

$$\frac{c_m}{m} \geq \frac{c_n}{n}\left(1 - \frac{n}{m}\right). \qquad \text{(A8)}$$

But $-n/m > -\epsilon$ and $c_n/n \geq c - \epsilon$, so

$$\frac{c_m}{m} \geq (c - \epsilon)(1 - \epsilon). \qquad \text{(A9)}$$

This equation holds for all sufficiently large m, and thus

$$\lim_{n} \inf \frac{c_n}{n} \geq (c - \epsilon)(1 - \epsilon). \qquad \text{(A10)}$$

But ϵ was an arbitrary number greater than 0, so letting $\epsilon \to 0$ we see that

$$\lim_{n} \inf \frac{c_n}{n} \geq c = \lim_{n} \sup \frac{c_n}{n}. \qquad \text{(A11)}$$

It follows that $\lim_n c_n/n$ exists, as claimed.

APPENDIX B: MAXIMA OF THE COHERENT INFORMATION

Various convexity and concavity properties are useful in calculating classical channel capacities, and the same is true in the quantum situation. This appendix is devoted to an explication of the basic properties of convexity and concavity related to the coherent information and the relation of these properties to expressions such as Eq. (9.12).

A *convex set* S is a subset of a vector space such that

given any two points $s_1, s_2 \in S$ and any λ such that $0 < \lambda < 1$, then the *convex combination*, $\lambda s_1 + (1 - \lambda)s_2$, is also an element of S. Geometrically, this means that given any two points in the set, the line joining them is also in the set. An *extremal point* of S is a point s that cannot be formed from the convex combination of any other two points in the set. A *convex function* f on S is a real-valued function such that for any λ satisfying $0 < \lambda < 1$,

$$f(\lambda s_1 + (1 - \lambda)s_2) \leq \lambda f(s_1) + (1 - \lambda)f(s_2); \qquad \text{(B1)}$$

a concave function satisfies the same condition but with the inequality reversed.

The first useful fact about maxima is the following.

Local maximum is a global maximum. Suppose f is a concave function on a convex set S. Then a local maximum of f is also a global maximum of f. This follows by supposing that s_1 and s_2 are distinct local maxima. If $f(s_1) < f(s_2)$, say, then

$$f(\lambda s_1 + (1 - \lambda)s_2) \geq \lambda f(s_1) + (1 - \lambda)f(s_2) \qquad \text{(B2)}$$

$$> f(s_1), \qquad \text{(B3)}$$

by concavity of f. By choosing sufficiently small values of λ we see that this violates the fact that s_1 is a local maximum. Thus f has the same value for all local maxima, from which it follows that all local maxima are also global maxima for the function.

The following lemma, from [32], is extremely useful in computing the maxima of convex functions on convex sets.

Convexity Lemma. Suppose f is a continuous convex function on a compact, convex set S. Then there is an extremal point at which f attains its global maximum.

The proof is obvious. The reason for our interest in the result is because for fixed ρ and trace-preserving operations $\mathcal{E}$, the coherent information $I(\rho, \mathcal{E})$ is a convex, continuous function of the operation $\mathcal{E}$. The set of trace-preserving quantum operations forms a compact, convex set, and thus by the convexity lemma $I(\rho, \mathcal{E})$ attains its maximum for a quantum operation $\mathcal{E}$, which is extremal in the set of all trace-preserving quantum operations.

Choi [13] has proved that any extremal point in the set of trace-preserving quantum operations has a set of operation elements $\{A_i\}$ such that (1) there are at most d elements A_i—this is to be contrasted with the general situation, where there may be up to d^2 elements; (2) the A_i are linearly independent.

This result provides a considerable saving in the class of quantum operations that must be optimized over in order to numerically calculate expressions of the form (9.12). Unfortunately, this only takes us part of the way towards proving that the expressions (9.12) and (9.2) are identically equal, or, alternatively, it suggests a starting point for a search for counterexamples to the proposition that the two quantities are equal. If the extremal points of the set of quantum operations were the unitary operations we would be done. However, that is not the case, as the above theorem shows.

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

[2] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

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

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

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

[6] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[7] S. Lloyd, Phys. Rev. A. **56**, 1613 (1997).

[8] C. H. Bennett, D. P. DiVincenzo, and J. Smolin, Phys. Rev. Lett. **78**, 3217 (1997).

[9] C. H. Bennett *et al.*, Phys. Rev. Lett. **70**, 1895 (1993).

[10] M. A. Nielsen and C. M. Caves, Phys. Rev. A **55**, 2547 (1997).

[11] K. Kraus, *States, Effects, and Operations* (Springer-Verlag, Berlin, 1983).

[12] K. Hellwig and K. Kraus, Commun. Math. Phys. **16**, 142 (1970).

[13] M.-D. Choi, Linear Algebr. Appl. **10**, 285 (1975).

[14] M. A. Nielsen, C. M. Caves, B. Schumacher, and H. Barnum, Proc. R. Soc. London, Ser. A **454**, 277 (1998).

[15] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic, Dordrecht, 1993).

[16] C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum Mechanics* (Wiley, New York, 1977).

[17] R. I. G. Hughes, *The Structure and Interpretation of Quantum Mechanics* (Harvard University Press, Cambridge, 1989).

[18] G. Lüders, Ann. Phys. (Leipzig) **8**, 323 (1951).

[19] R. Jozsa, J. Mod. Opt. **41**, 2315 (1995).

[20] E. Knill and R. Laflamme, Phys. Rev. A **55**, 900 (1997).

[21] M. A. Nielsen, e-print quant-ph/9606012.

[22] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).

[23] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes* (North-Holland, New York, 1977).

[24] J. Preskill, Proc. R. Soc. London, Ser. A **454**, 385 (1998).

[25] P. W. Shor and J. A. Smolin, e-print quant-ph/9604006.

[26] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin, Phys. Rev. A **57**, 830 (1998).

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

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

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

[30] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390 (1997).

[31] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).

[32] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities* (Dover, New York, 1992).