Contemporary Mathematics

Relative Entropy in Quantum Information Theory

Benjamin Schumacher and Michael D. Westmoreland

Abstract. We review the properties of the quantum relative entropy function and discuss its application to problems of classical and quantum information transfer and to quantum data compression. We then outline further uses of relative entropy to quantify quantum entanglement and analyze its manipulation.

1. Quantum relative entropy

In this paper we discuss several uses of the quantum relative entropy function in quantum information theory. Relative entropy methods have a number of advantages. First of all, the relative entropy functional satisfies some strong identities and inequalities, providing a basis for good theorems. Secondly, the relative entropy has a natural interpretation in terms of the statistical distinguishability of quantum states; closely related to this is the picture of relative entropy as a "distance" measure between density operators. These interpretations of the relative entropy give insight about the meaning of the mathematical constructions that use it. Finally, relative entropy has found a wide variety of applications in quantum information theory.

The usefulness of relative entropy in quantum information theory should come as no surprise, since the classical relative entropy has shown its power as a unifying concept in classical information theory [**1**]. Indeed, some of the results we will describe have close analogues in the classical domain. Nevertheless, the quantum relative entropy can provide insights in contexts (such as the quantification of quantum entanglement) that have no parallel in classical ideas of information.

Let Q be a quantum system described by a Hilbert space $\mathcal{H}$. (Throughout this paper, we will restrict our attention to systems with Hilbert spaces having a finite number of dimensions.) A pure state of Q can be described by a normalized vector $|\psi\rangle$ in $\mathcal{H}$, but a general (mixed) state requires a density operator ρ, which is a positive semi-definite operator on $\mathcal{H}$ with unit trace. For the pure state $|\psi\rangle$, the density operator ρ is simply the projection operator $|\psi\rangle\langle\psi|$; otherwise, ρ is a convex combination of projections. The entropy $S(\rho)$ is defined to be

$$S(\rho) = -\mathrm{Tr}\,\rho\log\rho. \tag{1}$$

2000 *Mathematics Subject Classification.* Primary 81P68.

©2002 American Mathematical Society

The entropy is non-negative and equals zero if and only if ρ is a pure state. (By "log" we will mean a logarithm with base 2.)

Closely related to the entropy of a state is the relative entropy of a pair of states. Let ρ and σ be density operators, and define the quantum relative entropy $\mathcal{S}(\rho||\sigma)$ to be

$$\mathcal{S}(\rho||\sigma) = \operatorname{Tr}\rho\log\rho - \operatorname{Tr}\rho\log\sigma. \tag{2}$$

(We read this as "the relative entropy of ρ with respect to σ".) This function has a number of useful properties: [**2**]

(1) $\mathcal{S}(\rho||\sigma) \geq 0$, with equality if and only if $\rho = \sigma$.
(2) $\mathcal{S}(\rho||\sigma) < \infty$ if and only if $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$. (Here "$\operatorname{supp}\rho$" is the subspace spanned by eigenvectors of ρ with non-zero eigenvalues.)
(3) The relative entropy is continuous where it is not infinite.
(4) The relative entropy is jointly convex in its arguments [**3**]. That is, if ρ_1, ρ_2, σ_1 and σ_2 are density operators, and p_1 and p_2 are non-negative numbers that sum to unity (i.e., probabilities), then

$$\mathcal{S}(\rho||\sigma) \leq p_1\mathcal{S}(\rho_1||\sigma_1) + p_2\mathcal{S}(\rho_2||\sigma_2) \tag{3}$$

where $\rho = p_1\rho_1 + p_2\rho_2$ and $\sigma = p_1\sigma_1 + p_2\sigma_2$. Joint convexity automatically implies convexity in each argument, so that (for example)

$$\mathcal{S}(\rho||\sigma) \leq p_1\mathcal{S}(\rho_1||\sigma) + p_2\mathcal{S}(\rho_2||\sigma). \tag{4}$$

The properties, especially property (1), motivate us to think of the relative entropy as a kind of "distance" between density operators. The relative entropy, which is not symmetric and which lacks a triangle inequality, is not technically a metric; but it is a positive definite directed measure of the separation of two density operators.

Suppose the density operator ρ_k occurs with probability p_k, yielding an average state $\rho = \sum_k p_k\rho_k$, and suppose σ is some other density operator. Then

$$\begin{aligned}
\sum_k p_k\mathcal{S}(\rho_k||\sigma) &= \sum_k p_k\left(\operatorname{Tr}\rho_k\log\rho_k - \operatorname{Tr}\rho_k\log\sigma\right) \\
&= \sum_k p_k\left(\operatorname{Tr}\rho_k\log\rho_k - \operatorname{Tr}\rho_k\log\rho + \operatorname{Tr}\rho_k\log\rho - \operatorname{Tr}\rho_k\log\sigma\right) \\
&= \sum_k p_k\left(\operatorname{Tr}\rho_k\log\rho_k - \operatorname{Tr}\rho_k\log\rho\right) + \operatorname{Tr}\rho\log\rho - \operatorname{Tr}\rho\log\sigma \\
\sum_k p_k\mathcal{S}(\rho_k||\sigma) &= \sum_k p_k\mathcal{S}(\rho_k||\rho) + \mathcal{S}(\rho||\sigma). \qquad (5)
\end{aligned}$$

Equation 5 is known as Donald's identity. [**4**]

The classical relative entropy of two probability distributions is related to the probability of distinguishing the two distributions after a large but finite number of independent samples. This is called Sanov's theorem [**1**], and this result has quantum analogue [**5**]. Suppose ρ and σ are two possible states of the quantum system Q, and suppose we are provided with N identically prepared copies of Q. A measurement is made to determine whether the prepared state is ρ, and the probability P_N that the state σ passes this test—in other words, is confused with ρ—is

$$P_N \approx 2^{-N\mathcal{S}(\rho||\sigma)} \tag{6}$$

as $N \to \infty$. (We have assumed that the measurement made is an optimal one for the purpose, and it is possible to show that an asymptotically optimal measurement strategy can be found that depends on ρ but not σ.)

The quantum version of Sanov's theorem tells us that the quantum relative entropy governs the asymptotic distinguishability of one quantum state from another by means of measurements. This further supports the view of $\mathcal{S}(\cdot||\cdot)$ as a measure of "distance"; two states are "close" if they are difficult to distinguish, but "far apart" if the probability of confusing them is small.

The remainder of this paper is organized as follows. Sections 2–5 apply relative entropy methods to the problem of sending classical information by means of a (possibly noisy) quantum channel. Sections 6–7 consider the transmission and compression of quantum information. Sections 8–9 then apply relative entropy methods to the discussion of quantum entanglement and its manipulation by local operations and classical communication. We conclude with a few remarks in Section 10.

2. Classical communication via quantum channels

One of the oldest problems in quantum information theory is that of sending classical information via quantum channels. A sender ("Alice") wishes to transmit classical information to a receiver ("Bob") using a quantum system as a communication channel. Alice will represent the message a, which occurs with probability p_a, by preparing the channel in the "signal state" represented by the density operator ρ_a. The average state of the channel will thus be $\rho = \sum_a p_a \rho_a$. Bob will attempt to recover the message by making a measurement of some "decoding observable" on the channel system.

The states ρ_a should be understood here as the "output" states of the channel, the states that Bob will attempt to distinguish in his measurement. In other words, the states ρ_a already include the effects of the dynamical evolution of the channel (including noise) on its way from sender to receiver. The dynamics of the channel will be described by a trace-preserving, completely positive map $\mathcal{E}$ on density operators [**6**]. The effect of $\mathcal{E}$ is simply to restrict the set of output channel states that Alice can arrange for Bob to receive. If $\mathcal{D}$ is the set of all density operators, then Alice's efforts can only produce output states in the set $\mathcal{A} = \mathcal{E}(\mathcal{D})$, a convex, compact set of density operators.

Bob's decoding observable is represented by a set of positive operators E_b such that $\sum_b E_b = 1$. If Bob makes his measurement on the state ρ_a, then the conditional probability of measurement outcome b is

$$P(b|a) = \operatorname{Tr} \rho_a E_b. \tag{7}$$

This yields a joint distribution over Alice's input messages a and Bob's decoded messages b:

$$P(a,b) = p_a P(b|a). \tag{8}$$

Once a joint probability distribution exists between the input and output messages (random variables A and B, respectively), the information transfer can be analyzed by classical information theory. The information obtained by Bob is given by the

mutual information $I(A:B)$:

$$I(A:B) = H(A) + H(B) - H(A,B) \tag{9}$$

where H is the Shannon entropy function

$$H(X) = -\sum_x p(x) \log p(x). \tag{10}$$

Shannon showed that, if the channel is used many times with suitable error-correcting codes, then any amount of information up to $I(A:B)$ bits (per use of the channel) can be sent from Alice to Bob with arbitrarily low probability of error [**1**]. The classical capacity of the channel is $C = \max I(A:B)$, where the maximum is taken over all input probability distributions. C is thus the maximum amount of information that may be reliably conveyed per use of the channel.

In the quantum mechanical situation, for a given ensemble of signal states ρ_a, Bob has many different choices for his decoding observable. Unless the signal states happen to be orthogonal, no choice of observable will allow Bob to distinguish perfectly between them. A theorem stated by Gordon[**7**] and Levitin[**8**] and first proved by Holevo[**9**] states that the amount of information accessible to Bob is limited by $I(A:B) \leq \chi$, where

$$\chi = S(\rho) - \sum_a p_a S(\rho_a). \tag{11}$$

The quantity χ is non-negative, since the entropy S is concave.

More recently, Holevo [**10**] and Schumacher and Westmoreland [**11**] have shown that this upper bound on $I(A:B)$ is asymptotically achievable. If Alice uses the same channel many times and prepares long codewords of signal states, and Bob uses an entangled decoding observable to distinguish these codewords, then Alice can convey to Bob up to χ bits of information per use of the channel, with arbitrarily low probability of error. (This fact was established for pure state signals $\rho_a = |\psi_a\rangle\langle\psi_a|$ in [**12**]. In this case, $\chi = S(\rho)$.)

The Holevo bound χ can be expressed in terms of the relative entropy:

$$\begin{aligned} \chi &= -\operatorname{Tr}\rho\log\rho + \sum_a p_a \operatorname{Tr}\rho_a \log\rho_a \\ &= \sum_a p_a \left(\operatorname{Tr}\rho_a\log\rho_a - \operatorname{Tr}\rho_a\log\rho\right) \\ \chi &= \sum_a p_a \mathcal{S}\left(\rho_a||\rho\right). \end{aligned} \tag{12}$$

In geometric terms, χ is the average relative entropy "directed distance" from the average state ρ to the members of the signal ensemble.

Donald's identity (Equation 5) has a particularly simple form in terms of χ. Given an ensemble and an additional state σ,

$$\sum_a p_a \mathcal{S}\left(\rho_a||\sigma\right) = \chi + \mathcal{S}\left(\rho||\sigma\right). \tag{13}$$

This implies, among other things, that

$$\chi \leq \sum_a p_a \mathcal{S}\left(\rho_a||\sigma\right) \tag{14}$$

with equality if and only if $\sigma = \rho$, the ensemble average state.

3. Thermodynamic cost of communication

In this section and the next, we focus on the transfer of classical information by means of a quantum channel.

Imagine a student who attends college far from home [**13**]. Naturally, the student's family wants to know that the student is passing his classes, and so they want the student to report to them frequently over the telephone. But the student is poor and cannot affort very many long-distance telephone calls. So they make the following arrangement: each evening at the same time, the poor student will call home only if he is failing one or more of this classes. Otherwise, he will save the phone charges by *not* calling home.

Every evening that the poor student does not call, therefore, the family is receiving a message *via the telephone* that his grades are good. (That the telephone is being used for this message can be seen from the fact that, if the phone lines are knocked out for some reason, the family can no longer make any inference from the absence of a phone call.)

For simplicity, imagine that the student's grades on successive days are independent and that the probability that the student will be failing on a given evening is p. Then the information conveyed each evening by the presence or absence of a phone call is

$$H(p) = -p \log p - (1-p) \log(1-p). \tag{15}$$

The cost of making a phone call is c, while not making a phone call is free. Thus, the student's average phone charge is cp per evening. The number of bits of information per unit cost is thus

$$\frac{H(p)}{cp} = \frac{1}{c} \left(-\log p - \left(\frac{1}{p} - 1 \right) \log(1-p) \right). \tag{16}$$

If the poor student is very successful in his studies, so that $p \to 0$, then this ratio becomes unboundedly large, even though both $H(p) \to 0$ and $cp \to 0$. That is, the student is able to send an arbitrarily large number of bits per unit cost. There is no irreducible cost for sending one bit of information over the telephone.

The key idea in the story of the poor student is that one possible signal—no phone call at all—has no cost to the student. The student can exploit this fact to use the channel in a cost-effective way, by using the zero-cost signal almost all of the time.

Instead of a poor student using a telephone, we can consider an analogous quantum mechanical problem. Suppose that a sender can manipulate a quantum channel to produce (for the receiver) one of two possible states, ρ_0 or ρ_1. The state ρ_0 can be produced at "zero cost", while the state ρ_1 costs a finite amount $c_1 > 0$ to produce. In the signal ensemble, the signal state ρ_1 is used with probability η and ρ_0 with probability $1-\eta$, leading to an average state

$$\rho = (1-\eta)\rho_0 + \eta\rho_1. \tag{17}$$

The average cost of creating a signal is thus $c = \eta c_1$. For this ensemble,

$$\chi = (1-\eta)\mathcal{S}\left(\rho_0 || \rho\right) + \eta \mathcal{S}\left(\rho_1 || \rho\right). \tag{18}$$

As discussed in the previous section, χ is an asymptotically achievable upper bound for the information transfered by the channel.

An upper bound for χ can be obtained from Donald's identity. Letting ρ_0 be the "additional" state,

$$\chi \leq (1-\eta)\mathcal{S}\left(\rho_0||\rho_0\right) + \eta\mathcal{S}\left(\rho_1||\rho_0\right) = \eta\mathcal{S}\left(\rho_1||\rho_0\right). \tag{19}$$

Combining this with a simple lower bound, we obtain

$$\eta\mathcal{S}\left(\rho_1||\rho\right) \leq \chi \leq \eta\mathcal{S}\left(\rho_1||\rho_0\right). \tag{20}$$

If we divide χ by the average cost, we find an asymptotically achievable upper bound for the number of bits sent through the channel per unit cost. That is,

$$\frac{\chi}{c} \leq \frac{1}{c_1}\mathcal{S}\left(\rho_1||\rho_0\right). \tag{21}$$

Furthermore, equality holds in the limit that $\eta \to 0$. Thus,

$$\sup\frac{\chi}{c} = \frac{1}{c_1}\mathcal{S}\left(\rho_1||\rho_0\right). \tag{22}$$

In short, the relative entropy "distance" between the signal state ρ_1 and the "zero cost" signal ρ_0 gives the largest possible number of bits per unit cost that may be sent through the channel—the "cost effectiveness" of the channel. If the state ρ_0 is a pure state, or if we can find a usable signal state ρ_1 whose support is not contained in the support of ρ_0, then $\mathcal{S}\left(\rho_1||\rho_0\right) = \infty$ and the cost effectiveness of the channel goes to infinity as $\eta \to 0$. (This is parallel to the situation of the poor student, who can make the ratio of "bits transmitted" to "average cost" arbitrarily large.)

What if there are many possible signal states ρ_1, ρ_2, etc., with positive costs c_1, c_2, and so on? If we assign the probability ηq_k to ρ_k for $k = 1, 2, \ldots$ (where $\sum_k q_k = 1$), and use ρ_0 with probability $1-\eta$, then we obtain

$$\eta\sum_k q_k\mathcal{S}\left(\rho_k||\rho\right) \leq \chi \leq \eta\sum_k q_k\mathcal{S}\left(\rho_k||\rho_0\right). \tag{23}$$

The average cost of the channel is $c = \eta\sum_k q_k c_k$. This means that

$$\frac{\chi}{c} \leq \frac{\sum_k q_k\mathcal{S}\left(\rho_k||\rho_0\right)}{\sum_k q_k c_k}. \tag{24}$$

We now note the following fact about real numbers. Suppose $a_n, b_n > 0$ for all n. Then

$$\frac{\sum_n a_n}{\sum_n b_n} \leq \max_n \frac{a_n}{b_n}. \tag{25}$$

This can be proven by letting $R = \max(a_n/b_n)$ and pointing out that $a_n \leq R b_n$ for all n. Then

$$\begin{aligned} \sum_n a_n &\leq R\sum_n b_n \\ \frac{\sum_n a_n}{\sum_n b_n} &\leq R. \end{aligned}$$

In our context, this implies that

$$\frac{\sum_k q_k\mathcal{S}\left(\rho_k||\rho_0\right)}{\sum_k q_k c_k} \leq \max_k \frac{q_k\mathcal{S}\left(\rho_k||\rho_0\right)}{q_k c_k} \tag{26}$$

and thus

$$\frac{\chi}{c} \leq \max_k \frac{\mathcal{S}\left(\rho_k||\rho_0\right)}{c_k}. \tag{27}$$

By using only the "most efficient state" (for which the maximum on the right-hand side is achieved) and adopting the "poor student" strategy of $\eta \to 0$, we can show that

$$\sup \frac{\chi}{c} = \max_k \frac{\mathcal{S}\left(\rho_k||\rho_0\right)}{c_k}. \tag{28}$$

These general considerations of an abstract "cost" of creating various signals have an especially elegant development if we consider the thermodynamic cost of using the channel. The thermodynamic entropy S_θ is related to the information-theoretic entropy $S(\rho)$ of the state ρ of the system by

$$S_\theta = k \ln 2 \; S(\rho). \tag{29}$$

The constant k is Boltzmann's constant. If our system has a Hamiltonian operator H, then the thermodynamic energy E of the state is the expectation of the Hamiltonian:

$$E = \langle H \rangle = \mathrm{Tr}\, \rho H. \tag{30}$$

Let us suppose that we have access to a thermal reservoir at temperature T. Then the "zero cost" state ρ_0 is the thermal equilibrium state

$$\rho_0 = \frac{1}{Z} e^{-\beta H}, \tag{31}$$

where $\beta = 1/kT$ and $Z = \mathrm{Tr}\, e^{-\beta H}$. ($Z$ is the partition function.)

The free energy of the system in the presence of a thermal reservoir at temperature T is $F = E - TS_\theta$. For the equilibrium state ρ_0,

$$\begin{aligned} F_0 &= \mathrm{Tr}\, \rho_0 H + kT \ln 2 \left(- \log Z - \frac{\beta}{\ln 2} \mathrm{Tr}\, \rho_0 H \right) \\ &= -kT \ln 2 \, \log Z \end{aligned} \tag{32}$$

The thermodynamic cost of the state ρ_1 is just the difference $F_1 - F_0$ between the free energies of ρ_1 and the equilibrium state ρ_0. But this difference has a simple relation to the relative entropy. First, we note

$$\mathrm{Tr}\, \rho_1 \log \rho_0 = - \log Z - \beta \mathrm{Tr}\, \rho_1 H, \tag{33}$$

from which it follows that [**14**]

$$\begin{aligned} F_1 - F_0 &= \mathrm{Tr}\, \rho_1 H + kT \ln 2 \, \mathrm{Tr}\, \rho_1 \log \rho_1 + kT \ln 2 \, \log Z \\ &= kT \ln 2 \left(\mathrm{Tr}\, \rho_1 \log \rho_1 - \mathrm{Tr}\, \rho_1 \log \rho_0 \right) \\ F_1 - F_0 &= kT \ln 2 \; \mathcal{S}\left(\rho_1||\rho_0\right). \end{aligned} \tag{34}$$

If we use the signal state ρ_1 with probability η, then the average thermodynamic cost is $f = \eta(F_1 - F_0)$. The number of bits sent per unit free energy is therefore

$$\frac{\chi}{f} \leq \eta \frac{\mathcal{S}\left(\rho_1||\rho_0\right)}{f} = \frac{1}{kT \ln 2}. \tag{35}$$

The same bound holds for all choices of the state ρ_1, and therefore for all ensembles of signal states.

We can approach this upper bound if we make η small, so that

$$\sup \frac{\chi}{f} = \frac{1}{kT \ln 2} \tag{36}$$

In short, for *any* coding and decoding scheme that makes use of the quantum channel, the maximum number of bits that can be sent per unit free energy is just $(kT \ln 2)^{-1}$. Phrased another way, the minimum free energy cost per bit is $kT \ln 2$.

This analysis can shed some light on Landauer's principle [**15**], which states that the minimum thermodynamic cost of information erasure is $kT \ln 2$ per bit. From this point of view, information erasure is simply information transmission into the environment, which requires the expenditure of an irreducible amount of free energy.

4. Optimal signal ensembles

Now we consider χ-maximizing ensembles of states from a given set $\mathcal{A}$ of available (output) states, without regard to the "cost" of each state. Our discussion in Section 2 tells us that the χ-maximizing ensemble is the one to use if we wish to maximize the classical information transfer from Alice to Bob via the quantum channel. Call an ensemble that maximizes χ an "optimal" signal ensemble, and denote the maximum value of χ by χ^*. (The results of this section are developed in more detail in [**16**].)

The first question is, of course, whether an optimal ensemble exists. It is conceivable that, though there is a least upper bound χ^* to the possible values of χ, no particular ensemble in $\mathcal{A}$ achieves it. (This would be similar to the results in the last section, in which the optimal cost effectiveness of the channel is only achieved in a limit.) However, an optimal ensemble does exist. Uhlmann [**17**] has proven a result that goes most of the way. Suppose our underlying Hilbert space $\mathcal{H}$ has dimension d and the set $\mathcal{A}$ of available states is convex and compact. Then given a fixed average state ρ, there exists an ensemble of at most d^2 signal states ρ_a that achieves the maximum value of χ for that particular ρ. The problem we are considering is to maximize χ over all choices of ρ in $\mathcal{A}$. Since Uhlmann has shown that each ρ-fixed optimal ensemble need involve no more than d^2 elements, we only need to maximize χ over ensembles that contain d^2 or fewer members. The set of such ensembles is compact and χ is a continuous function on this set, so χ achieves its maximum value χ^* for some ensemble with at most d^2 elements.

Suppose that the state ρ_a occurs with probability p_a in some ensemble, leading to the average state ρ and a Holevo quantity χ. We will now consider how χ changes if we modify the ensemble slightly. In the modified ensemble, a new state ω occurs with probability η and the state ρ_a occurs with probability $(1-\eta)p_a$. For the modified ensemble,

$$\rho' = \eta\omega + (1-\eta)\rho \tag{37}$$

$$\chi' = \eta \mathcal{S}\left(\omega||\rho'\right) + (1-\eta) \sum_a p_a \mathcal{S}\left(\rho_a||\rho'\right). \tag{38}$$

We can apply Donald's identity to these ensembles in two different ways. First, we can take the original optimal ensemble and treat ρ' as the other state (σ in Eq. 5), obtaining:

$$\sum_a p_a \mathcal{S}\left(\rho_a||\rho'\right) = \chi + \mathcal{S}\left(\rho||\rho'\right). \tag{39}$$

Substituting this expression into the expression for χ' yields:

$$\begin{aligned} \chi' &= \eta \mathcal{S}(\omega||\rho') + (1-\eta)(\chi + \mathcal{S}(\rho||\rho')) \\ \Delta\chi &= \chi' - \chi \\ &= \eta(\mathcal{S}(\omega||\rho') - \chi) + \eta\mathcal{S}(\rho||\rho') \end{aligned} \tag{40}$$

Our second application of Donald's identity is to the modified ensemble, taking the original average state ρ to play the role of the other state:

$$\eta\mathcal{S}(\omega||\rho) + (1-\eta)\chi = \chi' + \mathcal{S}(\rho'||\rho) \tag{41}$$

$$\Delta\chi = \eta(\mathcal{S}(\omega||\rho) - \chi) - \mathcal{S}(\rho'||\rho). \tag{42}$$

Since the relative entropy is never negative, we can conclude that

$$\eta(\mathcal{S}(\omega||\rho') - \chi) \leq \Delta\chi \leq \eta(\mathcal{S}(\omega||\rho) - \chi). \tag{43}$$

This gives upper and lower bounds for the change in χ if we mix in an additional state ω to our original ensemble. The bounds are "tight", since as $\eta \to 0$, $\mathcal{S}(\omega||\rho') \to \mathcal{S}(\omega||\rho)$.

Very similar bounds for $\Delta\chi$ apply if we make more elaborate modifications of our original ensemble, involving more than one additional signal state. This is described in [**16**].

We say that an ensemble has the *maximal distance property* if and only if, for any ω in $\mathcal{A}$,

$$\mathcal{S}(\omega||\rho) \leq \chi, \tag{44}$$

where ρ is the average state and χ is the Holevo quantity for the ensemble. This property gives an interesting characterization of optimal ensembles:

> **Theorem:** *An ensemble is optimal if and only if it has the maximum distance property.*

We give the essential ideas of the proof here; further details can be found in [**16**].

Suppose our ensemble has the maximum distance property. Then, if we add the state ω with probability η, the change $\Delta\chi$ satisfies

$$\Delta\chi \leq \eta(\mathcal{S}(\omega||\rho) - \chi) \leq 0. \tag{45}$$

In other words, we cannot increase χ by mixing in an additional state. Consideration of more general changes to the ensemble leads to the same conclusion that $\Delta\chi \leq 0$. Thus, the ensemble must be optimal, and $\chi = \chi^*$.

Conversely, suppose that the ensemble is optimal (with $\chi = \chi^*$). Could there be a state ω in $\mathcal{A}$ such that $\mathcal{S}(\omega||\rho) > \chi^*$? If there were such an ω, then by choosing η small enough we could make $\mathcal{S}(\omega||\rho') > \chi^*$, and so

$$\Delta\chi \geq \eta(\mathcal{S}(\omega||\rho') - \chi^*) > 0. \tag{46}$$

But this contradicts the fact that, if the original ensemble is optimal, $\Delta\chi \leq 0$ for any change in the ensemble. Thus, no such ω exists and the optimal ensemble satisfies the maximal distance property.

Two corollaries follow immediately from this theorem. First, we note that the support of the average state ρ of an optimal ensemble must contain the support of every state ω in $\mathcal{A}$. Otherwise, the relative entropy $\mathcal{S}(\omega||\rho) = \infty$, contradicting the maximal distance property. The fact that ρ has the largest support possible could be called the *maximal support property* of an optimal ensemble.

Second, we recall that χ^* is just the average relative entropy distance of the members of the optimal ensemble from the average state ρ:

$$\chi^* = \sum_a p_a \mathcal{S}\left(\rho_a || \rho\right).$$

Since $\mathcal{S}\left(\rho_a||\rho\right) \leq \chi^*$ for each a, it follows that whenever $p_a > 0$ we must have

$$\mathcal{S}\left(\rho_a||\rho\right) = \chi^*. \tag{47}$$

We might call this the *equal distance property* of an optimal ensemble.

We can now give an explicit formula for χ^* that does not optimize over ensembles, but only over states in $\mathcal{A}$. ¿From Equation 14, for any state σ,

$$\chi \leq \sum_a p_a \mathcal{S}\left(\rho_a||\sigma\right) \tag{48}$$

and thus

$$\chi \leq \max_\omega \mathcal{S}\left(\omega||\sigma\right) \tag{49}$$

where the maximum is taken over all ω in $\mathcal{A}$. We apply this inequality to the optimal ensemble, finding the lowest such upper bound for χ^*:

$$\chi^* \leq \min_\sigma \left(\max_\omega \mathcal{S}\left(\omega||\sigma\right) \right). \tag{50}$$

But since the optimal ensemble has the maximal distance property, we know that

$$\chi^* = \max_\omega \mathcal{S}\left(\omega||\rho\right) \tag{51}$$

for the optimal average state ρ. Therefore,

$$\chi^* = \min_\sigma \left(\max_\omega \mathcal{S}\left(\omega||\sigma\right) \right). \tag{52}$$

5. Additivity for quantum channels

The quantity χ^* is an asymptotically achievable upper bound to the amount of classical information that can be sent using available states of the channel system Q. It is therefore tempting to identify χ^* as the classical capacity of the quantum channel. But there is a subtlety here, which involves an important unsolved problem of quantum information theory.

Specifically, suppose that two quantum systems A and B are available for use as communication channels. The two systems evolve independently according the product map $\mathcal{E}^A \otimes \mathcal{E}^B$. Each system can be considered as a separate channel, or the joint system AB can be analyzed as a single channel. It is not known whether the following holds in general:

$$\chi^{AB*} \stackrel{?}{=} \chi^{A*} + \chi^{B*}. \tag{53}$$

Since separate signal ensembles for A and B can be combined into a product ensemble for AB, it is clear that $\chi^{AB*} \geq \chi^{A*} + \chi^{B*}$. However, the joint system AB also has other possible signal ensembles that use entangled input states and that might perhaps have a Holevo bound for the output states greater than $\chi^{A*} + \chi^{B*}$.

Equation 53 is the "additivity conjecture" for the classical capacity of a quantum channel. If the conjecture is false, then the use of entangled input states would sometimes increase the amount of classical information that can be sent over two or more independent channels. The classical capacity of a channel (which is defined asymptotically, using many instances of the same channel) would thus be greater

than χ^* for a single instance of a channel. On the other hand, if the conjecture holds, then χ^* is the classical capacity of the quantum channel.

Numerical calculations to date [**18**] support the additivity conjecture for a variety of channels. Recent work [**19, 20**] gives strong evidence that Equation 53 holds for various special cases, including channels described by unital maps. We present here another partial result: χ^* is additive for any "half-noisy" channel, that is, a dual channel that is represented by an map of the form $\mathcal{I}^A \otimes \mathcal{E}^B$, where $\mathcal{I}^A$ is the identity map on A.

Suppose the joint system AB evolves according to the map $\mathcal{I}^A \otimes \mathcal{E}^B$, and let ρ^A and ρ^B be the average output states of optimal signal ensembles for A and B individually. We will show that the product ensemble (with average state $\rho^A \otimes \rho^B$) is optimal by showing that this ensemble has the maximal distance property. That is, suppose we have another, possibly entangled input state of AB that leads to the output state ω^{AB}. Our aim is to prove that $\mathcal{S}\left(\omega^{AB}||\rho^A \otimes \rho^B\right) \leq \chi^{A*} + \chi^{B*}$. From the definition of $\mathcal{S}\left(\cdot||\cdot\right)$ we can show that

$$\begin{aligned} \mathcal{S}\left(\omega^{AB}||\rho^A \otimes \rho^B\right) &= -S\left(\omega^{AB}\right) - \operatorname{Tr}\omega^A \log \rho^A - \operatorname{Tr}\omega^B \log \rho^B \\ &= S\left(\omega^A\right) + S\left(\omega^B\right) - S\left(\omega^{AB}\right) \\ &\quad + \mathcal{S}\left(\omega^A||\rho^A\right) + \mathcal{S}\left(\omega^B||\rho^B\right). \end{aligned} \tag{54}$$

(The right-hand expression has an interesting structure; $S(\omega^A) + S(\omega^B) - S(\omega^{AB})$ is clearly analogous to the mutual information defined in Equation 9.)

Since A evolves according to the identity map $\mathcal{I}^A$, it is easy to see that $\chi^{A*} = d = \dim \mathcal{H}^A$ and

$$\rho^A = \left(\frac{1}{d}\right) 1^A. \tag{55}$$

¿From this it follows that

$$S\left(\omega^A\right) + \mathcal{S}\left(\omega^A||\rho^A\right) = \log d = \chi^{A*} \tag{56}$$

for any ω^A. This accounts for two of the terms on the right-hand side of Equation 54. The remaining three terms require a more involved analysis.

The final joint state ω^{AB} is a mixed state, but we can always introduce a third system C that "purifies" the state. That is, we can find $\left|\Omega^{ABC}\right\rangle$ such that

$$\omega^{AB} = \operatorname{Tr}_C \left|\Omega^{ABC}\right\rangle\left\langle\Omega^{ABC}\right|. \tag{57}$$

Since the overall state of ABC is a pure state, $S(\omega^{AB}) = S(\omega^C)$, where ω^C is the state obtained by partial trace over A and B. Furthermore, imagine that a complete measurement is made on A, with the outcome k occuring with probability p_k. For a given measurement outcome k, the subsequent state of the remaining system BC will be $\left|\Omega_k^{BC}\right\rangle$. Letting

$$\begin{aligned} \omega_k^B &= \operatorname{Tr}_C \left|\Omega_k^{BC}\right\rangle\left\langle\Omega_k^{BC}\right| \\ \omega_k^C &= \operatorname{Tr}_B \left|\Omega_k^{BC}\right\rangle\left\langle\Omega_k^{BC}\right|, \end{aligned} \tag{58}$$

we have that $S(\omega_k^B) = S(\omega_k^C)$ for all k. Furthermore, by locality,

$$\begin{aligned} \omega^B &= \sum_k p_k \omega_k^B \\ \omega^C &= \sum_k p_k \omega_k^C. \end{aligned} \tag{59}$$

In other words, we have written both ω^B and ω^C as ensembles of states.

We can apply this to get an upper bound on the remaining terms in Equation 54

$$\begin{aligned} & S\left(\omega^B\right) - S\left(\omega^{AB}\right) + \mathcal{S}\left(\omega^B||\rho^B\right) \\ &= \quad S\left(\omega^B\right) - \sum_k p_k S\left(\omega_k^B\right) \\ &\qquad -S\left(\omega^C\right) + \sum_k p_k S\left(\omega_k^C\right) + \mathcal{S}\left(\omega^B||\rho^B\right) \\ &\leq \quad \chi_\omega^B + \mathcal{S}\left(\omega^B||\rho^B\right), \end{aligned} \tag{60}$$

where χ_ω^B is the Holevo quantity for the ensemble of ω_k^B states. Donald's identity permits us to write

$$S\left(\omega^B\right) - S\left(\omega^{AB}\right) + \mathcal{S}\left(\omega^B||\rho^B\right) = \sum_k p_k \mathcal{S}\left(\omega_k^B||\rho^B\right). \tag{61}$$

The B states ω_k^B are all available output states of the B channel. These states are obtained by making a complete measurement on system A when the joint system AB is in the state ω^{AB}. But this state was obtained from some initial AB state and a dynamical map $\mathcal{I}^A \otimes \mathcal{E}^B$. This map commutes with the measurement operation on A alone, so we could equally well make the measurement *before* the action of $\mathcal{I}^A \otimes \mathcal{E}^B$. The A-measurement outcome k would then determine the input state of B, which would evolve into ω_k^B. Thus, for each k, ω_k^B is a possible output of the $\mathcal{E}^B$ map.

Since ρ^B has the maximum distance property and the states ω_k^B are available outputs of the channel, $\mathcal{S}\left(\omega_k^B||\rho^B\right) \leq \chi^{B*}$ for every k. Combining Equations 54, 56 and 61, we find the desired inequality:

$$\mathcal{S}\left(\omega^{AB}||\rho^A \otimes \rho^B\right) \leq \chi^{A*} + \chi^{B*}. \tag{62}$$

This demonstrates that the product of optimal ensembles for A and B also has the maximum distance property for the possible outputs of the joint channel, and so this product ensemble must be optimal. It follows that $\chi^{AB*} = \chi^{A*} + \chi^{B*}$ in this case.

Our result has been phrased for the case in which A undergoes "trivial" dynamics $\mathcal{I}^A$, but the proof also works without modification if the time evolution of A is unitary—that is, A experiences "distortion" but not "noise". If only one of the two systems is noisy, then χ^* is additive.

The additivity conjecture for χ^* is closely related to another additivity conjecture, the "minimum output entropy" conjecture [**19, 20**]. Suppose A and B are systems with independent evolution described by $\mathcal{E}^A \otimes \mathcal{E}^B$, and let $\rho^A B$ be an output state of the channel with minimal entropy $S(\rho^A B)$. Is ρ^{AB} a product state $\rho^A \otimes \rho^B$? The answer is not known in general; but it is quite easy to show this in the half-noisy case that we consider here.

6. Maximizing coherent information

When we turn from the transmission of classical information to the transmission of quantum information, it will be helpful to adopt an explicit description of the channel dynamics, instead of merely specifying the set of available output states $\mathcal{A}$. Suppose the quantum system Q undergoes a dynamical evolution described by the map $\mathcal{E}$. Since $\mathcal{E}$ is a trace-preserving, completely positive map, we can always

find a representation of $\mathcal{E}$ as a unitary evolution of a larger system [**6**]. In this representation, we imagine that an additonal "environment" system E is present, initially in a pure state $\left|\breve{0}^E\right\rangle$, and that Q and E interact via the unitary evolution operator U^{QE}. That is,

$$\rho^Q = \mathcal{E}(\breve{\rho}^Q) = \mathrm{Tr}_E U^{QE} \left(\breve{\rho}^Q \otimes \left|\breve{0}^E\right\rangle\left\langle\breve{0}^E\right|\right) U^{QE\dagger}. \tag{63}$$

For convenience, we denote an initial state of a system by the breve accent (as in $\breve{\rho}^Q$), and omit this symbol for final states.

The problem of sending quantum information through our channel can be viewed in one of two ways:

(1) An unknown pure quantum state of Q is to be transmitted. In this case, our criterion of success is the *average fidelity* $\bar{F}$, defined as follows. Suppose the input state $\left|\breve{\phi}_k\right\rangle$ occurs with probability p_k and leads to the output state ρ_k. Then

$$\bar{F} = \sum_k p_k \left\langle \breve{\phi}_k \right| \rho_k \left| \breve{\phi}_k \right\rangle. \tag{64}$$

In general, $\bar{F}$ depends not only on the average input state $\breve{\rho}^Q$ but also on the particular pure state input ensemble. [**21**]

(2) A second "bystander" system R is present, and the joint system RQ is initially in a pure entangled state $\left|\breve{\Psi}^{RQ}\right\rangle$. The system R has "trivial" dynamics described by the identity map $\mathcal{I}$, so that the joint system evolves according to $\mathcal{I} \otimes \mathcal{E}$, yielding a final state ρ^{RQ}. Success is determined in this case by the *entanglement fidelity* F_e, defined by

$$F_e = \left\langle \breve{\Psi}^{RQ} \right| \rho^{RQ} \left| \breve{\Psi}^{RQ} \right\rangle. \tag{65}$$

It turns out, surprisingly, that F_e is only dependent on $\mathcal{E}$ and the input state $\breve{\rho}^Q$ of Q alone. That is, F_e is an "intrinsic" property of Q and its dynamics. [**22**]

These two pictures of quantum information transfer are essentially equivalent, since F_e approaches unity if and only if $\bar{F}$ approaches unity for every ensemble with the same average input state $\breve{\rho}^Q$. For now we adopt the second point of view, in which the transfer of quantum information is essentially the transfer of quantum entanglement (with the bystander system R) through the channel.

The quantum capacity of a channel should be defined as the amount of entanglement that can be transmitted through the channel with $F_e \to 1$, if we allow ourselves to use the channel many times and employ quantum error correction schemes [**23**]. At present it is not known how to calculate this *asymptotic* capacity of the channel in terms of the properties of a single instance of the channel.

Nevertheless, we can identify some quantities that are useful in describing the quantum information conveyed by the channel [**24**]. A key quantity is the *coherent information* I^Q, defined by

$$I^Q = S\left(\rho^Q\right) - S\left(\rho^{RQ}\right). \tag{66}$$

This quantity is a measure of the final entanglement between R and Q. (The initial entanglement is measured by the entropy $S(\breve{\rho}^Q)$ of the initial state of Q, which of course equals $S(\breve{\rho}^R)$. See Section 7 below.) If we adopt a unitary representation

for $\mathcal{E}$, then the overall system RQE including the environment remains in a pure state from beginning to end, and so $S(\rho^{RQ}) = S(\rho^E)$. Thus,

$$I^Q = S\left(\rho^Q\right) - S\left(\rho^E\right). \tag{67}$$

Despite the apparent dependence of I^Q on the systems R and E, it is in fact a function only of the map $\mathcal{E}$ and the initial state $\breve{\rho}^Q$ of Q. Like the entanglement fidelity F_e, it is an "intrinsic" characteristic of the channel system Q and its dynamics.

It can be shown that the coherent information I^Q does not increase if the map $\mathcal{E}$ is followed by a second independent map $\mathcal{E}'$, giving an overall dynamics described by $\mathcal{E}' \circ \mathcal{E}$. That is, the coherent information cannot be increased by any "quantum data processing" on the channel outputs. The coherent information is also closely related to quantum error correction. Perfect quantum error correction—resulting in $F_e = 1$ for the final state—is possible if and only if the channel loses no coherent information, so that $I^Q = S(\breve{\rho}^Q)$. These and other properties lead us to consider I^Q as a good measure of the quantum information that is transmitted through the channel [**24**].

The coherent information has an intriguing relation to the Holevo quantity χ, and thus to classical information transfer (and to relative entropy) [**25**]. Suppose we describe that the input state $\breve{\rho}^Q$ by an ensemble of pure states $\left|\breve{\phi}_k^Q\right\rangle$:

$$\breve{\rho}^Q = \sum_k p_k \left|\breve{\phi}_k^Q\right\rangle \left\langle\breve{\phi}_k^Q\right|. \tag{68}$$

We adopt a unitary representation for the evolution and note that the initial pure state $\left|\breve{\phi}_k^Q\right\rangle \otimes \left|\breve{0}^E\right\rangle$ evolves into a pure, possibly entangled state $\left|\phi_k^{QE}\right\rangle$. Thus, for each k the entropies of the final states of Q and E are equal:

$$S\left(\rho_k^Q\right) = S\left(\rho_k^E\right). \tag{69}$$

It follows that

$$\begin{aligned} I^Q &= S\left(\rho^Q\right) - S\left(\rho^E\right) \\ &= S\left(\rho^Q\right) - \sum_k p_k S\left(\rho_k^Q\right) - S\left(\rho^E\right) + \sum_k p_k S\left(\rho_k^E\right) \\ I^Q &= \chi^Q - \chi^E. \end{aligned} \tag{70}$$

Remarkably, the difference $\chi^Q - \chi^E$ depends only on $\mathcal{E}$ and the average input state $\breve{\rho}^Q$, not the details of the environment E or the exact choice of pure state input ensemble.

The quantities χ^Q and χ^E are related to the classical information transfer to the output system Q and to the environment E, respectively. Thus, Equation 70 relates the classical and quantum information properties of the channel. This relation has been used to analyze the privacy of quantum cryptographic channels [**25**]. We will use it here to give a relative entropy characterization of the the input state $\breve{\rho}^Q$ that maximizes the coherent information of the channel.

Let us suppose that $\breve{\rho}^Q$ is an input state that maximizes the coherent information I^Q. If we change the input state to

$$\breve{\rho}^{Q\prime} = (1-\eta)\breve{\rho}^Q + \eta\breve{\omega}^Q, \tag{71}$$

for some pure state $\breve{\omega}^Q$, we produces some change ΔI^Q in the coherent information. Viewing $\breve{\rho}^Q$ as an ensemble of pure states, this change amounts to a modification

of that ensemble; and such a modification leads to changes in the output ensembles for both system Q and system E. Thus,

$$\Delta I^Q = \Delta\chi^Q - \Delta\chi^E. \tag{72}$$

We can apply Equation 43 to bound both $\Delta\chi^Q$ and $\Delta\chi^E$ and obtain a lower bound for ΔI^Q:

$$\begin{aligned} \Delta I^Q &\geq \eta\left(\mathcal{S}\left(\omega^Q||\rho^{Q\prime}\right) - \chi^Q\right) - \eta\left(\mathcal{S}\left(\omega^E||\rho^E\right) - \chi^E\right) \\ \Delta I^Q &\geq \eta\left(\mathcal{S}\left(\omega^Q||\rho^{Q\prime}\right) - \mathcal{S}\left(\omega^E||\rho^E\right) - I^Q\right). \end{aligned} \tag{73}$$

Since we assume that I^Q is maximized for the input $\breve{\rho}^Q$, then $\Delta I^Q \leq 0$ when we modify the input state. This must be true for every value of η in the relation above. Whenever $\mathcal{S}\left(\omega^Q||\rho^Q\right)$ is finite, we can conclude that

$$\mathcal{S}\left(\omega^Q||\rho^Q\right) - \mathcal{S}\left(\omega^E||\rho^E\right) \leq I^Q. \tag{74}$$

This is analogous to the maximum distance property for optimal signal ensembles, except that it is the difference of two relative entropy distances that is bounded above by the maximum of I^Q.

Let us write Equation 70 in terms of relative entropy, imagining that the input state $\breve{\rho}^Q$ is written in terms of an ensemble of pure states $\left|\breve{\phi}_k^Q\right\rangle$:

$$I^Q = \sum_k p_k \left(\mathcal{S}\left(\rho_k^Q||\rho^Q\right) - \mathcal{S}\left(\rho_k^E||\rho^E\right)\right). \tag{75}$$

Every input pure state $\left|\breve{\phi}_k^Q\right\rangle$ in the input ensemble with $p_k > 0$ will be in the support of $\breve{\rho}^Q$, and so Equation 74 holds. Therefore, we can conclude that

$$I^Q = \mathcal{S}\left(\rho_k^Q||\rho^Q\right) - \mathcal{S}\left(\rho_k^E||\rho^E\right) \tag{76}$$

for every such state in the ensemble. Furthermore, *any* pure state in the support of $\breve{\rho}^Q$ is a member of some pure state ensemble for $\breve{\rho}^Q$.

This permits us to draw a remarkable conclusion. If $\breve{\rho}^Q$ is the input state that maximizes the coherent information I^Q of the channel, then for any pure state $\breve{\omega}^Q$ in the support of $\breve{\rho}^Q$,

$$I^Q = \mathcal{S}\left(\omega^Q||\rho^Q\right) - \mathcal{S}\left(\omega^E||\rho^E\right). \tag{77}$$

This result is roughly analogous to the equal distance property for optimal signal ensembles. Together with Equation 74, it provides a strong characterization of the state that maximizes coherent information.

The additivity problem for χ^* leads us to ask whether the maximum of the coherent information is additive when independent channels are combined. In fact, there are examples known where $\max I^{AB} > \max I^A + \max I^B$; in other words, entanglement between independent channels can increase the amount of coherent information that can be sent through them [**26**]. The asymptotic behavior of coherent information and its precise connection to quantum channel capacities are questions yet to be resolved.

7. Indeterminate length quantum coding

In the previous section we saw that the relative entropy can be used to analyze the coherent information "capacity" of a quantum channel. Another issue in quantum information theory is *quantum data compression* [**21**], which seeks to represent quantum information using the fewest number of qubits. In this section we will see that the relative entropy describes the cost of suboptimal quantum data compression.

One approach to classical data compression is to use variable length codes, in which the codewords are finite binary strings of various lengths [**1**]. The best-known examples are the Huffman codes. The Shannon entropy $H(X)$ of a random variable X is a lower bound to the average codeword length in such codes, and for Huffman codes this average codeword length can be made arbitrarily close to $H(X)$. Thus, a Huffman code optimizes the use of a communication resources (number of bits required) in classical communication without noise.

There are analogous codes for the compression of quantum information. Since coherent superpositions of codewords must be allowed as codewords, these are called *indeterminate length* quantum codes [**27**]. A quantum analogue to Huffman coding was recently described by Braunstein et al. [**28**] An account of the theory of indeterminate length quantum codes, including the quantum Kraft inequality and the condensability condition (see below), will be presented in a forthcoming paper [**29**]. Here we will outline a few results and demonstrate a connection to the relative entropy.

The key idea in constructing an indeterminate length code is that the codewords themselves must carry their own length information. For a classical variable length code, this requirement can be phrased in two ways. A *uniquely decipherable* code is one in which any string of N codewords can be correctly separated into its individual codewords, while a *prefix-free* code is one in which no codeword is an initial segment of another codeword. The lengths of the codewords in each case satisfy the Kraft-McMillan inequality:

$$\sum_k 2^{-l_k} \leq 1, \tag{78}$$

where is the sum is over the codewords and l_k is the length of the kth codeword. Every prefix-free code is uniquely decipherable, so the prefix-free property is a more restrictive property. Nevertheless, it turns out that any uniquely decipherable code can be replaced by a prefix-free code with the same codeword lengths.

There are analogous conditions for indeterminate length quantum codes, but these properties must be phrased carefully because we allow coherent superpositions of codewords. For example, a classical prefix-free code is sometimes called an "instantaneous" code, since as soon as a complete codeword arrives we can recognize it at once and decipher it immediately. However, if an "instantaneous" decoding procedure were to be attempted for a quantum prefix-free code, it would destroy coherences between codewords of different lengths. Quantum codes require that the entire string of codewords be deciphered together.

The property of an indeterminate length quantum code that is analogous to unique decipherability is called *condensability*. We digress briefly to describe the condensability condition. We focus on *zero-extended forms* (zef) of our codewords. That is, we cosider that our codewords occupy an initial segment of a qubit register

of fixed length n, with $|0\rangle$'s following. (Clearly n must be chosen large enough to contain the longest codeword.) The set of all zef codewords spans a subspace of the Hilbert space of register states. We imagine that the output of a quantum information source has been mapped unitarily to the zef codeword space of the register. Our challenge is to take N such registers and "pack" them together in a way that can exploit the fact that some of the codewords are shorter than others.

If codeword states must carry their own length information, there must be a *length observable* Λ on the zef codeword space with the following two properties:

- The eigenvalues of Λ are integers $1, \ldots, n$, where n is the length of the register.
- If $\left|\psi_{\mathsf{zef}}\right\rangle$ is an eigenstate of Λ with eigenvalue l, then it has the form

$$\left|\psi_{\mathsf{zef}}\right\rangle = \left|\psi^{1\cdots l}0^{l+1\cdots n}\right\rangle. \tag{79}$$

  That is, the last $n-l$ qubits in the register are in the state $|0\rangle$ for a zef codeword of length l.

For register states not in the zef subspace, we can take $\Lambda = \infty$.

A code is *condensable* if the following condition holds: For any N, there is a unitary operator U (depending on N) that maps

$$\underbrace{\left|\psi_{1,\mathsf{zef}}\right\rangle \otimes \cdots \otimes \left|\psi_{N,\mathsf{zef}}\right\rangle}_{Nn\,\text{qubits}} \to \underbrace{\left|\Psi_{1\cdots N}\right\rangle}_{Nn\,\text{qubits}}$$

with the property that, if the individual codewords are all length eigenstates, then U maps the codewords to a zef string of the Nn qubits—that is, one with $|0\rangle$'s after the first $L = l_1 + \cdots + l_N$ qubits:

$$\left|\psi_1^{1\cdots l_1}0^{l_1+1\cdots n}\right\rangle \otimes \cdots \otimes \left|\psi_N^{1\cdots l_N}0^{l_N+1\cdots n}\right\rangle \to \left|\Psi^{1\cdots L}0^{L+1\cdots Nn}\right\rangle.$$

The unitary operator U thus "packs" N codewords, given in their zef forms, into a "condensed" string that has all of the trailing $|0\rangle$'s at the end. The unitary character of the packing protocol automatically yields an "unpacking" procedure given by U^{-1}. Thus, if the quantum code is condensable, a packed string of N codewords can be coherently sorted out into separated zef codewords.

The quantum analogue of the Kraft-McMillan inequality states that, for any indeterminate length quantum code that is condensable, the length observable Λ on the subspace of zef codewords must satisfy

$$\operatorname{Tr} 2^{-\Lambda} \leq 1, \tag{80}$$

where we have restricted our trace to the zef subspace. We can construct a density operator ω (a positive operator of unit trace) on the zef subspace by letting $K = \operatorname{Tr} 2^{-\Lambda} \leq 1$ and

$$\omega = \frac{1}{K}\, 2^{-\Lambda}. \tag{81}$$

The density operator ω is generally not the same as the actual density operator ρ of the zef codewords produced by the quantum information source. The average

codeword length is

$$\begin{aligned} \bar{l} &= \operatorname{Tr} \rho \Lambda \\ &= -\operatorname{Tr} \rho \log \left(2^{-\Lambda} \right) \\ &= -\operatorname{Tr} \rho \log \omega - \log K \\ \bar{l} &= S(\rho) + \mathcal{S}\left(\rho||\omega\right) - \log K. \end{aligned} \tag{82}$$

Since $\log K \leq 0$ and the relative entropy is positive definite,

$$\bar{l} \geq S(\rho). \tag{83}$$

The average codeword length must always be at least as great as the von Neuman entropy of the information source.

Equality for Equation 83 can be approached asymptotically using block coding and a quantum analogue of Huffman (or Shannon-Fano) coding. For special cases in which the eigenvalues of ρ are of the form 2^{-m}, then a code exists for which $\bar{l} = S(\rho)$, without the asymptotic limit. In either case, we say that a code satisfying $\bar{l} = S(\rho)$ is a length optimizing quantum code. Equation 82 tells us that, if we have a length optimizing code, $K = 1$ and

$$\rho = \omega = 2^{-\Lambda}. \tag{84}$$

The condensed string of N codewords has Nn qubits, but we can discard all but about $N\bar{l}$ of them and still retain high fidelity. That is, $\bar{l}$ is the asymptotic number of qubits that must be used per codeword to represent the quantum information faithfully.

Suppose that we have an indeterminate length quantum code that is designed for the wrong density operator. That is, our code is length optimizing for some other density operator ω, but $\rho \neq \omega$. Then (recalling that $K = 1$ for a length optimizing code, even if it is optimizing for the wrong density operator),

$$\bar{l} = S(\rho) + \mathcal{S}\left(\rho||\omega\right). \tag{85}$$

$S(\rho)$ tells us the number of qubits necessary to represent the quantum information if we used a length optimizing code for ρ. (As we have mentioned, such codes always exist in an asymptotic sense.) However, to achieve high fidelity in the situation where we have used a code designed for ω, we have to use at least $\bar{l}$ qubits per codeword, an additional cost of $\mathcal{S}\left(\rho||\omega\right)$ qubits per codeword.

This result gives us an interpretation of the relative entropy function $\mathcal{S}\left(\rho||\omega\right)$ in terms of the physical resources necessary to accomplish some task—in this case, the additional cost (in qubits) of representing the quantum information described by ρ using a coding scheme optimized for ω. This is entirely analogous to the situation for classical codes and classical relative entropy [**1**]. A fuller development of this analysis will appear in [**29**].

8. Relative entropy of entanglement

One recent application of relative entropy has been to quantify the entanglement of a mixed quantum state of two systems [**30**]. Suppose Alice and Bob share a joint quantum system AB in the state ρ^{AB}. This state is said to be *separable* if it is a product state or else a probabilistic combination of product states:

$$\rho^{AB} = \sum_k p_k \rho_k^A \otimes \rho_k^B. \tag{86}$$

Without loss of generality, we can if we wish take the elements in this ensemble of product states to be pure product states. Systems in separable states display statistical correlations having perfectly ordinary "classical" properties—that is, they do not violate any sort of Bell inequality. A separable state of A and B could also be created from scratch by Alice and Bob using only local quantum operations (on A and B separately) and the exchange of classical information.

States which are not separable are said to be *entangled*. These states cannot be made by local operations and classical communication; in other words, their creation requires the exchange of *quantum* information between Alice and Bob. The characterization of entangled states and their possible transformations has been a central issue in much recent work on quantum information theory.

A key question is the quantification of entanglement, that is, finding numerical measures of the entanglement of a quantum state ρ^{AB} that have useful properties. If the joint system AB is in a pure state $\left|\Psi^{AB}\right\rangle$, so that the subsystem states are

$$\begin{aligned} \rho^A &= \mathrm{Tr}_B \left|\Psi^{AB}\right\rangle\left\langle\Psi^{AB}\right| \\ \rho^B &= \mathrm{Tr}_A \left|\Psi^{AB}\right\rangle\left\langle\Psi^{AB}\right| \end{aligned} \tag{87}$$

then the entropy $S(\rho^A) = S(\rho^B)$ can be used to measure the entanglement of A and B. This measure has many appealing properties. It is zero if and only if $\left|\Psi^{AB}\right\rangle$ is separable (and thus a product state). For an "EPR pair" of qubits—that is, a state of the general form

$$\left|\phi^{AB}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0^A 0^B\right\rangle + \left|1^A 1^B\right\rangle\right), \tag{88}$$

the susbsystem entropy $S(\rho^A) = 1$ bit.

The subsystem entropy is also an asymptotic measure, both of the resources necessary to create the particular entangled pure state, and of the value of the state as a resource [**31**]. That is, for sufficiently large N,

- approximately $NS(\rho^A)$ EPR pairs are required to create N copies of $\left|\Psi^{AB}\right\rangle$ by local operations and classical communication; and
- approximately $NS(\rho^A)$ EPR pairs can be created from N copies of $\left|\Psi^{AB}\right\rangle$ by local operations and classical communication.

For mixed entangled states ρ^{AB} of the joint system AB, things are not so well-established. Several different measures of entanglement are known, including [**32**]

- the *entanglement of formation* $E(\rho^{AB})$, which is the minimum asymptotic number of EPR pairs required to create ρ^{AB} by local operations and classical communication; and
- the distillable entanglement $D(\rho^{AB})$, the maximum asymptotic number of EPR pairs that can be created from ρ^{AB} by entanglement purification protocols involving local operations and classical communication.

Bennett et al. [**32**] further distinguish D_1 and D_2, the distillable entanglements with respect to purification protocols that allow one-way and two-way classical communication, respectively. All of these measures reduce to the subsystem entropy $S(\rho^A)$ if ρ^{AB} is a pure entangled state.

These entanglement measures are not all equal; furthermore, explicit formulas for their calculation are not known in most cases. This motivates us to consider

alternate measures of entanglement with more tractable properties and which have useful relations to the asymptotic measures E, D_1 and D_2.

A state ρ^{AB} is entangled inasmuch as it is not a separable state, so it makes sense to adopt as a measure of entanglement a measure of the distance of ρ^{AB} from the set Σ^{AB} of separable states of AB. Using relative entropy as our "distance", we define the *relative entropy of entanglement* E_r to be [**30**]

$$E_r\left(\rho^{AB}\right) = \min_{\sigma^{AB}\in\Sigma^{AB}} \mathcal{S}\left(\rho^{AB}||\sigma^{AB}\right). \tag{89}$$

The relative entropy of entanglement has several handy properties. First of all, it reduces to the subsystem entropy $S(\rho^A)$ whenever ρ^{AB} is a pure state. Second, suppose we write ρ^{AB} as an ensemble of pure states $\left|\psi_k^{AB}\right\rangle$. Then

$$E_r\left(\rho^{AB}\right) \leq \sum_k p_k S\left(\rho_k^A\right) \tag{90}$$

where $\rho_k^A = \mathrm{Tr}_B \left|\psi_k^{AB}\right\rangle\left\langle\psi_k^{AB}\right|$. It follows from this that $E_r \leq E$ for any state ρ^{AB}.

Even more importantly, the relative entropy of entanglement E_r can be shown to be non-increasing on average under local operations by Alice and Bob together with classical communication between them.

The quantum version of Sanov's theorem gives the relative entropy of entanglement an interpretation in terms of the statistical distinguishability of ρ^{AB} and the "least distinguishable" separable state σ^{AB}. The relative entropy of entanglement is thus a useful and well-motivated measure of the entanglement of a state ρ^{AB} of a joint system, both on its own terms and as a surrogate for less tractable asymptotic measures.

9. Manipulating multiparticle entanglement

The analysis in this section closely follows that of Linden et al. [**33**], who provides a more detailed discussion of the main result here and its applications.

Suppose Alice, Bob and Claire initially share three qubits in a "GHZ state"

$$\left|\Psi^{ABC}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0^A 0^B 0^C\right\rangle + \left|1^A 1^B 0^C\right\rangle\right). \tag{91}$$

The mixed state ρ^{BC} shared by Bob and Claire is, in fact, not entangled at all:

$$\rho^{BC} = \frac{1}{2}\left(\left|0^B 0^C\right\rangle\left\langle 0^B 0^C\right| + \left|1^B 1^C\right\rangle\left\langle 1^B 1^C\right|\right). \tag{92}$$

No local operations performed by Bob and Claire can produce an entangled state from this starting point. However, Alice can create entanglement for Bob and Claire. Alice measures her qubit in the basis $\{\left|+^A\right\rangle, \left|-^A\right\rangle\}$, where

$$\left|\pm^A\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0^A\right\rangle \pm \left|1^A\right\rangle\right). \tag{93}$$

It is easy to verify that the state of Bob and Claire's qubits after this measurement, depending on the measurement outcome, must be one of the two states

$$\left|\phi_\pm^{BC}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0^A 0^B\right\rangle \pm \left|1^A 1^B\right\rangle\right), \tag{94}$$

both of which are equivalent (up to a local unitary transformation by either Bob or Claire) to an EPR pair. In other words, if Alice makes a local measurement on her

qubit and then announces the result by classical communication, the GHZ triple can be converted into an EPR pair for Bob and Claire.

When considering the manipulation of quantum entanglement shared among several parties, we must therefore bear in mind that the entanglement between subsystems can both increase and decrease, depending on the situation. This raises several questions: Under what circumstances can Alice increase Bob and Claire's entanglement? How much can she do so? Are there any costs involved in the process?

To study these questions, we must give a more detailed account of "local operations and classical communication". It turns out that Alice, Bob and Claire can realize any local operation on their joint system ABC by a combination of the following:

- Local unitary transformations on the subsystems A, B and C;
- Adjoining to a subsystem additional local "ancilla" qubits in a standard state $|0\rangle$;
- Local ideal measurements on the (augmented) subsystems A, B and C; and
- Discarding local ancilla qubits.

Strictly speaking, though, we do not need to include the last item. That is, any protocol that involves discarding ancilla qubits can be replaced by one in which the ancillas are simply "set aside"—not used in future steps, but not actually gotten rid of. In a similar vein, we can imagine that the ancilla qubits required are already present in the subsystems A, B and C, so the second item in our list is redundant. We therefore need to consider only local unitary transformations and local ideal measurements.

What does classical communication add to this? It is sufficient to suppose that Alice, Bob and Claire have complete information—that is, they are aware of all operations and the outcomes of all measurements performed by each of them, and thus know the global state of ABC at every stage. Any protocol that involved an incomplete sharing of information could be replaced by one with complete sharing, simply by ignoring some of the messages that are exchanged.

Our local operations (local unitary transformations and local ideal measurements) always take an initial pure state to a final pure state. That is, if ABC starts in the joint state $\left|\Psi^{ABC}\right\rangle$, then the final state will be a pure state $\left|\Psi_k^{ABC}\right\rangle$ that depends on the joint outcome k of all the measurements performed. Thus, ABC is always in a pure state known to all parties.

It is instructive to consider the effect of local operations on the entropies of the various subsystems of ABC. Local unitary transformations leave $S(\rho^A)$, $S(\rho^B)$ and $S(\rho^C)$ unchanged. But suppose that Alice makes an ideal measurement on her subsystem, obtaining outcome k with probability p_k. The initial global state is $\left|\Psi^{ABC}\right\rangle$ and the final global state is $\left|\Psi_k^{ABC}\right\rangle$, depending on k. For the initial subsystem states, we have that

$$S\left(\rho^A\right) = S\left(\rho^{BC}\right) \tag{95}$$

since the overall state is a pure state. Similarly, the various final subsystem states satisfy

$$S\left(\rho_k^A\right) = S\left(\rho_k^{BC}\right). \tag{96}$$

But an operation on A cannot change the average state of BC:

$$\rho^{BC} = \sum_k p_k \rho_k^{BC}. \tag{97}$$

Concavity of the entropy gives

$$S\left(\rho^{BC}\right) \geq \sum_k p_k S\left(\rho_k^{BC}\right) \tag{98}$$

and therefore

$$S\left(\rho^{A}\right) \geq \sum_k p_k S\left(\rho_k^{A}\right). \tag{99}$$

Concavity also tells us that $S(\rho^B) \geq \sum_k p_k S(\rho_k^B)$, etc., and similar results hold for local measurements performed by Bob or Claire.

We now return to the question of how much Alice can increase the entanglement shared by Bob and Claire. Let us measure the bipartite entanglement of the system BC (which may be in a mixed state) by the relative entropy of entanglement $E_r(\rho^{BC})$, and let σ^{BC} be the separable state of BC for which

$$E_r(\rho^{BC}) = \mathcal{S}\left(\rho^{BC}||\sigma^{BC}\right). \tag{100}$$

No local unitary operation can change $E_r(\rho^{BC})$; furthermore, no local measurement by Bob or Claire can increase $E_r(\rho^{BC})$ on average. We need only consider an ideal measurement performed by Alice on system A. Once again we suppose that outcome k of this measurement occurs with probability p_k, and once again Equation 97 holds. Donald's identity tells us that

$$\sum_k p_k \mathcal{S}\left(\rho_k^{BC}||\sigma^{BC}\right) = \sum_k p_k \mathcal{S}\left(\rho_k^{BC}||\rho^{BC}\right) + \mathcal{S}\left(\rho^{BC}||\sigma^{BC}\right). \tag{101}$$

But $E_r(\rho_k^{BC}) \leq \mathcal{S}\left(\rho_k^{BC}||\sigma^{BC}\right)$ for every k, leading to the following inequality:

$$\sum_k p_k E_r(\rho_k^{BC}) - E_r(\rho^{BC}) \leq \sum_k p_k \mathcal{S}\left(\rho_k^{BC}||\rho^{BC}\right). \tag{102}$$

We recognize the left-hand side of this inequality χ for the ensemble of post-measurement states of BC, which we can rewrite using the definition of χ in Equation 11. This yields:

$$\begin{aligned} \sum_k p_k E_r(\rho_k^{BC}) - E_r(\rho^{BC}) &\leq S\left(\rho^{BC}\right) - \sum_k p_k S\left(\rho_k^{BC}\right) \\ &= S\left(\rho^{A}\right) - \sum_k p_k S\left(\rho_k^{A}\right), \end{aligned} \tag{103}$$

since the overall state of ABC is pure at every stage.

To summarize, in our model (in which all measurements are ideal, all classical information is shared, and no classical or quantum information is ever discarded), the following principles hold:

- The entropy of any subsystem A cannot be increased on average by any local operations.
- The relative entropy of entanglement of two subsystems B and C cannot be increased on average by local operations on those two subsystems.

- The relative entropy of entanglement of B and C can be increased by a measurement performed on a third subsystem A, but the average increase in $E_r^B C$ is no larger than the average decrease in the entropy of A.

We say that a joint state $\left|\Psi_1^{ABC}\right\rangle$ can be transformed *reversibly* into $\left|\Psi_2^{ABC}\right\rangle$ if, for sufficiently large N, N copies of $\left|\Psi_1^{ABC}\right\rangle$ can be transformed with high probability (via local operations and classical communication) to approximately N copies of $\left|\Psi_2^{ABC}\right\rangle$, and *vice versa*. The qualifiers in this description are worth a comment or two. "High probability" reflects the fact that, since the local operations may involve measurements, the actual final state may depend on the exact measurement outcomes. "Approximately N copies" means more than $(1-\epsilon)N$ copies, for some suitably small ϵ determined in advance. We denote this reversibility relation by

$$\left|\Psi_1^{ABC}\right\rangle \leftrightarrow \left|\Psi_2^{ABC}\right\rangle .$$

Two states that are related in this way are essentially equivalent as "entanglement resources". In the large N limit, they may be interconverted with arbitrarily little loss.

Our results for entropy and relative entropy of entanglement allow us to place necessary conditions on the reversible manipulation of multiparticle entanglement. For example, if $\left|\Psi_1^{ABC}\right\rangle \leftrightarrow \left|\Psi_2^{ABC}\right\rangle$, then the two states must have exactly the same subsystem entropies. Suppose instead that $S(\rho_1^A) < S(\rho_2^A)$. Then the transformation of N copies of $\left|\Psi_1^{ABC}\right\rangle$ into about N copies of $\left|\Psi_2^{ABC}\right\rangle$ would involve an increase in the entropy of subsystem A, which cannot happen on average.

In a similar way, we can see that $\left|\Psi_1^{ABC}\right\rangle$ and $\left|\Psi_2^{ABC}\right\rangle$ must have the same relative entropies of entanglement for every pair of subsystems. Suppose instead that $E_{r,1}^{BC} < E_{r,2}^{BC}$. Then the transformation of N copies of $\left|\Psi_1^{ABC}\right\rangle$ into about N copies of $\left|\Psi_2^{ABC}\right\rangle$ would require an increase in E_r^{BC}. This can take place if a measurement is performed on A, but as we have seen this would necessarily involve a decrease in $S(\rho^A)$. Therefore, reversible transformations of multiparticle entanglement must preserve both subsystem entropies and the entanglement (measured by E_r) of pairs of subsystems.

As a simple example of this, suppose Alice, Bob and Claire share two GHZ states. Each subsystem has an entropy of 2.0 bits. This would also be the case if Alice, Bob and Claire shared three EPR pairs, one between each pair of participants. Does it follow that two GHZs can be transformed reversibly (in the sense described above) into three EPRs?

No. If the three parties share two GHZ triples, then Bob and Claire are in a completely unentangled state, with $E_r^{BC} = 0$. But in the "three EPR" situation, the relative entropy of entanglement E_r^{BC} is 1.0 bits, since they share an EPR pair. Thus, two GHZs cannot be reversibly transformed into three EPRs; indeed, $2N$ GHZs are inequivalent to $3N$ EPRs.

Though we have phrased our results for three parties, they are obviously applicable to situations with four or more separated subsystems. In reversible manipulations of multiparticle entanglement, all subsystem entropies (including the entropies of clusters of subsystems) must remain constant, as well as the relative entropies of entanglement of all pairs of subsystems (or clusters of subsystems).

10. Remarks

The applications discussed here show the power and the versatility of relative entropy methods in attacking problems of quantum information theory. We have derived useful fundamental results in classical and quantum information transfer, quantum data compression, and the manipulation of quantum entanglement. In particular, Donald's identity proves to be an extremely useful tool for deriving important inequalities.

One of the insights provided by quantum information theory is that the von Neumann entropy $S(\rho)$ has an interpretation (actually several interpretations) as a measure of the resources necessary to perform an information task. We have seen that the relative entropy also supports such interpretations. We would especially like to draw attention to the results in Sections 3 on the cost of communication and Section 7 on quantum data compression, which are presented here for the first time.

We expect that relative entropy techniques will be central to further work in quantum information theory. In particular, we think that they show promise in resolving the many perplexing additivity problems that face the theory at present. Section 5, though not a very strong result in itself, may point the way along this road.

The authors wish to acknowledge the invaluable help of many colleagues. T. Cover, M. Donald, M. Neilsen, M. Ruskai, A. Uhlmann and V. Vedral have given us indispensible guidance about the properties and meaning of the relative entropy function. Our work on optimal signal ensembles and the additivity problem was greatly assisted by conversations with C. Fuchs, A. Holevo, J. Smolin, and W. Wootters. Results described here on reversibility for transformations of multiparticle entanglement were obtained in the course of joint work with N. Linden and S. Popescu. We would like to thank the organizers of the AMS special session on "Quantum Information and Computation" for a stimulating meeting and an opportunity to pull together several related ideas into the present paper. We hope it will serve as a spur for the further application of relative entropy methods to problems of quantum information theory.

References

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

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

[3] E. Leib and M. B. Ruskai, *Phys. Rev. Lett.* **30**, 434 (1973); E. Leib and M. B. Ruskai, *J. Math. Phys.* **14**, 1938 (1973).

[4] M. J. Donald, Math. Proc. Cam. Phil. Soc. **101**, 363 (1987).

[5] F. Hiai and D. Petz, Comm. Math. Phys. **143**, 99 (1991). V. Vedral, M. B. Plenio, K. Jacobs and P. L. Knight, Phys. Rev. A **56**, 4452 (1997).

[6] W. F. Stinespring, *Proc. of Am. Math. Soc.* **6**, 211 (1955); K. Kraus, *Annals of Phys.* **64**, 311 (1971); K. Hellwig and K. Kraus, *Comm. Math. Phys.* **16**, 142 (1970); M.-D. Choi, *Lin. Alg. and Its Applications* **10**, 285 (1975); K. Kraus, *States, Effects and Operations: Fundamental Notions of Quantum Theory* (Springer-Verlag, Berlin, 1983).

[7] J. P. Gordon, "Noise at optical frequencies; information theory," in *Quantum Electronics and Coherent Light; Proceedings of the International School of Physics "Enrico Fermi," Course XXXI*, P. A. Miles, ed., (Academic Press, New York, 1964), pp. 156–181.

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

[9] A. S. Holevo, *Probl. Inform. Transmission* **9**, 177 (1973) (translated from *Problemy Peredachi Informatsii*).
[10] A. S. Holevo, IEEE Trans. Inform. Theory **44**, 269 (1998).
[11] B. Schumacher and M. Westmoreland, *Phys. Rev. A* **51**, 2738 (1997).
[12] P. Hausladen, R. Josza, B. Schumacher, M. Westmoreland, and W. K. Wootters, *Phys. Rev. A* **54**, 1869 (1996).
[13] B. Schumacher, *Communication, Correlation and Complementarity*, Ph.D. thesis, the University of Texas at Austin (1990).
[14] G. Lindblad, *Non-Equilibrium Entropy and Irreversibility*, (Reidel, Dordrecht, 1983); M. Donald, *J. Stat. Phys.* **49**, 81 (1987); H. M. Partovi, *Phys. Lett. A* **137**, 440 (1989).
[15] R. Landauer, *IBM J. Res. Develop.* **5**, 183 (1961); V. Vedral, *Proc. Royal Soc.* (to appear, 2000). LANL e-print quant-ph/9903049.
[16] B. Schumacher and M. Westmoreland, "Optimal signal ensembles", submitted *Phys. Rev. A*. LANL e-print quant-ph/9912122.
[17] A. Uhlmann, Open Sys. and Inf. Dynamics **5**, 209 (1998).
[18] C. H. Bennett, C. Fuchs and J. A. Smolin, "Entanglement enhanced classical communication on a noisy quantum channel", in: Proc. 3d Int. Conf. on Quantum Communication and Measurement, ed. by C. M. Caves, O. Hirota, A. S. Holevo, Plenum, NY 1997. LANL e-print quant-ph/9611006.
[19] C. King and M. B. Ruskai, "Minimal entropy of states emerging from noisy quantum channels". LANL e-print quant-ph/9911079.
[20] G. G. Amosov, A. S. Holveo and R. F. Werner, "On some additivity problems in quantum information theory", LANL e-print quant-ph/0003002.
[21] B. Schumacher, *Phys. Rev. A* **51**, 2738 (1995); R. Jozsa and B. Schumacher, *J. Mod. Opt.* **41**, 2343 (1994); H. Barnum, C. A. Fuchs, R. Jozsa and B. Schumacher, *Phys. Rev. A* **54**, 4707 (1996).
[22] B. Schumacher, *Phys. Rev. A* **54**, 2614 (1996).
[23] H. Barnum, M. A. Nielsen and B. Schumacher, *Phys. Rev. A* **57**, 4153 (1998).
[24] B. Schumacher and M. A. Nielsen, *Phys. Rev. A* **54**, 2629 (1996).
[25] B. Schumacher and M. Westmoreland, *Physical Review Letters*, **80** (June, 1998), 5695 - 5697.
[26] D. P. DiVincenzo, P. Shor and J. Smolin, *Phys. Rev. A* **57**, 830 (1998).
[27] B. Schumacher, presentation at the Santa Fe Institute workshop on Complexity, Entropy and the Physics of Information (1994).
[28] S. L. Braunstein, C. A. Fuchs, D. Gottesman and H.-K. Lo, *IEEE Trans. Inf. Theory* (to appear, 2000).
[29] B. Schumacher and M. Westmoreland, "Indeterminate Length Quantum Coding" (in preparation).
[30] V. Vedral, M. B. Plenio, M. A. Rippin and P. L. Knight, *Phys. Rev. Lett.* **78**, 2275.
[31] C. H. Bennett, H. Bernstein, S. Popescu and B. W. Schumacher, *Phys. Rev. A* **53**, 3824 (1996).
[32] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin and W. K. Wootters, *Phys. Rev. A* **54**, 3824 (1996).
[33] N. Linden, S. Popescu, B. Schumacher and M. Westmoreland, "Reversibility of local transformations of multiparticle entanglement", submitted to *Phys. Rev. Lett.* LANL e-print quant-ph/9912039.

(Benjamin Schumacher) Department of Physics, Kenyon College, Gambier, OH 43022 USA

(Michael D. Westmoreland) Department of Mathematical Sciences, Denison University, Granville, OH 43023 USA