Entropic bounds on coding for noisy quantum channels

Nicolas J. Cerf[1,2,3]
[1]W. K. Kellogg Radiation Laboratory, California Institute of Technology, Pasadena, California 91125
[2]Information and Computing Technologies Research Section, Jet Propulsion Laboratory, Pasadena, California 91109
[3]Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, 1050 Bruxelles, Belgium

(Received 11 July 1997)

In analogy with its classical counterpart, a noisy quantum channel is characterized by a *loss*, a quantity that depends on the channel input and the quantum operation performed by the channel. The loss reflects the transmission quality: if the loss is zero, *quantum* information can be perfectly transmitted at a rate measured by the quantum source entropy. By using *block coding* based on sequences of n entangled symbols, the *average* loss (defined as the overall loss of the joint n-symbol channel divided by n, when $n\to\infty$) can be made lower than the loss for a *single* use of the channel. In this context, we examine several upper bounds on the rate at which quantum information can be transmitted reliably via a noisy channel, that is, with an asymptotically vanishing average loss while the *one-symbol* loss of the channel is nonzero. These bounds on the channel capacity rely on the entropic Singleton bound on quantum error-correcting codes [Phys. Rev. A **56**, 1721 (1997)]. Finally, we analyze the Singleton bounds when the noisy quantum channel is supplemented with a classical auxiliary channel. [S1050-2947(98)09404-9]

PACS number(s): 03.67.Hk, 03.65.Bz, 89.70.+c

I. INTRODUCTION

Within recent years, the quantum theory of information and communication has undergone a dramatic evolution (see, e.g., [1]). Major progress has been made toward the extension to the quantum regime of the classical theory of information pioneered by Shannon [2]. In particular, the use of *quantum* communication channels in order to transmit not only classical information but also intact quantum states (or *quantum* information) has received a considerable amount of attention, following the proof of the quantum analog of Shannon's fundamental theorem for noiseless coding by Schumacher [3]. It has been shown that the von Neumann entropy plays the role of a quantum information-theoretic entropy in the sense that it characterizes the minimum amount of quantum resources (e.g., number of quantum bits) that is necessary to code an ensemble of quantum states with an asymptotically vanishing distortion in the absence of noise. This result suggests that a general quantum theory of information, paralleling Shannon theory, can be developed based on this concept. While such a full theory does not exist as of yet, a great deal of effort has been devoted to this issue over the past few years, and several fundamental results have been obtained, ranging from entanglement-based communication schemes [4] to quantum error-correcting codes [5]. In particular, a substantial amount of work has been devoted recently to the transmission of arbitrary states (or quantum information) through *noisy* quantum channels (see, e.g., [6–9]). A quantum state processed by such a channel undergoes *decoherence* by interacting with an external system or environment, which effects an alteration of quantum information. A natural question that arises in this context concerns the possibility of transmitting quantum information *reliably*, in spite of quantum noise, if it is suitably encoded as sequences of quantum bits in analogy with the standard construction used for classical channels. More specifically, a fundamental issue is to understand the quantum analog of Shannon's noisy channel coding theorem and to define the *capacity* of a noisy quantum channel, i.e., an upper limit to the amount of quantum information that can be processed with an arbitrarily high fidelity. While several attempts have been made to define a quantum analog of Shannon mutual information that would be a natural candidate for such a quantum measure of capacity (see the concepts of coherent information [7,8] or von Neumann mutual entropy [9,10]), the problem of characterizing in general the capacity of a noisy quantum channel is still unsolved.

The purpose of this paper is to further clarify the description of noisy quantum channels centered on the von Neumann mutual entropy (see [9]). It has been shown recently that a consistent information-theoretic framework that closely parallels Shannon's construction can be developed, based on von Neumann *conditional* and *mutual* entropies [10–13]. The central peculiarity of this framework is that it involves *negative* conditional entropies in order to account for quantum nonlocal correlations between entangled variables. This is in contrast with Shannon information theory in which marginal and conditional entropies are all nonnegative quantities. Negative quantum conditional entropies simply reflect the nonmonotonicity of the von Neumann entropy [14] (the entropy of a composite system can be lower than that of its components if the latter are entangled). The resulting information-theoretic formalism provides grounds for the quantum extension of the usual algebraic relations between Shannon entropies in multipartite systems [11–13]. Surprisingly, many concepts of Shannon theory can be straightforwardly translated to the quantum regime by extending the range for quantum (conditional and mutual) entropies with respect to the classical one in order to encompass entanglement [10]. This is very helpful in analyzing quantum information processes in a *unified* framework, paralleling Shannon theory. For example, entanglement-based quantum communication processes [10], quantum channels [9], and quantum error-correcting codes [15] can be described along these lines.

In this paper, we focus on the application of this information-theoretic framework to the issue of finding upper bounds on the capacity of quantum codes and quantum channels. In Sec. II, we outline the general treatment of noisy quantum channels based on quantum entropies [9], and extend it to the characterization of consecutive uses of a quantum memoryless channel (cf. the notions of one-symbol and average loss explained in Sec. II D). This provides a simple framework to consider *block coding* with quantum channels. Note that, just as in Shannon information theory, quantum entropic considerations alone do not result in *constructive* methods for building codes. Rather, they are useful to derive bounds on what can possibly be achieved or not, from basic principles. Accordingly, we analyze in Sec. III several upper bounds (based on the Singleton bound on quantum codes [15]) for standard quantum channels such as the quantum erasure or depolarizing channel. This confirms bounds on the quantum capacity that were derived otherwise, but places this problem in a unified context. Finally, we examine in Sec. IV the extension of this quantum entropic treatment of noisy quantum channels to the case where an auxiliary *classical* channel is available. Quantum teleportation appears then as a special case of this construction when no block coding is applied.

II. ENTROPIC CHARACTERIZATION OF NOISY QUANTUM CHANNELS

A. Notations

Let us start by summarizing the basic definitions that will be useful in the rest of this paper when considering noisy quantum channels. The entropy of a quantum system X (of arbitrary dimension) is defined as the von Neumann entropy of the density operator ρ_X that characterizes the state of X, i.e.,

$$S(X)=S[\rho_X]\equiv-\mathrm{Tr}(\rho_X\log_2\rho_X). \tag{2.1}$$

It can be viewed as the uncertainty about X in the sense that it measures (asymptotically) the minimum number of quantum bits (qubits) necessary to specify X [3]. This definition can be extended to the notions of *conditional* and *mutual* von Neumann entropies, based on a simple parallel with their classical counterparts which is motivated in [10–12]. For a bipartite system XY characterized by ρ_{XY}, the *conditional* von Neumann entropy is

$$S(X|Y)=S(XY)-S(Y) \tag{2.2}$$

while the *mutual* von Neumann entropy is

$$\begin{aligned}S(X{:}Y)&=S(X)-S(X|Y)\\&=S(Y)-S(Y|X)\\&=S(X)+S(Y)-S(XY),\end{aligned} \tag{2.3}$$

where $S(XY)$ is calculated from ρ_{XY} while $S(X)$ and $S(Y)$ are obtained from the reduced density operators $\rho_X=\mathrm{Tr}_Y(\rho_{XY}$ and $\rho_Y=\mathrm{Tr}_X(\rho_{XY})$. Subadditivity of quantum entropies implies $S(X{:}Y)\geqslant 0$, where the equality holds if X and Y are independent (i.e., $\rho_{XY}=\rho_X\otimes\rho_Y$). Note that, when $S(XY)=0$ (i.e., the joint system XY is in a pure state), we have $S(X{:}Y)=2S(X)=2S(Y)$ as a consequence of the Schmidt decomposition. This property will be useful in the following. Several quantum entropies can also be defined for characterizing *multipartite* quantum systems. Consider, for instance, a tripartite system XYZ. The von Neumann conditional mutual entropy (of X and Y, conditionally on Z) can be defined as

$$\begin{aligned}S(X{:}Y|Z)&=S(X|Z)-S(X|YZ)\\&=S(X|Z)+S(Y|Z)-S(XY|Z)\\&=S(XZ)+S(YZ)-S(Z)-S(XYZ)\end{aligned} \tag{2.4}$$

in perfect analogy with the classical expressions. Note that the strong subadditivity of quantum entropies implies $S(X{:}Y|Z)\geqslant 0$ [12]. We can also define the von Neumann *ternary* mutual entropy as

$$S(X{:}Y{:}Z)=S(X{:}Y)-S(X{:}Y|Z). \tag{2.5}$$

Note that, if $S(XYZ)=0$ (i.e., the ternary system is in a pure state), then $S(X{:}Y{:}Z)=0$ [12], or, equivalently, $S(X{:}Y)=S(X{:}Y|Z)$, a property which is very useful in the analysis of quantum channels. Also, *chain rules* for quantum entropies can be written, such as

$$S(X{:}YZ)=S(X{:}Y)+S(X{:}Z|Y), \tag{2.6}$$

which parallel the classical relations [12]. The motivation for building such a quantum entropic framework is that it provides an *information-theoretic* formulation of quantum entanglement in multipartite systems, unified with Shannon's description of classical correlation. It is an extension of Shannon's formalism beyond its original range, as reflected, for example, by the fact that the quantum mutual entropy can reach *twice* the maximum value allowed for classical entropies [10], that is,

$$0\leqslant S(X{:}Y)\leqslant 2\min[S(X),S(Y)]. \tag{2.7}$$

This factor 2 appears in many quantum information-theoretic relations (see below), and originates from the Araki-Lieb inequality for quantum entropies [10–12].

B. Quantum mutual entropy, loss, and noise

Let us now outline the entropic treatment of a noisy quantum channel (see also Ref. [9]). Such a treatment explicitly displays the correspondence with the standard description of noisy classical channels (see Appendix A), thereby unifying classical and quantum channels. Our description involves three quantum systems of arbitrary dimensions: Q (the quantum system whose processing by the channel is concerned), R (a ''reference'' system that Q is initially entangled with), and E (an external system or environment that Q is interacting with in the noisy channel). More specifically, we assume that Q is initially entangled with R, so that the joint state of Q and R is the *pure* state $|\Psi_{RQ}\rangle$. We may as well regard Q as a quantum source, being initially in a mixed state ρ_Q (realized by a given ensemble of quantum states associated with some probability distribution). The ''purification'' of ρ_Q into $|\Psi_{RQ}\rangle$ can always be achieved by extending the Hilbert space $\mathcal{H}_Q$ to $\mathcal{H}_{RQ}$, so that we have $\rho_Q=\mathrm{Tr}_R(|\Psi_{RQ}\rangle\langle\Psi_{RQ}|)$. The corresponding reduced von Neumann entropies are

$$S(R)=S(Q)\equiv S, \tag{2.8}$$

where S is called the *source entropy*. In the dual picture where an *arbitrary* pure state of Q (rather than entanglement) is sent through the channel, S then measures the ''arbitrariness'' of Q (it can be viewed as the average number of quantum bits that are to be processed by the channel in order to transmit the state of Q). In what follows, we prefer to consider a quantum input Q that is entangled with R, so that the preservation of entanglement—rather than of arbitrary states—will be the central feature of a quantum transmission channel. The initial *mutual entropy* to be transmitted is thus

$$S(R{:}Q)=2S, \tag{2.9}$$

that is, *twice*[1] the source entropy.

When it is processed by the channel, Q interacts with E (assumed to be initially in a pure state $|0\rangle$) according to the unitary transformation U_{QE}, inducing *decoherence*. This describes the most general (trace-preserving) operation of a quantum channel that is allowed by quantum mechanics. Roughly speaking, the resulting *noisy* quantum channel is such that, typically, only a fraction of the initial entanglement with R can be recovered after having been processed by the channel (the rest of the entanglement with R is lost, in the sense that it is transferred to the environment). More specifically, the decohered quantum system after interaction with E, denoted as Q', is in the state

$$\rho'_Q=\mathrm{Tr}_E(U_{QE}(\rho_Q\otimes|0\rangle\langle 0|)U^\dagger_{QE}), \tag{2.10}$$

where ρ_Q is the initial state of Q (with source entropy S). The completely positive linear map $\rho_Q\rightarrow\rho'_Q$ corresponds to the ''quantum operation'' performed by the noisy channel [7]. After such an environment-induced decoherence, the joint system $R'Q'E'$ is in the state $|\Psi_{R'Q'E'}\rangle=(1_R\otimes U_{QE})|\Psi_{RQ}\rangle|0_E\rangle$ whose entropy Venn diagram is represented in Fig. 1 (the primes refer to the systems *after* decoherence). Note that, as the reference is not involved in decoherence, we have $R'\equiv R$.

The entropy diagram of $R'Q'E'$ depends on three parameters, the von Neumann *mutual* entropy (or the quantum information) I, the *loss* L, and the *noise* N, these quantities being defined in analogy with their classical counterparts:

$$I=S(R{:}Q'), \tag{2.11}$$

$$L=S(R{:}E'|Q')=S(R{:}E'), \tag{2.12}$$

$$N=S(Q'{:}E'|R)=S(Q'{:}E'). \tag{2.13}$$

The classical correspondence can be made fully explicit by including an environment in the description of a classical channel, as shown in [9]. The second equality in Eqs. (2.12) and (2.13) has no classical analog, and results from the vanishing of the ternary mutual entropy $S(R{:}Q'{:}E')$ (see [9,15]). Physically, the quantum information I corresponds to the residual mutual entropy between the decohered quantum output Q' and the reference system R that purifies the quantum input Q. The loss L is the mutual entropy that has arisen between the environment after decoherence E' and the reference system R, while the noise N is the mutual entropy between the decohered quantum output Q' and the environment E'. Note that I, L, and N can be written as a function of reduced entropies only, without explicitly involving the environment E in the discussion, by making use of the Schmidt decomposition of the state of $R'Q'E'$, namely $S(E')=S(RQ')$:

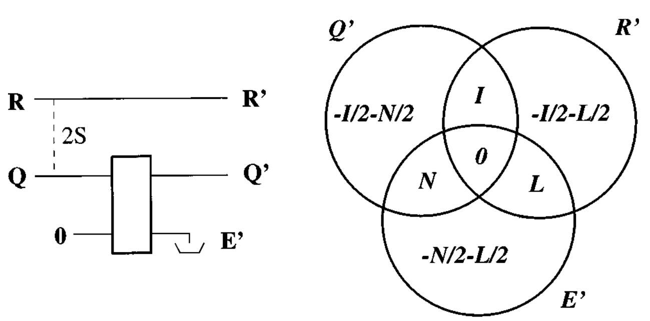

FIG. 1. Schematic representation of the quantum operation effected by a noisy quantum channel. The quantum system Q is initially entangled with the reference R, with a mutual entropy of twice the source entropy S (this is indicated by a dashed line). Then Q decoheres by interacting with an environment E (initially in a pure state $|0\rangle$). The entropy Venn diagram summarizes the entropic relations between Q' (output of the quantum channel), R' (reference), and E' (environment) *after* decoherence. The three parameters I, L, and N denote the von Neumann mutual entropy (quantum information), the loss, and the noise, respectively.

$$I=S(Q)+S(Q')-S(RQ'), \tag{2.14}$$

$$L=S(Q)+S(RQ')-S(Q'), \tag{2.15}$$

$$N=S(Q')+S(RQ')-S(Q). \tag{2.16}$$

It can also be shown that these three quantities are in fact independent of the choice of the reference system R whenever the latter purifies the quantum input Q, so that they provide a most concise entropic characterization of information flow in the channel. They depend in general on the channel input (i.e., ρ_Q) *and* on the quantum operation performed by the channel (i.e., the completely positive trace-preserving map on Q that is specified by U_{QE} in the joint space of Q and E). This exactly parallels the situation for the analog classical quantities. The information I, loss L, and noise N of a classical channel of input X and output Y (see Appendix A) indeed depend on the input distribution $p(x)$ and on the channel ''operation'' characterized by $p(y|x)$.

Among these three quantities, only I and L are relevant as far as (forward) information transmission through the channel is concerned (the noise N plays a role in the description of the ''reverse'' channel, just as for classical channels). Indeed, information processing is characterized by the balance between the von Neumann mutual entropy and the loss, these two quantities always summing to *twice* the source entropy:

[1] Note that this factor *two* reflects a fundamental difference between classical and quantum channels (see Appendix A for comparison). Such a factor is omnipresent in the quantum information-theoretic relations between entropies [12].

$$I+L=2S(Q)\equiv 2S. \tag{2.17}$$

The mutual entropy $I=S(R{:}Q')$ represents the amount of the initial mutual entropy with respect to R (i.e., $2S$) that has been processed by the channel, while the loss $L=S(R{:}E')$ corresponds to the fraction of it that is unavoidably lost in the environment. If the channel is *lossless* ($L=0$), then $I=2S$, so that the interaction with the environment can be perfectly ''undone,'' and the initial entanglement of Q can be fully recovered by an appropriate decoding [7,9]. (Equivalently, this means that an arbitrary initial state of Q can be recovered without error.) This can be understood by noting that R does not become entangled *directly* with the environment in a lossless channel, but only via the output Q' (see Fig. 1 when $L=0$). An operation on Q' only (namely, the decoding operation) is enough to transfer the unwanted entanglement with E' (measured by the noise N) to an ancilla, while preserving the entanglement $2S$ with R.

Thus, if $L=0$, a perfect transmission of information (*including* quantum information) can be achieved through the channel by applying an appropriate decoding. When $I=0$, on the other hand, no information at all (classical *or* quantum) can be processed by the channel. This is the case, for example, of the quantum depolarizing channel with $p=3/4$ (see Sec. III D). In between these limiting cases, classical information (and, up to some *restricted* extent, quantum information) can be reliably transmitted at the expense of a decrease in the rate by making use of *block coding*. The analysis of such a transmission of quantum information immune to noise is the main focus of this paper.

For completeness, let us mention that a channel with $N=0$ is the quantum analog of a *deterministic* channel [16], that is, a channel where the input fully determines the output (see Appendix A). The quantum output Q' is indeed not directly entangled with E' but only via R, which implies that its entanglement with R remains intact (see Fig. 1 when $N=0$). This does not mean, however, that perfect error correction is achievable, as an operation on the reference R is needed to recover the initial entanglement $2S$ between Q and R. A channel which is both lossless ($L=0$) and deterministic ($N=0$) is called *noiseless*; its action on Q is the identity operator (or any *fixed* unitary operator). For example, the overall channel including a noisy quantum channel along with the encoder and decoder is obviously noiseless if perfect error correction is achieved. (In other words, the decoder is used to eliminate the quantum noise $N\neq 0$ by transferring the entanglement with E to an ancilla, which then makes the overall channel noiseless *provided* that $L=0$.) It is worth noting here that the noise N and the loss L play symmetric roles when considering the ''reverse'' channel obtained by interchanging the input and output. (This is true for classical channels as well.) More specifically, N and I always sum to twice the *output* entropy,

$$I+N=2S(Q'), \tag{2.18}$$

in analogy with Eq. (2.17). Roughly speaking, N plays the role of the loss of the reverse channel, as shown in Sec. II C.

C. Properties of quantum *I*, *L*, and *N*

The above entropies definitions for a noisy quantum channel can be shown to fulfill several properties, akin to classical ones, which make them reasonable quantum measures of information, loss, or noise (see also Ref. [9]). First, the quantum mutual entropy I can be shown to be *concave* in the input ρ_Q for a fixed channel, i.e., a fixed quantum operation $\rho_Q\rightarrow\rho_{Q'}$ or a fixed U_{QE}. Therefore, any local maximum of I is the absolute maximum, that is, the von Neumann capacity of the channel. This parallels the concavity of the Shannon mutual entropy $H(X{:}Y)$ in the input probability distribution $p(x)$ for a fixed channel, i.e., fixed $p(y|x)$ [17]. Second, I is *convex* in the output $\rho_{Q'}$ for a fixed input ρ_Q. This property will be used in the next section when considering a ''probabilistic'' channel (the effective channel resulting from the probabilistic use of a family of channels). It is the quantum analog of the property that the information $H(X{:}Y)$ processed by a classical channel is a convex function of $p(y|x)$ for a fixed $p(x)$ [17]. These two properties are simple to prove by reexpressing the von Neumann mutual entropy I as

$$S(R{:}Q')=S(Q'E')+S(Q')-S(E')=S(Q')+S(Q'|E') \tag{2.19}$$

or as

$$S(R{:}Q')=S(R)+S(Q')-S(RQ')=S(R)-S(R|Q'). \tag{2.20}$$

If the input ρ_Q is a convex combination of density operators while the channel is fixed, it is easy to see that ρ_{QE} and therefore $\rho_{Q'E'}$ are also convex combinations (as the channel operation is linear). Since the conditional entropy $S(Q'|E')$ is concave in a convex combination of $\rho_{Q'E'}$ while $S(Q')$ is concave in $\rho_{Q'}$ [14], Eq. (2.19) implies the *concavity* of the quantum mutual entropy I in the input for a fixed channel. The second property can be proven the same way by noting that, if we have a ''probabilistic'' channel—a convex combination of quantum channels—acting on a fixed input, then $\rho_{RQ'}$ is a convex combination of density operators while ρ_R is constant. Thus, Eq. (2.20) together with the concavity of the conditional entropy $S(R|Q')$ in a convex combination of $\rho_{RQ'}$ implies that the quantum mutual entropy I is *convex* in the output for a fixed input.

A third important property is that the mutual entropy I and the quantum loss L are *subadditive* when considering a channel made of several independent quantum channels used *in parallel*. This will be shown when analyzing quantum block coding (cf. Sec. II D). Finally, it can be proved that I obeys (forward and reverse) data-processing inequalities when considering chained quantum channels. If we chain two channels by using the output of the first as an input for the second (see Fig. 2), the total (1+2) channel $\rho_Q\rightarrow\rho_{Q'}\rightarrow\rho_{Q''}$ is characterized by

$$I_{12}=S(R{:}Q''), \tag{2.21}$$

$$L_{12}=S(R{:}E'E''), \tag{2.22}$$

$$N_{12}=S(Q''{:}E'E''), \tag{2.23}$$

since we can regard the two environments E' and E'' as a global environment for this total channel.

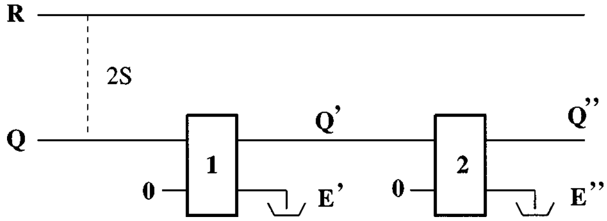

FIG. 2. Schematic view of the chaining of two noisy quantum channels. In each of them, the input state decoheres by interacting with a (separate) environment. The input of the first channel is initially entangled with R, with a source entropy of S (see the dashed line). The output of this channel Q' is then used as an input for the second channel. Since Q' is purified by RE' (not by R alone), the ''reference'' system that must be considered in the entropic characterization of the second channel is RE'.

Using the chain rule for quantum mutual entropies $S(R{:}E'E'')=S(R{:}E')+S(R{:}E''|E')$, and remembering that $S(R{:}E''|E')\geqslant 0$ as a result of strong subadditivity, we obtain

$$0\leqslant L_1\leqslant L_{12}, \tag{2.24}$$

where $L_1=S(R{:}E')$ is the loss of the first channel while L_{12} is the loss of the total channel. Thus, the loss can only increase by further processing of quantum information in the second channel. Since $I_1+L_1=2S(Q)$ and $I_{12}+L_{12}=2S(Q)$, we obtain the *forward* data-processing inequality

$$I_{12}\leqslant I_1\leqslant 2S(Q) \tag{2.25}$$

implying that the mutual entropy of the total channel cannot exceed the one of the first channel. This is the quantum analog of $H(X{:}Z)\leqslant H(X{:}Y)\leqslant H(X)$ for chained classical channels $X\rightarrow Y\rightarrow Z$ [17].

Now, if we use the chain rule $S(Q''{:}E'E'')=S(Q''{:}E'')+S(Q''{:}E'|E'')$ together with strong subadditivity, we obtain

$$0\leqslant N_2\leqslant N_{12}, \tag{2.26}$$

where $N_2=S(Q''{:}E'')$ is the noise of the second channel while N_{12} is the noise of the total channel. As $I_2+N_2=2S(Q'')$ and $I_{12}+N_{12}=2S(Q'')$, we obtain the *reverse* data-processing inequality

$$I_{12}\leqslant I_2\leqslant 2S(Q''), \tag{2.27}$$

where $I_2=S(RE'{:}Q'')$ is the mutual entropy processed by the second channel. (Note that the ''reference'' system that purifies the input Q' of the second channel is RE'.) This parallels the classical inequality $H(X{:}Z)\leqslant H(Y{:}Z)\leqslant H(Z)$ for chained channels [17]. Equations (2.24) and (2.26) emphasize that the loss L and the noise N play a symmetric role in this entropic description if one interchanges the input and the output of the quantum channel (''time-reversal''), just as for classical channels. This is reflected by the symmetry between the forward and the reverse data-processing inequalities.

D. One-symbol loss and average loss

The central idea of classical error correction by block coding is to introduce *correlations* between the bits that make a block, in order to have redundancy in the transmitted flow of data. This can make the transmission asymptotically immune to errors, up to some level of noise. In quantum error-correcting codes, the qubits that form a block are *entangled* in a specific way, so that a partial alteration due to decoherence can be recovered [5]. Even though entanglement gives rise to some qualitatively new features (see [15] for a detailed analysis), the objective is similar. Namely, when block coding is used, i.e., when say k ''logical'' qubits are encoded into blocks of n ''physical'' qubits, it is possible to achieve a situation where the *overall* loss of the joint (n-bit) channel is arbitrarily small, while the loss for individual qubits (for each use of the channel) is finite. In analogy with the classical construction, if blocks of n qubits that are initially entangled with respect to R (with a mutual entropy $2k$) can be transmitted through the channel with an asymptotically vanishing overall loss, we say that the channel processes $2k/n$ bits of entanglement *per qubit*. Equivalently, the channel is transmitting at a *rate* $R=k/n$ (on average, k arbitrary binary quantum states can be transmitted for n transmitted qubits). The maximum rate at which quantum information can be reliably sent through the noisy channel is defined as the quantum *channel capacity*. (This maximum has to be taken over all possible coding schemes, and for $n\rightarrow\infty$.) Whether a good (and operational) definition of such a ''purely quantum'' channel capacity exists is currently an open question. In the following, we restrict ourselves to the issue of finding upper bounds on the rate of perfect quantum information transmission (and therefore on such a ''purely quantum'' capacity).

Let us consider the asymptotic use of a quantum discrete memoryless channel, where n (tending to infinity) qubits are transmitted sequentially.[2] Each qubit may decohere due to an environment (quantum noise), the exact interaction depending on the considered noise model. The important point is that the environment for each qubit is initially *independent* of the one interacting with every other qubit. Thus the information process can be viewed as n sequential uses of a quantum *memoryless* channel (the environment being ''reset'' after each use) or, equivalently, as n *parallel* independent channels processing one qubit each (see Fig. 3). We assume that the set of n input symbols $(Q_1,\ldots,Q_n)$ are initially entangled with R, so that $S(R{:}Q_1\cdots Q_n)=2S$ and $S(R)=S(Q_1\cdots Q_n)=S$. If we consider these n symbols as the single input of a joint n-bit channel $Q_1\cdots Q_n\rightarrow Q_1'\cdots Q_n'$, information transmission is described by the mutual entropy

$$\begin{aligned} I&=S(R{:}Q_1'\cdots Q_n')\\ &=S(Q_1\cdots Q_n)+S(Q_1'\cdots Q_n')-S(E_1'\cdots E_n')\\ &=S(Q_1'\cdots Q_n'|E_1'\cdots E_n')+S(Q_1'\cdots Q_n') \end{aligned} \tag{2.28}$$

and the loss

[2] Throughout this paper, we use indistinctly the terms qubit or symbol to denote the quantum state that is sent in a single use of the channel. As a matter of fact, the reasoning is totally general, and applies to quantum states (or symbols) in a Hilbert space of arbitrary dimension $\geqslant 2$.

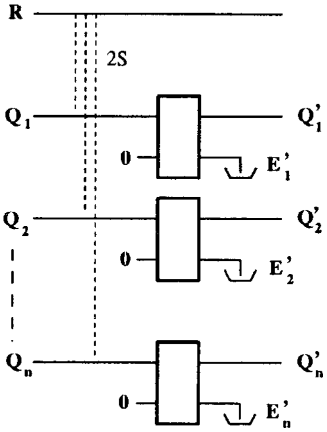

FIG. 3. Schematic view of a memoryless quantum channel. This channel is used n times, but the environment is ''reset'' after each use. This can be viewed as n parallel (independent) channels, each one being used for one of the input symbols. The n input symbols $(Q_1,Q_2,\ldots,Q_n)$ are initially entangled with R (as indicated by a dashed line), with a joint source entropy of S.

$$\begin{aligned} L&=S(R{:}E'_1\cdots E'_n)\\ &=S(Q_1\cdots Q_n)+S(E'_1\cdots E'_n)-S(Q'_1\cdots Q'_n)\\ &=S(E'_1\cdots E'_n|Q'_1\cdots Q'_n)+S(E'_1\cdots E'_n), \end{aligned} \tag{2.29}$$

where we have made use of the conservation of entropy imposed by the unitarity of the global interaction with the n environments $E_1,\ldots,E_n$. Obviously, we have $I+L=2S(Q_1\cdots Q_n)=2S(R)$, which is twice the source entropy S of the joint channel.

Each individual channel $Q_i\rightarrow Q'_i$ can be described in the same way, noting that each Q_i interacts with an environment E_i (initially in a pure state $|0\rangle$) which results in Q'_i and E'_i. The only difference here is that R has to be supplemented with all the input Q's *except* Q_i in order to purify Q_i. Thus, the mutual entropy characterizing the ith channel is

$$I_i=S(RQ_1\cdots Q_{i-1}Q_{i+1}\cdots Q_n{:}Q'_i) \tag{2.30}$$

while the corresponding loss is

$$L_i=S(RQ_1\cdots Q_{i-1}Q_{i+1}\cdots Q_n{:}E'_i). \tag{2.31}$$

These quantities can be reexpressed by using the fact that the environments $E_1,\ \ldots\ ,E_n$ are initially in a product state and that the Q_i and E_i interact pairwise. We have

$$I_i=S(Q_i)+S(Q'_i)-S(E'_i)=S(Q'_i|E'_i)+S(Q'_i) \tag{2.32}$$

and

$$L_i=S(Q_i)+S(E'_i)-S(Q'_i)=S(E'_i|Q'_i)+S(E'_i). \tag{2.33}$$

For each channel, the loss and the mutual entropy sum to twice the source entropy of the channel (tracing over all the other channels): $I_i+L_i=2S(Q_i)$. The subadditivity of von Neumann entropies,

$$S(E'_1\cdots E'_n)\leqslant S(E'_1)+\cdots+S(E'_n), \tag{2.34}$$

and the subadditivity of von Neumann conditional entropies,

$$S(E'_1\cdots E'_n|Q'_1\cdots Q'_n)\leqslant S(E'_1|Q'_1)+\cdots+S(E'_n|Q'_n), \tag{2.35}$$

combined with Eqs. (2.29) and (2.33) imply that the loss is *subadditive*:

$$L\leqslant L_1+\cdots+L_n. \tag{2.36}$$

The same reasoning can be made using Eqs. (2.28) and (2.32) and interchanging the Q'_i's and E'_i's in Eqs. (2.34) and (2.35), which results in the equivalent expression for mutual entropies:

$$I\leqslant I_1+\cdots+I_n. \tag{2.37}$$

The latter inequality corresponds to the *subadditivity* of the von Neumann mutual entropy for parallel channels. Finally, using the relation between the loss and the mutual entropy for individual channels and for the joint channel, we obtain

$$L_1+\cdots+L_n-2M\leqslant L\leqslant L_1+\cdots+L_n \tag{2.38}$$

with $M=S(Q_1)+\cdots+S(Q_n)-S(Q_1\cdots Q_n)\geqslant 0$. Equivalently, if we define the *average* loss[3] of the joint n-bit channel as $l=L/n$ for $n\rightarrow\infty$, we see that

$$l_1-2m\leqslant l\leqslant l_1, \tag{2.39}$$

where $m=M/n$ and $l_1=(L_1+\cdots+L_n)/n$ is the *one-symbol* loss, i.e., the loss for a single use of an individual channel averaged over all 1-bit channels. Thus, Eqs. (2.38) or (2.39) imply that the loss cannot increase by using block coding (using parallel channels). It typically decreases by an amount which is bounded by $2M$ (or $2m$), a quantity related to the entanglement between the input symbols. (Note that $M=0$ if the input symbols are independent.) The analog construction for a classical channel is presented in Appendix A in order to clarify the straightforward classical to quantum correspondence. As an example, let us consider the use of blocks of two qubits. Assume also that the two 1-bit channels are identical, i.e., $L_1=L_2\equiv l_1$. The *average* loss of the joint 2-bit channel, $l=L/2$, can be bounded by

$$l_1-S(Q_1{:}Q_2)\leqslant l\leqslant l_1. \tag{2.40}$$

This explicitly shows that block coding can decrease the average loss only when the symbols are entangled [i.e., $S(Q_1{:}Q_2)>0$].

Equations (2.38) or (2.39) allow us to derive a simple upper bound on the maximum achievable rate by block coding, as a function of the one-symbol loss (or mutual entropy) for a single use of the channel. Indeed, only if the lower bound on L (or l) extends to zero (that is, if $2M\geqslant L_1+\cdots+L_n$) is it possible that block coding makes the joint channel

[3] The average loss l reflects the effective loss *per qubit* processed in the noisy channel, that is, the loss affecting the overall process (encoding + joint channel viewed as n parallel one-bit noisy channels) divided by the number of physical qubits n when $n\rightarrow\infty$.

perfectly immune to noise while each 1-bit channel has a nonvanishing loss. Thus, we have the necessary condition for having a vanishing average loss ($l=0$):

$$2S(Q_1\cdots Q_n)\leqslant 2S(Q_1)+\cdots+2S(Q_n)-L_1-\cdots-L_n. \tag{2.41}$$

As a consequence, the rate of quantum information transmission through the joint channel, $R=S(Q_1\cdots Q_n)/n$, is bounded from above by *half* the averaged *one-symbol* mutual entropy for individual channels:

$$R\leqslant\frac{I_1+\cdots+I_n}{2n}. \tag{2.42}$$

Thus, the (averaged) mutual von Neumann entropy characterizing each use of the channel provides an upper bound on the achievable rate of transmission by block coding through the noisy channel. Except for the factor 1/2, this inequality parallels the one for a classical channel (see Appendix A). Remember that the quantum capacity of a channel is defined as the maximum rate that can be achieved through the channel (over all possible input and coding schemes) with a fidelity arbitrarily close to 1. The classical analogy suggests then that the (maximum) one-symbol von Neumann mutual entropy yields the quantum capacity. However, this upper bound appears not to be attainable in general (see, e.g., [6]), in contrast with the equivalent classical bound. (The physical meaning of the von Neumann mutual entropy is better understood in the context of noisy superdense coding, as shown in [9].) Therefore, it is necessary to derive more constraining entropic upper bounds on R, which is the main concern of the rest of this paper.

In the next section, we build on the entropic derivation of the Singleton bound for quantum codes presented in Ref. [15], and extend it to the treatment of noisy quantum channels in order to find better upper bounds on the rate of perfect quantum information transmission (and therefore on the ''purely quantum'' capacity). The bounds that we derive can be attained in some cases (e.g., for the quantum erasure channel), or not in other cases (e.g., for the quantum depolarizing channel). It is unknown whether such a purely entropic approach unifying classical and quantum channels can possibly yield the best (asymptotically attainable) upper bound, just as it is the case for classical channels, but this is not out of the question. This will be further investigated in future work.

III. ENTROPIC BOUNDS ON CODES AND CHANNEL CAPACITIES

In this section, we derive several bounds, either on quantum codes or on quantum channel capacities, using an entropic approach based on the Singleton bound (see Ref. [15]).

A. Quantum channel subject to a p-bounded fraction of erasures

We say that a quantum channel is subjected to a p-bounded fraction of erasures [18] if, among n uses of the channel, a fraction of pn qubits[4] (at most) are *erased* (or replaced by a distinguishable third state, e.g., $|2\rangle$). When considering *erasures* (rather than errors), the important point is that it is possible to perform an incomplete measurement of each qubit at the output of the channel, to check whether it is in the $|2\rangle$ (erasure) state, or in the subspace spanned by $|0\rangle$ and $|1\rangle$, without destroying superpositions in the latter subspace [19]. In this error model, transmission through the channel is considered successful if an arbitrary initial quantum state can be perfectly recovered (or the entanglement with R can be maintained), which can obviously be achieved if one uses a quantum e-erasure correcting code with $e=pn$, that is, a code that allows any pattern of e qubits of each codeword to be erased. The *rate* (i.e., the average number of logical qubits transmitted with arbitrarily high fidelity per physical qubit) of a channel subjected to a p-bounded fraction of error is thus equivalent to the rate of an $((n,k))$ quantum code correcting $e=pn$ erasures. The rate of an $((n,k))$ code, i.e., a code mapping k logical qubits into codewords of n qubits, is defined as $R=k/n$. Consequently, an upper bound on the rate of quantum codes is simply equivalent to an upper bound on the rate of a channel with this particular error model (or an upper bound on the capacity, which is the highest achievable rate through the channel). When considering a channel, k is simply the source entropy S of the joint channel (i.e., the number of arbitrary qubits that are sent).

It is known that an upper bound on the Hamming distance of *nondegenerate* quantum codes with fixed n and k can be derived from ''sphere-packing'' considerations [20]. However, as a bound on the rate (or capacity) of a quantum channel involves a maximization over all coding schemes, including those based on degenerate codes (which have been shown to exceed the Hamming bound [21]), only the bounds which are valid for all quantum codes are applicable to channels. As proven in a previous paper [15], an upper bound on the rate of (nondegenerate *and* degenerate) quantum codes can be derived using entropic considerations only. This is the quantum Singleton bound: $k\leqslant n-2e$ (see also [22]). Translated in the channel language, this implies that an upper bound on the rate (and therefore the capacity) of a lossless ($L=0$) channel subjected to a p-fraction of erasures is

$$R\leqslant 1-2p. \tag{3.1}$$

For completeness, we summarize the proof of the Singleton bound given in [15]. The basic idea of the proof will be useful in the following, when considering other channels.

As pictured in Fig. 4, for each pattern of erased qubits Q_e, the entropic condition $S(R{:}Q_e)=0$ must be fulfilled, so that the unerased qubits Q_u emerge from a lossless channel. This implies that the full entanglement of the codeword Q (with respect to R) must be ''concentrated'' in the unerased qubits Q_u:

$$S(R{:}Q_u)=S(R{:}Q)=2S(R). \tag{3.2}$$

[4] When $n\rightarrow\infty$, the number of erasures pn can be considered as an integer without loss of generality.

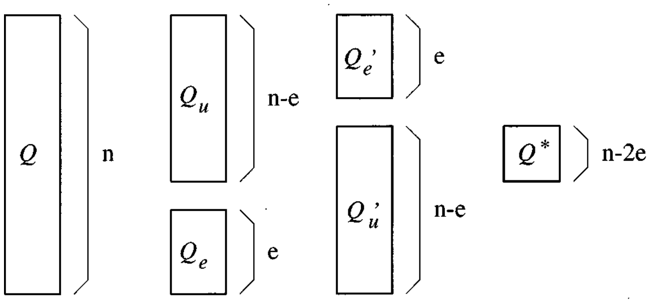

FIG. 4. Schematic representation of two possible partitions of Q into an erased piece Q_e (or Q_e') and unerased piece Q_u (or Q_u'). The ''overlap'' between the unerased pieces in both partitions is denoted by Q^*. The entropic erasure-correction condition for the first partition is $S(R{:}Q_e)=0$, while the condition for the second one is $S(R{:}Q_e')=0$.

Using the fact that the joint state of RQ_uQ_e is pure, along with the subadditivity of quantum entropies, we have

$$\begin{aligned} S(R{:}Q_u) &= S(R)+S(Q_u)-S(RQ_u) \\ &= S(R)+S(Q_e'Q^*)-S(Q_e) \\ &\leqslant S(R)+S(Q_e')+S(Q^*)-S(Q_e), \end{aligned} \tag{3.3}$$

where we have divided the unerased qubits Q_u into another pattern of e qubits, Q_e', and the remaining piece of $n-2e$ qubits, Q^*. Equations (3.2) and (3.3) provide the inequality

$$S(R)-S(Q^*)\leqslant S(Q_e')-S(Q_e). \tag{3.4}$$

Since this reasoning is symmetric in Q_u vs Q_u', the inequality corresponding to the division of Q_u' into Q_e and Q^*,

$$S(R)-S(Q^*)\leqslant S(Q_e)-S(Q_e'), \tag{3.5}$$

must also be satisfied. Combining these two inequalities yields the condition

$$S(R)\leqslant S(Q^*)\leqslant n-2e, \tag{3.6}$$

where the upper bound on $S(Q^*)$ simply results from the dimension of the Hilbert space of Q^*. Since the encoding of a k-qubit arbitrary state requires that $S(R)=S(Q)=k$, the above condition implies the quantum Singleton bound

$$k\leqslant n-2e. \tag{3.7}$$

The physical meaning of the Singleton bound is that, in order to have Q_e independent of R and, at the same time, $Q_u' \equiv Q_eQ^*$ fully entangled with R, a minimum Hilbert space for Q^* (minimum number of qubits) is necessary in order to accommodate the source entropy k.

B. Quantum channel subject to a *p*-bounded fraction of errors

Another possible error model for a quantum channel is the case where a fraction of p qubits (at most) are altered by interacting with an environment. The difference with the previous error model is that the location of the errors is unknown (by contrast with erasures), i.e., there is no ''flag'' indicating which are the qubits that have been altered.

Since a channel where the fraction of errors is bounded can be made lossless ($L=0$) using an *error*-correcting code (just as for erasures in the preceding section), it is enough to use the correspondence between error-correcting codes and erasure-correcting codes to derive an upper bound on the rate of perfect transmission of quantum information in such a channel. In analogy with the classical situation, one can show that any code that corrects t errors is also able to correct up to $e=2t$ erasures [23]. This enables us to reuse the result of the preceding section simply by replacing p by $2p$. Thus, we obtain the Singleton upper bound on the rate (or capacity) of a lossless ($L=0$) channel with a fixed fraction p of errors:

$$R\leqslant 1-4p. \tag{3.8}$$

(This bound, or rather the fact that the rate of a code is vanishing at $p=1/4$, is originally due to Knill and Laflamme [22].) To our knowledge, the only stronger bounds on R for quantum codes that have been displayed for some special cases are as follows. (i) $R\leqslant H[1/2+\sqrt{2p(1-2p)}]$ for additive (or stabilizer) codes, where H stands for the dyadic Shannon entropy [18]. This bound is based on an upper bound on classical linear codes.[5] (ii) $R=0$ for $p\geqslant 1/6$ for all quantum codes [25]. It is worthwhile looking for improvement of Eq. (3.8) using an entropic approach as presented above.

C. Quantum erasure channel

We now consider a quantum erasure channel with erasure probability p (see, e.g., [19]). In such a channel, each transmitted qubit has a probability p of being *erased* (and detectable at the output as an ''erased'' qubit). We are interested in the maximum rate of quantum information transmission that can be achieved by this channel. More precisely, our aim is to derive an upper bound on this rate using an entropic approach. It will appear that such an entropic bound exactly coincides with the capacity of a quantum erasure channel recently displayed in Ref. [19].

The central point of the reasoning is to describe the joint ''probabilistic'' channel (with n qubits at input and output) as a superposition of ''binomial'' n-bit channels (defined below). This allows us to make use of the convexity of the von Neumann mutual entropy in the output of a channel for a fixed input (see Sec. II C) in order to derive an upper bound on R. We consider a family of ''binomial'' n-bit channels C (labeled by the index c), each characterized by a pattern of e erased qubits, $Q_e(c)$, and the complementary pattern of $n-e$ unerased qubits, $Q_u(c)$. The probabilistic channel of interest here corresponds to a probabilistic use of these channels with a *binomial* distribution. More precisely, each channel c with e erased qubits is associated with a probability (or weight) $w_c=p^e(1-p)^{n-e}$, and there are obviously $\binom{n}{e}$ distinct channels with e erased qubits. The superposition means, physically, that the resulting probabilistic joint channel consists in using one of these 2^n distinct ''bi-

[5] Some slightly stronger bounds have been recently obtained in Ref. [24].

nomial'' n-bit channels with the appropriate probability. Thus, for a given n-bit input of the channel, say ρ, the output can thus be written as a convex combination

$$\rho' = \sum_c w_c \rho'_c \quad \text{with} \quad \sum_c w_c = 1, \tag{3.9}$$

where w_c is the weight of the cth (binomial) channel in the superposition, and ρ'_c is the output for that channel c. The convexity of the von Neumann mutual entropy in the output (for a given input) for the overall probabilistic channel implies that

$$\underbrace{S(R{:}Q')}_{I} \leq \sum_c w_c \underbrace{S(R{:}Q'_c)}_{I_c} \tag{3.10}$$

where $I = S(R{:}Q')$ is the quantum mutual entropy between the reference R and the output Q' of the *joint* (n-bit) quantum erasure channel, while $I_c = S(R{:}Q'_c)$ stands for the mutual entropy of the output of the cth channel, Q'_c, with respect to R. Note that we have $S(R{:}Q'_c) = S(R{:}Q_u(c))$ since only the unerased qubits of the cth channel contribute to mutual entanglement with R (the erased qubits are independent of R). As explained in Sec. II C, the mutual von Neumann entropy $S(R{:}Q') = S(R) - S(R|Q')$ is convex in the output (for a fixed input) because the conditional entropy $S(R|Q')$ is concave in a convex combination of $\rho_{RQ'}$ [i.e., a convex combination of quantum channels acting on a fixed input, as shown in Eq. (3.9)].

It is convenient to group the channels C into several classes, according to the number of erased qubits, e. Using Eq. (3.10), can then write an upper bound on the processed quantum mutual information in the joint (n-bit) erasure channel of probability p as

$$I(p) \leqslant \sum_{e=0}^{n} (1-p)^{n-e} p^e \sum_c S(R{:}Q_u(c)), \tag{3.11}$$

where the sum over c spans the $\binom{n}{e}$ channels where e qubits are erased. Before deriving a simpler expression of this upper bound using the Singleton bound, let us show that a simple relation between $I(p)$ and $I(1-p)$ can be obtained from Eq. (3.11). First, note that

$$\begin{aligned} S(R{:}Q'_c) &= S(R{:}Q_u(c)) = S(R) + S(Q_u(c)) - S(Q_e(c)) \\ &= 2S(R) - S(R{:}Q_e(c)), \end{aligned} \tag{3.12}$$

where we have used the fact that RQ_uQ_e is in a pure state for each channel C. Equation (3.11) can then be rewritten as

$$I(p) \leqslant 2S(R) - \sum_{e=0}^{n} (1-p)^{n-e} p^e \sum_c S(R{:}Q_e(c)). \tag{3.13}$$

The second term on the right-hand side of Eq. (3.13) can be interpreted as an upper bound on $I(1-p)$, i.e., the information processed through a ''dual'' erasure channel of probability $1-p$ where the erased qubits are replaced by unerased qubits and conversely. As a consequence, remembering that $S(R) = k$ if an arbitrary k-bit quantum state is sent in the channel (i.e., the source entropy of the n-bit channel is k bits), we have

$$I(p) + I(1-p) \leqslant 2k \tag{3.14}$$

for all n. Using $I(p) + L(p) = 2k$, the corresponding relation for the quantum losses of the dual (with probability p and $1-p$) erasure channels is

$$L(p) + L(1-p) \geqslant 2k. \tag{3.15}$$

[Remember that $0 \leqslant L(p) \leqslant 2k$.] This implies that a *necessary* condition for having a perfect channel at probability p, i.e., $L(p) \equiv 0$, is that $L(1-p) = 2k$, i.e., that the dual channel at probability $1-p$ is ''fully erasing'' (no information at all—either classical or quantum—is processed through it). Another way of expressing this condition is by writing a lower bound on the loss of the erasing channel

$$L(p) \geqslant I(1-p). \tag{3.16}$$

Only if no information is transmitted through the erasure channel of probability $1-p$, or if $I(1-p) = 0$, is it possible that the loss of the erasing channel of probability p vanishes. This is obviously compatible with $L(0) = I(1) = 0$. Equation (3.15) also implies that $L(1/2) \geqslant k$, so that the quantum erasure channel with probability 1/2 cannot be lossless for a nonzero source entropy. (The fact that it actually has a vanishing capacity—or a maximum loss—will be shown below. This result can also be derived from an argument based on the impossibility of cloning, as shown in Ref. [19].)

Let us now derive a general expression for an upper bound on the mutual entropy of the n-bit channel (or, equivalently, a lower bound on the overall loss). Using Eqs. (3.11) and (3.12), we have

$$I(p) \leqslant S(R) + \sum_{e=0}^{n} (1-p)^{n-e} p^e \sum_c [S(Q_u(c)) - S(Q_e(c))]. \tag{3.17}$$

We first rewrite Eq. (3.17) as a summation up to $n/2$ (we assume here that n is even), by combining each channel c in this sum with its dual channel where erased qubits are unerased and conversely:

$$\begin{aligned} I(p) \leqslant S(R) + \sum_{e=0}^{n/2} &[(1-p)^{n-e} p^e - (1-p)^e p^{n-e}] \\ &\times \sum_c [S(Q_u(c)) - S(Q_e(c))]. \end{aligned} \tag{3.18}$$

(Note that the term with $e = n/2$ is vanishing.) We now follow the reasoning that we used earlier to derive the Singleton bound, and group the channels in pairs (c and c') which ''overlap'' in $n-2e$ qubits denoted by Q^* (see Fig. 4). Thus, for a given value of e, we have to calculate the sum of terms $[S(Q_u(c)) - S(Q_e(c))]$ for channels c and c'. We have

$$
\begin{aligned}
S(Q_u)-S(Q_e)+S(Q_u')-S(Q_e') \\
&= S(Q_e'Q^*)-S(Q_e)+S(Q_eQ^*)-S(Q_e') \\
&\leqslant 2S(Q^*)\leqslant 2(n-2e), \qquad (3.19)
\end{aligned}
$$

where the last inequality reflects the limitation on $S(Q^*)$ imposed by the dimension of the Hilbert space of Q^*. Thus each *pair* of terms in the summation over c can be bounded from above by $2(n-2e)$, so that a simple calculation yields an upper bound on the mutual entropy of the joint (n-bit) channel

$$
\begin{aligned}
I(p) &\leqslant S(R)+\sum_{e=0}^{n}\binom{n}{e}(1-p)^{n-e}p^e(n-2e) \\
&\leqslant k+n(1-2p) \qquad (3.20)
\end{aligned}
$$

or a lower bound on the overall loss of the joint channel

$$L(p)\geqslant k+n(2p-1). \qquad (3.21)$$

This results in a lower bound on the *average* loss (per qubit) for a quantum erasure channel of probability p as a function of the rate $R=k/n$,

$$l(p)=\frac{L(p)}{n}\geqslant R+2p-1. \qquad (3.22)$$

(Note that this is only a lower bound on the loss, which is, by definition, a non-negative quantity.) This inequality is consistent with $l(p)+l(1-p)\geqslant 2R$ [cf. Eq. (3.15)]. It implies that a vanishing average loss (i.e., the reliable transmission of information through the n-bit probabilistic channel) is only possible if the rate

$$R\leqslant 1-2p. \qquad (3.23)$$

Thus, a quantum erasure channel with $p=1/2$ (i.e., if the channel is erasing 50% of the qubits) has a zero capacity. Equation (3.23) confirms the linear interpolation[6] between the 50%-erasure channel (for which the capacity is zero) and a noiseless channel (for which the capacity is 1) that was used in Refs. [6,19]. In addition, since it is shown in Refs. [6,19] that this upper bound coincides with a lower bound obtained from one-way random hash coding, Eq. (3.23) therefore describes the *exact* capacity of the quantum erasure channel $C=1-2p$ [19].

D. Quantum depolarizing channel

We now consider a quantum depolarizing channel with error probability p. In this channel, each qubit interacts with the environment such that it undergoes a bit-flip (σ_x rotation), a phase-flip (σ_z rotation), or the combination of both (σ_y rotation) with probability $p/3$ each.

In Secs. III A and III B, we have seen that the simple connection between quantum error- and erasure-correcting codes provides a trivial relation between the resulting upper bounds on the rate of channels subjected to a bounded fraction of errors or erasures. Unfortunately, there is no such simple relationship when comparing the quantum erasure and depolarizing channels. As a matter of fact, the upper bound on error-correcting codes, Eq. (3.8), is not immediately applicable to the quantum depolarizing channel. Such a situation results from the fact that the definition of an error-correcting code requires that *all* error patterns (of at most $t=pn$ errors) are perfectly corrected, while a rate will be said to be attainable through a channel (i.e., it is below the quantum channel capacity) whenever the fraction of uncorrected error patterns is asymptotically vanishing (for $n\rightarrow\infty$). Still, the reasoning used to calculate the Singleton bound on the quantum erasure channel is applicable to the quantum depolarizing channel as well.

Assume that an individual qubit, Q, is initially entangled with the reference, so that RQ is in the state $|\Psi_{RQ}\rangle$ (for example the singlet state). After processing by the channel, the system RQ' is in a Werner state [26] of ‘‘entanglement fidelity’’ $F=1-p$, that is, the mixed state:

$$\rho_{RQ'}=(1-4p/3)|\Psi_{RQ}\rangle\langle\Psi_{RQ}|+4p/3\left(\frac{\mathbf{1}_R}{2}\otimes\frac{\mathbf{1}_Q}{2}\right). \qquad (3.24)$$

In other words, the qubit emerges at the output of the channel either in a random state (having totally lost the entanglement with R) with probability $4p/3$, or in its intact original state (fully entangled with R) with probability $1-4p/3$. When $p=3/4$, the channel is 100% depolarizing, i.e., its output is random.

In the joint n-bit channel, each qubit undergoes the above evolution independently of the other ones. As before, the resulting n-bit probabilistic channel can be described as a superposition of binomial channels, in which each qubit is either kept unchanged or ‘‘randomized.’’ The distribution of the underlying channels is thus a binomial one, just as in the preceding section, the only difference being that p is replaced here by $4p/3$. The entire calculation of the preceding section can then be repeated, because the ‘‘randomization’’ or ‘‘erasure’’ of a qubit is equivalent as far as the mutual entropy with R is concerned. Indeed, for a channel c, we have

$$I_c=S(R{:}Q_c')=S(R{:}Q_u(c)), \qquad (3.25)$$

where $Q_u(c)$ correspond to qubits that are not randomized (rather than not erased) in channel c. This is obvious because the randomized qubits [in state $(|0\rangle\langle 0|+|1\rangle\langle 1|)/2$] are independent of R, just as the erased qubits (in state $|2\rangle$).

First, in analogy with Eq. (3.15), we have

$$L(p)+L(3/4-p)\geqslant 2k \qquad (3.26)$$

implying that $L(3/8)\geqslant k$, so that the quantum depolarizing channel with probability $p=3/8$ cannot be lossless (in fact, it has a vanishing capacity, as shown below). Equivalently, we have $L(p)\geqslant I(3/4-p)$, showing that the $p=3/4$ channel cannot transmit classical or quantum information, i.e.,

[6] In Ref. [6], it is shown that the capacity of a composite channel (which is a convex combination of a perfect and an imperfect quantum channel) cannot exceed the appropriately averaged capacity of these two component channels. In other words, the quantum capacity cannot be superadditive when ‘‘mixing’’ a perfect channel with an imperfect one.

$I(3/4)=L(0)=0$. The resulting Singleton lower bound on the *average* loss (per qubit) is

$$l(p)=\frac{L(p)}{n}\geqslant R+8p/3-1 \tag{3.27}$$

in analogy with Eq. (3.22). Consequently, the quantum depolarizing channel can have a vanishing average loss (i.e., allow an asymptotically reliable transmission of information by using blocks of n qubits) at the condition that

$$R\leqslant 1-8p/3 \tag{3.28}$$

as originally shown in Ref. [6]. Thus the quantum depolarizing channel with $p=3/8$ has a zero capacity for the transmission of quantum information. (Note that such a channel corresponds in fact to a 50%-depolarizing channel, where 50% of the qubits are replaced by a random qubit. This channel can obviously not have a nonzero capacity, as a consequence of the no-cloning theorem [6].) As for the quantum erasure channel, a linear interpolation between the perfect channel and the 50%-depolarizing channel can be used (and also results independently from our reasoning). Note that Eq. (3.28) yields only an upper bound on the capacity of the quantum depolarizing channel, which is provably not achievable (in contrast with the equivalent bound for the quantum erasure channel). Indeed, a tighter bound for the depolarizing channel has been obtained very recently [27],

$$R\leqslant 1-4p, \tag{3.29}$$

which is based on the Buzek-Hillery universal cloning machine [28]. While Eq. (3.29) happens to be equivalent to Eq. (3.8), there appears to be no direct relation between them. A simple intuitive reason why this bound is stronger than Eq. (3.28) can be understood by realizing that the two quantum channels underlying the universal cloning machine (from the single input to both outputs) cannot be described classically.[7] Indeed, when tracing over one of the outputs of the universal cloning machine, the other output appears to emerge from a $p=1/4$ (or $F=3/4$) channel, i.e., a 33%-depolarizing channel. This looks as if the qubit was sent with probability 2/3 to each output of the cloning machine, which is obviously not understandable in classical terms. Only a quantum superposition, involving the cloning machine and both outputs, can account for this situation and results in a stronger upper bound [27]. (and related entropic bounds) The information-theoretic analysis of quantum cloning (and related entropic bounds) will be the subject of further investigation.

To our knowledge, the only stronger upper bound on the capacity (for a restricted range of p values) of the depolarizing channel is $R\leqslant 1-H(p)$, based on a connection between quantum additive (or stabilizer) codes and classical linear codes [18].

IV. QUANTUM CHANNEL WITH AN AUXILIARY CLASSICAL CHANNEL

In this section, we consider a quantum channel which is supplemented with a classical channel (assumed to be noiseless and of unlimited classical capacity). We are interested in calculating the Singleton upper bound on the rate of reliable $(L=0)$ transmission of *quantum* information through a noisy quantum channel, knowing that a classical side channel can be used simultaneously for forward communication only. In particular, we aim at analyzing how our quantum information-theoretic formalism accounts for the property that such a classical (one-way) communication channel does *not* increase the quantum capacity of the noisy channel [6].

A. Entropic treatment of the channel

We first consider the problem in the language of quantum codes, following closely Ref. [15]. As explained earlier, the result will then be immediately applicable to channels with an error model where the fraction of errors or erasures is bounded. Also, the expression of a ''probabilistic'' joint channel as a binomial superposition of underlying channels will then yield the corresponding bounds for the quantum erasure or depolarizing channel. We start with a ''logical'' system L (logical words), which is initially entangled with the reference R (the initial mutual entropy being $2k$, so that k arbitrary qubits are transmitted). As before, the system RL is initially in a pure entangled state $|\Psi_{RL}\rangle$. We assume that the encoding operation on L includes a partial measurement, so that the encoder has a quantum and a classical output, as shown in Fig. 5.

Note that the encoding process can be described in terms of a unitary transformation applied on L (supplemented with an ancilla initially in a $|0\rangle$ state) without any projection operators, so that the classical output is simply described as a

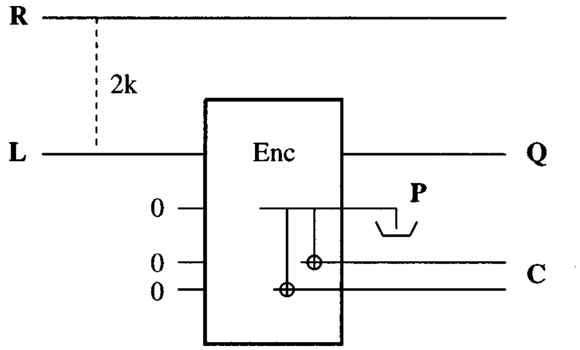

FIG. 5. Schematic representation of the ''encoder'' performing a unitary transformation on the ''logical'' system L (initially entangled with R with a mutual entropy $2k$). The outputs are the n ''physical'' qubits Q, the classical bits C, and the ''precursor'' P (''microscopic'' classical bits, before amplification). Each classical bit can be thought of as a set of qubits that are classically correlated when tracing over P, appearing then as a collective classical variable (see [29]).

[7] Remember that a standard no-cloning argument to show that the rate vanishes at $p=3/8$ is based on a classical machine that is transmitting an input qubit to one of two outputs with probability 1/2, a random qubit being sent on the other output, which results in two 50%-depolarizing channels [6]. Clearly, this is the most constraining bound on R that can be constructed by use of such a ''classical cloning.''

mixture of orthogonal quantum states. More precisely, we have a quantum output Q (consisting of n ''physical'' qubits), P, which is the ''precursor'' of the classical variable, and C, which represents the *amplified* classical variable. In general, C is a discrete variable of arbitrary dimension, but we often use the term classical bit (two-dimensional variable) in the following. Each classical bit C can be thought of as a set of qubits that become *classically* correlated when tracing over the corresponding ''precursor,'' and can then be viewed as a collective variable. The resulting *amplification* gives to C the appearance of a classical variable (see [29]). It is convenient to keep the classical output C separate from its precursor P for reasons that will appear later.

To fix the ideas, let us write the intermediate wave function of RQP before amplification:

$$|\Psi_{RQP}\rangle=\sum_i \sqrt{p_i}\;|\psi^i_{RQ}\rangle\otimes|\phi^i_P\rangle. \tag{4.1}$$

Here, the $|\phi^i_P\rangle$ form a set of orthogonal states for P, and the $|\psi^i_{RQ}\rangle$ are orthogonal states of RQ. Equation (4.1) is simply the Schmidt decomposition of $|\Psi_{RQP}\rangle$ divided in RQ versus P, and implies $S(RQ)=S(P)$ before amplification. It is important to note that the orthogonal states $|\phi^i_P\rangle$ for P correspond to the *classical* information that will be amplified later (the amplification of P will be performed in this basis), the classical variable being distributed according to p_i. Equation (4.1) is the most general expression of the output of the encoder (before amplification of P) if we require that its overall operation is unitary (the joint state of RQP must be pure). We also have obviously $S(R{:}QP)=S(R{:}L)=2k$ as a result of the conservation of mutual entropy under a local unitary operation on both subsystems [12].

The amplification of the precursor (symbolically represented by the XOR gates inside the encoder) in the $|\phi^i_P\rangle$ basis gives rise to a total wave function for the system $RQPC$ (at the output of the encoder) of the form

$$|\Psi_{RQPC}\rangle=\sum_i \sqrt{p_i}\;|\psi^i_{RQ}\rangle\otimes|\phi^i_P\rangle\otimes|\phi^i_C\rangle, \tag{4.2}$$

where the $|\phi^i\rangle$ correspond to an orthogonal set of states for both P and C. The classical information in P has been ''amplified,'' so that the precursor P and the collective set of qubits C are fully *classically* correlated when tracing over the remaining variables. The entropy of the classical channel is $S(C)=S(P)=H[p_i]$. In fact, P and C are interchangeable, but we need to keep them separate to account for the fact that amplifying the classical bits, i.e., tracing over the precursor P, results in a mixed state for the systems R, Q, and C. The density matrix for the system RQC is given by

$$\rho_{RQC}=\sum_i p_i|\psi^i_{RQ}\rangle\langle\psi^i_{RQ}|\otimes|\phi^i_C\rangle\langle\phi^i_C|. \tag{4.3}$$

This can be viewed as a *classical* mixture of orthogonal states of RQ and C. Thus, conditionally on the classical bits C, the system RQ is in a pure (generally entangled) state $|\psi^i_{RQ}\rangle$. The question now will be, roughly speaking, to determine under which circumstances the quantum output Q (possibly altered by decoherence or partially erased) retains the full entanglement with R when it is *supplemented* with the classical information C. In other words, the question will be whether decoding *using* C allows to perfectly recover the logical words (and is more effective than in the absence of classical information).

We define the two parameters k and c as

$$S(R)=k, \tag{4.4}$$

$$S(C)=S(P)=S(CP)=c, \tag{4.5}$$

where the second equation is due to the fact that C and P are *fully* classically correlated, i.e., $S(C{:}P)=c$. The parameter $c=H[p_i]$ simply represents the Shannon entropy processed by the classical side channel. As before, the parameter k stands for the number of logical qubits (i.e., k-qubit arbitrary states are encoded), or, equivalently, the source entropy when considering a quantum channel.

A necessary condition for the global (classical + quantum) channel to be lossless is clearly that the amplification of P is a *lossless* process, that is, does not destroy the quantum coherence of the logical words. More precisely, the constraint we must express is that the channel $L\rightarrow QC$ (considered as a quantum channel) is *lossless*, that is,

$$S(R{:}P)=0, \tag{4.6}$$

where P plays the role of an ''environment'' for this channel. This means that, when amplifying the classical bits C (ignoring the precursor P) no entanglement with R is lost. In other words, the mutual entropy $2k$ with R is entirely found in QC, i.e., $S(R{:}QC)=2k$, so that no entanglement with R leaks out in P when tracing over P. [This is so even though the joint system QPC is fully entangled with R, i.e., $S(R{:}QPC)=2k$.] The condition (4.6) implies that

$$S(QC)=S(RP)=S(R)+S(P)-S(R{:}P)=k+c. \tag{4.7}$$

Thus, the parameters k and c fully determine the ternary entropy diagram[8] for R, QC, and P, as shown in Fig. 6. Note that C and P are interchangeable, so that we have also $S(QP)=k+c$.

A fourth parameter is necessary to fully describe the entropies of the 4-partite system (R, Q, C, and P):

$$S(Q)=s. \tag{4.8}$$

It corresponds to the von Neumann entropy of the quantum input of the noisy channel (or quantum output of the encoder). Grouping C and P, we can also display the ternary entropy diagram for R, Q, and CP, using the fact that the joint system is in a pure state (cf. Fig. 7). It shows that neither the quantum output Q nor the classical one CP *before* amplification (i.e., including the precursor) is unentangled with R. In short, the situation is that P alone is

[8] The ternary diagram of a tripartite system in a pure state is determined in general by three parameters, for example the reduced entropy of each of the three components [12].

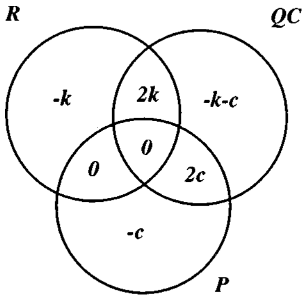

FIG. 6. Entropy diagram characterizing the reference R, the system QC (quantum and classical output of the encoder), and P (precursor or the classical bits). The condition $S(R{:}P)=0$ means that the full mutual entropy $2k$ with R is found in QC and does not leak out when tracing over P, so that the amplification is *lossless*. The two parameters are $S(R)=k$ and $S(P)=c$.

unentangled with R (it can be traced over without altering the entanglement with R), while P and C together are entangled with R. Moreover, even though the classical bits can be amplified without losing coherence, the classical information C is in general necessary (together with Q) in order to recover the initial entanglement of L with respect to R, as implied by $S(R{:}QC)=2k$.

We now want to describe the entropic situation *after* amplification of the classical bits, i.e., after tracing over the precursor P. The density matrix for the system RQC (or equivalently for the system RQP) is given by

$$\rho_{RQC}=\sum_i p_i|\psi^i_{RQ}\rangle\langle\psi^i_{RQ}|\otimes|\phi^i_C\rangle\langle\phi^i_C|. \tag{4.9}$$

The system RQ is thus *classically* correlated with C, and we have

$$S(RQ{:}C)=S(RQ)=S(C)=c. \tag{4.10}$$

For each value i of the classical bits (occurring with probability p_i), RQ is in a given pure (generally entangled) state $|\psi^i_{RQ}\rangle$. Now, using $S(QP)=S(QC)=k+c$, we can show that R is *independent* of C:

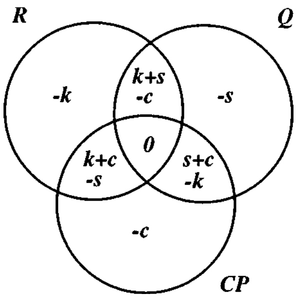

FIG. 7. Entropy diagram characterizing the reference R, the quantum system Q (output of the encoder), and PC (the classical output and precursor *before* amplification). The three parameters are $S(R)=k$, $S(Q)=s$, and $S(CP)=c$.

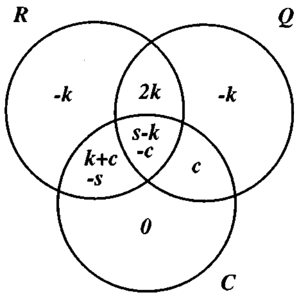

FIG. 8. Entropy diagram characterizing the reference R, the quantum system Q (output of the encoder), and the classical output C *after* amplification. The joint system is in a *mixed* state of entropy $S(RQC)=c$. Note that R is independent of C (i.e., C contains no information about the encoded logical word), while RQ is (fully) classically correlated with C.

$$\begin{aligned}S(R{:}C)&=S(R)+S(C)-S(RC)\\&=S(R)+S(C)-S(QP)\\&=k+c-(k+c)=0.\end{aligned} \tag{4.11}$$

This means that, if the amplification of the classical bits is *lossless* (that is, cannot result in an irrecoverable loss of mutual entanglement with R), then the ''amplified'' classical bits C must be statistically *independent* of R. In other words, no information about the encoded logical word is found in the ''amplified'' classical bits C even though C is in general necessary to recover the entanglement with R. Using $S(QC)=S(R)+S(C)$, the mutual entropy between Q and C can be written as

$$S(Q{:}C)=S(Q)+S(C)-S(QC)=S(Q)-S(R)=s-k. \tag{4.12}$$

Thus, the quantum output Q is in general *not* independent of the classical bits C (in contrast with R). This simply means that, in general, the encoder can introduce some extra entanglement between Q and CP, additionally to the initial entanglement $2k$ between Q and R, giving rise to a nonvanishing mutual entropy. However, this additional entanglement is useless as far as the transmission of quantum information is concerned, and we will see below that the interesting situation corresponds to $s=k$, in which case $S(Q{:}C)=0$. Finally, it is easy to see that RQC is in a mixed state of entropy $S(RQC)=S(P)=c$. The latter condition, together with Eqs. (4.4), (4.5), (4.8), (4.10), (4.11), and (4.12), fully describes the entropies of the tripartite system RQC after tracing over P (i.e., after amplification). The corresponding ternary entropy diagram is presented in Fig. 8.

In summary, we have $S(R)=k$, $S(Q)=s$, $S(C)=c$, $S(RQ)=c$, $S(RC)=k+c$, $S(QC)=k+c$, and $S(RQC)=c$, which implies that

$$\begin{aligned}S(R{:}Q{:}C)=&S(R)+S(Q)+S(C)-S(RQ)-S(RC)\\&-S(QC)+S(RQC)=s-k-c.\end{aligned} \tag{4.13}$$

As visible from this diagram, the system RQ is in a pure entangled state *conditionally* on the amplified classical variable C, i.e., $S(RQ|C)=0$, with a characteristic diagram

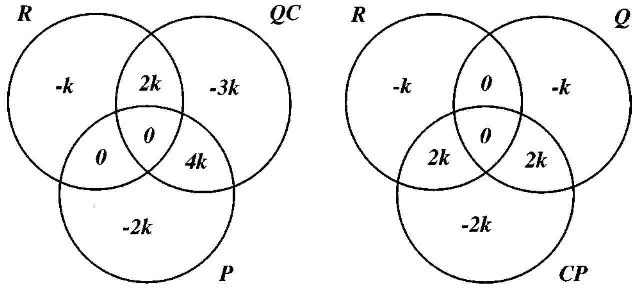

FIG. 9. Entropy diagrams *before* amplification of the classical bits in the case where the processed classical information is maximum ($c=2s=2k$). As before, $S(R{:}P)=0$, so that the amplification of the precursor P does not affect quantum coherence. Note that $S(R{:}Q)=0$, so that the quantum output Q can be erased if no amplification is performed.

$(-k,2k,-k)$. Also, RQ is classically correlated with C, with a diagram $(0,c,0)$, so that the simultaneous knowledge of R and Q yields C, i.e., $S(C|RQ)=0$. Finally, it is easy to check that the diagram for R vs C is $(k,0,c)$, i.e., C is independent of R.

Several inequalities relating the parameters k, c, and s must be satisfied. First, the subadditivity of entropies implies that

$$S(R{:}Q)=k+s-c\geqslant 0, \tag{4.14}$$

$$S(Q{:}C)=s-k\geqslant 0. \tag{4.15}$$

Similarly, the strong subadditivity of entropies implies that

$$S(R{:}C|Q)=k+c-s\geqslant 0. \tag{4.16}$$

These inequalities can be summarized as

$$0\leqslant s-k\leqslant c\leqslant s+k \tag{4.17}$$

implying, namely, that $0\leqslant c\leqslant 2s$. The two limiting cases are (i) $c=0$ and $s=k$, which corresponds to a quantum channel without a classical side channel; and (ii) $c=2s=2k$, which corresponds to the situation where the entropy of the classical channel is maximum. In the following, we will focus on case (ii), that is, when a maximum amount of classical information is processed, since it is supposedly the case where the classical side channel might help the transmission of quantum information the most.

We display in Figs. 9 and 10 the entropy diagrams corresponding to the limiting case $c=2k=2s$. Note that $S(R{:}Q)=0$ as shown in Fig. 9, so that the channel $L\rightarrow CP$ is lossless. Thus, the entire entanglement with R is retained in the *unamplified* classical variable CP. We will discuss this below. Figure 10 implies that $S(R{:}C)=S(Q{:}C)=0$, i.e., Q and R are both *independent* of the classical variable C. This emphasizes the fact that R and Q play exactly the same role in this limiting case (ii). The peculiar feature here is that Q and R together are fully correlated with C, according to the diagram $(0,2k,0)$, although Q or R taken separately is independent of C. In other words, the classical variable C contains the information about which entangled state RQ is in (i.e., the mutual entropy $2k$) while it contains no information about Q or R alone.

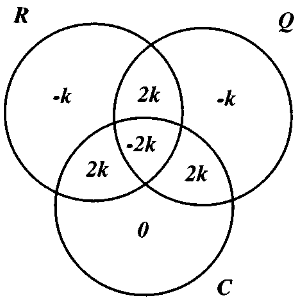

FIG. 10. Entropy diagram *after* amplification of the classical bits in the case where the processed classical information is maximum ($c=2s=2k$). Here, R and Q play the same role. The classical variable C contains the information about which entangled state RQ is in, while R or Q alone is independent of C. This diagram with $k=1$ plays a crucial role in the entropic description of superdense coding and teleportation, as shown elsewhere [30].

The diagram in Fig. 10 plays an important role in the information-theoretic description of quantum teleportation and superdense coding, as shown in a further work [30]. In the special case where $k=1$ (no block coding is used), it describes quantum teleportation in the following sense: Q is the particle that is initially sent to Bob (one half of the Bell state shared with Alice), and C are the 2 classical bits that Alice sends to Bob. If the teleported particle (here L) is initially entangled with R, so that RL is in a Bell state, then Q ends up in an entangled state with R which is one of the four Bell states, *conditionally* on C. (Teleportation is completed by having Bob applying a unitary transformation on Q that is specified by C.) Thus, the 2 classical bits code for one of the four Bell states, and the corresponding diagram is shown in Fig. 10 with $k=1$. The amplification of the classical bits by Alice does *not* destroy coherence, since we have $S(R{:}P)=0$, and bringing Q and C together yields the initial entanglement with respect to R, i.e., $S(R{:}QC)=2$. The nonclassical feature here is that the latter equation can be satisfied even though $S(R{:}Q)=0$ holds at the same time, that is, the particle that Bob received initially is independent of R (this must be true as a consequence of causality). Thus, the entire entanglement with R is carried by C, as reflected by $S(R{:}C|Q)=2$. This will be discussed in more details elsewhere [30].

Note finally that the vanishing mutual entropy between R and Q implies that no entanglement with R is found in Q alone. Therefore, the quantum output Q can be erased without losing the entanglement with R, provided that the classical variable is not amplified (by keeping CP). In other words, since $S(R{:}Q)=0$, the knowledge of the classical bits C (along with the precursor P) is sufficient to recover L, even in the absence of Q. For example, in teleportation, the *unamplified* classical bits alone are enough to teleport an arbitrary state, so that the qubit Q that Bob received initially can be erased. (Of course, this is unrealistic, since one never has access to all the microscopic degrees of freedom making the classical bits. Tracing over one of them is enough to lose the quantum information if Q is erased.)

B. Singleton bound on a quantum channel with a classical side channel

Let us now repeat the reasoning which results in the Singleton bound on a quantum code (Sec. III), but taking into account the classical auxiliary channel. Thus, we assume again that the quantum output Q of n qubits is partitioned into an erased piece Q_e (of e qubits) and an unerased one Q_u (of $n-e$ qubits). We are seeking for a necessary condition for the possibility of recovering the erasure of Q_e when the decoder has access to the classical information C (so the decoding operation can be conditional on C). As before, we consider two different partitions of Q (see Fig. 4), and express a lower bound on the entropy of the ''overlap'' Q^*. The basic constraints (which must be satisfied simultaneously) are

$$S(R{:}Q_eP)=0, \tag{4.18}$$

$$S(R{:}Q_e'P)=0, \tag{4.19}$$

expressing the fact that no entanglement (with respect to R) is lost when amplifying the classical bits *and* erasing Q_e (or Q_e'). Equivalently, the full initial entanglement of L must be ''squeezed'' into Q_u (or Q_u') and C:

$$S(R{:}Q_uC)=S(R{:}QC)=2S(R), \tag{4.20}$$

$$S(R{:}Q_u'C)=S(R{:}QC)=2S(R). \tag{4.21}$$

In other words, the knowledge of the unerased part Q_u (or Q_u') is sufficient to reconstruct (using C) the initial logical word. Since the system RQ_uQ_ePC is in a pure state, we have

$$S(RQ_uC)=S(Q_eP)=S(Q_eC), \tag{4.22}$$

where we used the fact that P and C are interchangeable. Now, dividing Q_u into Q_e' and Q^*, we can write an upper bound on the mutual entropy between R and Q_uC,

$$\begin{aligned} S(R{:}Q_uC)&=S(R)+S(Q_uC)-S(RQ_uC)\\ &=S(R)+S(Q_e'Q^*C)-S(Q_eC)\\ &\leqslant S(R)+S(Q_e'C)+S(Q^*)-S(Q_eC), \end{aligned} \tag{4.23}$$

where we have used the subadditivity of quantum entropies. Eqs. (4.20), (4.21), (4.23), and its counterpart (when Q_u is replaced by Q_u') thus give

$$S(R)-S(Q^*)\leqslant S(Q_e'C)-S(Q_eC), \tag{4.24}$$

$$S(R)-S(Q^*)\leqslant S(Q_eC)-S(Q_e'C). \tag{4.25}$$

Combining these two last inequalities results in the same inequality as in the case where no classical auxiliary channel is used:

$$S(R)\leqslant S(Q^*)\leqslant n-2e. \tag{4.26}$$

Therefore, we obtain the *same* quantum Singleton bound for a quantum code supplemented with a noiseless classical channel as in the absence of such a channel:

$$k\leqslant n-2e. \tag{4.27}$$

In other words, the classical side channel (transmitting data with an entropy up to twice the quantum source entropy k) does not increase the Singleton bound on the maximum attainable distance for quantum codes.

This result can be immediately applied to a quantum channel characterized by a p-bounded fraction of erasures (cf. Sec. III A) and supplemented with an auxiliary classical channel, since, in that case, the use of a $((n,k))$ code protecting for $e=np$ erasures is enough to guarantee reliable transmission. Therefore, the upper bound on the rate is given by

$$R\leqslant 1-2p \tag{4.28}$$

similar to Eq. (3.1), confirming the fact that the classical side channel does not enhance the quantum information transmission through the quantum channel [6].

In the case of a quantum erasure channel (with erasure probability p) supplemented with a classical channel, the entire reasoning of Sec. III C can be repeated, the only difference being that one has to calculate the sum of $[S(Q_uC)-S(Q_eC)]$ for two overlapping channels (c and c'):

$$\begin{aligned} &S(Q_uC)-S(Q_eC)+S(Q_u'C)-S(Q_e'C)\\ &\qquad\leqslant 2S(Q^*)\leqslant 2(n-2e). \end{aligned} \tag{4.29}$$

The resulting bound on the mutual entropy is thus the same as Eq. (3.20), so that we have the same upper bound on the rate of reliable transmission of quantum information:

$$R\leqslant 1-2p. \tag{4.30}$$

Finally, the cases of a channel with a p-bounded fraction of errors and the quantum depolarizing channel can be treated exactly as in Secs. III B and III D, so that the classical side channel does not modify the Singleton upper bound on the rate in both cases.

V. CONCLUSION

The search for better bounds on the capacity of quantum channels such as the depolarizing channel is still a major endeavor in quantum information theory today. Clearly, entropic considerations alone do not suffice to prove that a reliable quantum coding scheme exists that achieves a transmission rate arbitrarily close to the capacity. As a matter of fact, a similar situation prevails for classical channels as well. Nevertheless, an entropic approach appears to be helpful in order to derive bounds on the capacity of classical or quantum channels from similar principles, and to analyze classical and quantum communication in a unified framework, as shown in this work. More generally, the leading idea underlying the approach to quantum information presented in this work and in Refs. [9–13] is to build a theory that extends Shannon's concepts to the quantum regime. Rather than attempting to define a *distinct* (purely quantum) information theory that would apply to the transmission of quantum states only, we prefer to consider an *extended* Shannon theory, which should account for the processing of classical as well as quantum information (arbitrary classical or quantum states). After all, any classical information-processing system should be describable, in principle, in terms of its underlying quantum mechanical degrees of freedom. In this sense, Shannon theory should simply be viewed as a special case of a more general theory of information in quantum mechanics that remains to be built.

The central characteristics of a quantum extension of Shannon theory happens to be that *negative* von Neumann conditional entropies must be used in order to encompass quantum entanglement in an information-theoretic description. It is pointed out in Ref. [10] that this apparently innocent observation should be viewed as the central novelty of a quantum mechanical extension of Shannon theory beyond its original range. Since most of the classical concepts of Shannon theory have a straightforward quantum analog, it is possible to repeat a great part of the classical reasoning and apply it to quantum information processes, as shown in this paper and our previous work. More specifically, we have shown here that an information-theoretic description of noisy quantum channels following closely Shannon theory provides insight into the derivation of entropic bounds on the quantum capacity (the maximum rate at which quantum information can be reliably processed in spite of the noise). Namely, the entropic Singleton bound on quantum error-correcting codes [15] can be used in order to investigate standard quantum channels such as the quantum erasure channel or the quantum depolarizing channel. The same formalism can be extended in order to account for an auxiliary classical channel used for forward communication besides the noisy quantum channel. Entropic Singleton bounds can be derived in the latter case, too, showing that the classical channel does not enhance the quantum capacity (or an upper bound on it), in agreement with what was proven in Ref. [6].

The central part of the reasoning consists in calculating a lower bound on the *average loss* of the channel (i.e., the loss of the joint channel made of n consecutive uses of a memoryless channel, divided by the number of processed symbols $n\to\infty$) which characterizes the ‘‘quality’’ of the transmission. If the use of block coding makes the joint (n-bit) channel lossless (i.e., the average loss is zero), then reliable transmission of quantum information is achievable. This is true even though the *one-symbol loss* (for a single use of the channel) is nonzero, reflecting the alteration due to noise in each use of the channel. Perfect transmission by block coding is thus possible provided that this lower bound on the average loss is zero or less, which results in an upper bound on the attainable rate.

Obviously, there remains much to be done in order to derive better bounds on the rate (or perhaps the exact capacity) using such an entropic approach. We have made progress in this direction, as illustrated by the entropic Singleton bound on the capacity of the quantum erasure channel ($C\leq 1-2p$), which happens to be the exact capacity calculated in [19]. For the quantum depolarizing channel, however, we obtain a well-known bound on the capacity ($C\leq 1-8p/3$, see [6]), which has been recently shown not to be attainable [27]. Nevertheless, the characterization of the exact quantum capacity of the depolarizing channel is still an open problem, and it is possible that the entropic approach presented here could be further improved. Also, the issue of the attainable capacity of a general noisy quantum channel might be explored along the same lines. The search for better entropic bounds on the capacity of quantum channels will be the subject of future work.

ACKNOWLEDGMENTS

We acknowledge C. Adami for numerous useful discussions. This research was supported in part by the National Science Foundation under Grant Nos. PHY 94-12818 and PHY 94-20470, and by a grant from DARPA/ARO through the QUIC Program (No. DAAH04-96-1-3086).

APPENDIX A: INFORMATION-THEORETIC CHARACTERIZATION OF A NOISY CLASSICAL CHANNEL

In this appendix, we outline the information-theoretic description of noisy classical channels, for the sake of clarifying the correspondence with the treatment of noisy quantum channels used throughout this paper. At first sight, a classical channel seems very different from a quantum channel as no classical ‘‘reference’’ is used to purify the input. Also, the classical input-output joint probability distribution has no quantum equivalent, since there is no joint state for the initial quantum system Q and the final system Q' (it is the *same* system). However, a classical channel can be thought of as a device which processes classical *correlation* (with respect to a reference R). If the input X is initially fully correlated with a reference R, then the residual mutual entropy between the output Y and R is a measure of the amount of correlation (or information) transmitted through the channel. For a quantum channel, we consider the processing of *entanglement* (with respect to R) rather than correlation, so that the residual mutual entropy between the decohered quantum system Q' (the quantum output) and R is the interesting quantity. This makes the classical-quantum correspondence easier to understand.

A noisy classical channel with input X and output Y is characterized by

$$I=H(X{:}Y), \tag{A1}$$

$$L=H(X|Y), \tag{A2}$$

$$N=H(Y|X), \tag{A3}$$

where I, L, and N denote the *information*, the *loss*, and the *noise*, respectively (see, e.g., [16]). Information processing through the channel is measured by the balance between I and L, these two quantities summing to the source entropy:

$$I+L=H(X). \tag{A4}$$

The loss measures the inherent uncertainty in the process of inferring the input of the channel from the altered output (decoding), that is, the entropy of the input conditional on the output. When the loss L is zero (*lossless* channel), the information I is maximum so that classical information is perfectly transmitted through the channel. Conversely, when $I=0$ (and L is maximum), no classical information is processed by the channel. The noise N reflects the uncertainty of the output symbol for a given input symbol, and is irrelevant as far as the transmission of information is concerned (it corresponds to the loss of the *reverse* channel where the input and output are interchanged). A channel with $N=0$ is called *deterministic*, and a channel which is both deterministic and lossless is named *noiseless*. We follow the same nomenclature for quantum channels in this paper.

If n *independent* channels are used in parallel with $X_1 \cdots X_n$ as an input string and $Y_1 \cdots Y_n$ as an output string, it can be shown that the information and the loss are *subadditive*:

$$I \leqslant I_1 + \cdots + I_n, \tag{A5}$$

$$L \leqslant L_1 + \cdots + L_n, \tag{A6}$$

where I (L) is the information (loss) of the joint (n-symbol) channel, while I_i (L_i) is the information (loss) of the ith individual channel $X_i \rightarrow Y_i$. Property (A5) results from the subadditivity of Shannon entropies

$$H(Y_1 \cdots Y_n) \leqslant H(Y_1) + \cdots + H(Y_n) \tag{A7}$$

and from the fact that the channels are independent (each Y_i depends on X_i only) [17]:

$$\begin{aligned} I &= H(X_1 \cdots X_n : Y_1 \cdots Y_n) \\ &= H(Y_1 \cdots Y_n) - H(Y_1 \cdots Y_n | X_1 \cdots X_n) \\ &\leqslant \sum_{i=1}^{n} [H(Y_i) - H(Y_i | Y_1 \cdots Y_{i-1}, X_1 \cdots X_n)] \\ &\leqslant \sum_{i=1}^{n} [H(Y_i) - H(Y_i | X_i)]. \end{aligned} \tag{A8}$$

Property (A6) is an immediate consequence of the subadditivity of Shannon conditional entropies

$$H(X_1 \cdots X_n | Y_1 \cdots Y_n) \leqslant H(X_1 | Y_1) + \cdots + H(X_n | Y_n). \tag{A9}$$

Since the information and the loss of each individual (one-symbol) channel sum to the source entropy of that channel

$$I_i + L_i = H(X_i), \tag{A10}$$

the allowed range for the overall loss of the joint channel is

$$L_1 + \cdots + L_n - M \leqslant L \leqslant L_1 + \cdots + L_n \tag{A11}$$

with $M = H(X_1) + \cdots + H(X_n) - H(X_1 \cdots X_n) \geqslant 0$. Consequently, the loss cannot increase by using block coding (i.e., using parallel channels with correlated input symbols), but it can decrease by an amount which is bounded by M. Note that M vanishes when the input symbols are independent, while a positive value of M reflects the correlation between the input symbols.

It is simple to obtain a necessary condition for perfect transmission (i.e., with a vanishing overall loss) by block coding through a noisy channel. Clearly, the condition

$$H(X_1 \cdots X_n) \leqslant H(X_1) + \cdots + H(X_n) - L_1 - \cdots - L_n \tag{A12}$$

must be fulfilled for the lower bound on L to extend to zero. Therefore, the rate of transmission through the joint channel, $R = H(X_1 \cdots X_n)/n$, is bounded from above by the averaged *one-symbol* information of the individual channels:

$$R \leqslant \frac{I_1 + \cdots + I_n}{n}. \tag{A13}$$

This is related to the weak converse of Shannon's noisy coding theorem [17]: the transmission cannot be perfect (or lossless, i.e., $L=0$) if the rate of transmission exceeds the (averaged) mutual Shannon entropy characterizing each use of the channel. This is consistent with Shannon's result that the classical channel capacity is the maximum (over all input distributions) of the mutual information between channel input and output for a *single* use of the channel. Note that entropy considerations alone do not suffice to prove that a reliable coding scheme exists that achieves a transmission rate arbitrarily close to the capacity. A similar situation is found for quantum channels as well, as shown throughout this paper. Still, an entropic approach is very helpful in order to derive bounds on classical or quantum channels from basic principles, and to analyze classical and quantum communication in a unified manner.

[1] C. H. Bennett, Phys. Today **48** (10), 24 (1995); D. P. Di Vincenzo, Science **270**, 255 (1995).

[2] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

[3] B. Schumacher, Phys. Rev. A **51**, 2738 (1995); R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).

[4] C. H. Bennett and S. J. Wiesner, Phys. Rev. Lett. **69**, 2881 (1992); C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *ibid.* **70**, 1895 (1993).

[5] P. W. Shor, Phys. Rev. A **52**, R2493 (1995); A. R. Calderbank and P. W. Shor, *ibid.* **54**, 1098 (1996); A. Steane, Phys. Rev. Lett. **77**, 793 (1996); D. Gottesman, Phys. Rev. A **54**, 1862 (1996).

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

[7] B. Schumacher, Phys. Rev. A **54**, 2614 (1996); B. Schumacher and M. A. Nielsen, *ibid.* **54**, 2629 (1996): H. Barnum, M. A. Nielsen, and B. Schumacher, Report No. quant-ph/9702049.

[8] S. Lloyd, Phys. Rev. A **55**, 1613 (1997).

[9] C. Adami and N. J. Cerf, Phys. Rev. A **56**, 3470 (1997).

[10] N. J. Cerf and C. Adami, Phys. Rev. Lett. **79**, 5194 (1997).

[11] N. J. Cerf and C. Adami, in *New Developments on Fundamental Problems in Quantum Physics*, Fundamental Theories of Physics Vol. 81, edited by M. Ferrero and A. van der Merwe (Kluwer Academic, Dordrecht, 1997), p. 77; also in Report No. quant-ph/9610005.

[12] N. J. Cerf and C. Adami, Physica D (to be published); also in Report No. quant-ph/9605039.

[13] N. J. Cerf and C. Adami, Report No. quant-ph/9611032.

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

[15] N. J. Cerf and R. Cleve, Phys. Rev. A **56**, 1721 (1997).

[16] R. B. Ash, *Information Theory* (Dover, New York, 1965).

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

[18] R. Cleve, Phys. Rev. A **55**, 4054 (1997).

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

[20] A. Ekert and C. Macchiavello, Phys. Rev. Lett. **77**, 2585 (1996).

[21] P. W. Shor and J. A. Smolin, Report No. quant-ph/9604006.

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

[23] M. Grassl, T. Beth, and T. Pellizzari, Phys. Rev. A **56**, 33 (1997).

[24] A. Ashikhmin and S. Litsyn, Report No. quant-ph/9709049.

[25] E. Rains, IEEE Trans. Inf. Theory (to be published); also in Report No. quant-ph/9611001.

[26] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).

[27] D. Bruss, D. P. DiVincenzo, A. Ekert, C. A. Fuchs, C. Macchiavello, and J. A. Smolin, Report No. quant-ph/9705038.

[28] V. Buzek and M. Hillery, Phys. Rev. A **54**, 1844 (1996).

[29] N. J. Cerf and C. Adami, Report No. quant-ph/9605002.

[30] N. J. Cerf and C. Adami (unpublished).