Quantum Coding

Benjamin Schumacher

Department of Physics

Kenyon College

Gambier, OH 43022

e-mail: schumacb@kenyon.edu

Abstract

A theorem is proven for quantum information theory that is analogous to the noiseless coding theorem of classical information theory. In the quantum result, the von Neumann entropy S of the density operator describing an ensemble of pure quantum signal states is equal to the number of spin-1/2 systems (quantum bits or "qubits") necessary to represent the signal faithfully. The theorem holds whether or not the signal states are orthogonal. Related results are also presented about the fidelity of quantum coding and about representing entangled quantum states.

PACS numbers: 03.65, 05.30

I. Entropy and Information

In the classical information theory developed by Shannon and others, the central problem is *coding* [1]. For example, suppose that A is a message source that produces the message a with probability $p(a)$, and further suppose that we wish to represent the messages with sequences of binary digits (bits) that are as short as possible. It can be shown that the mean length $\bar{L}$ of these bit sequences is bounded below by the Shannon entropy $H(A)$ of the source: $\bar{L} \geq H(A)$, where

$$H(A) = -\sum_{a} p(a) \log p(a). \tag{1}$$

(Throughout this paper, we use base-2 logarithms.) Furthermore, if we allow ourselves to code entire blocks of independent messages together, it turns out that the mean number $\bar{L}$ of bits per message can be brought arbitrarily close to $H(A)$.

This *noiseless coding theorem* shows the importance of the Shannon entropy $H(A)$ for information theory. It also provides an *interpretation* of $H(A)$ as the mean number of bits necessary to code the output of A using an ideal code. We might imagine that each bit has a fixed "cost" (in units of energy or space or money), so that $H(A)$ is a measure of the tangible resources necessary to represent the information produced by A.

The "entropy" is also of central importance in statistical mechanics, where it is a measure of the disorder of a physical system. In classical statistical mechanics, in fact, the statistical entropy is formally identical to the Shannon entropy. This has led to a considerable effort to give statistical mechanics an information-theoretic foundation [2]. In this approach, the entropy of a macrostate can be interpreted as the number of bits that would be required to specify the microstate of the system. (Of course, in classical statistical mechanics the phase space of states is a continuum, so that the number of bits needed to specify a microstate *completely* is infinite. This can be avoided in the usual way by specifying the microstate to a finite, and arbitrary, resolution in phase space.)

In quantum systems, however, the expression for entropy (first proposed by von Neumann [3]) is not identical to the Shannon entropy. Suppose ρ is the density operator describing an ensemble of states of a quantum system; then the von Neumann entropy $S(\rho)$ is

$$S(\rho) = -\mathrm{Tr}\,\rho \log \rho. \tag{2}$$

This has obvious analogies to the Shannon entropy; in fact, if we can interpret the probabilities $p(a)$ in Eq. (1) as eigenvalues of the density operator ρ, then $S(\rho)$ is numerically equal to $H(A)$.

Despite their formal similarity, however, these two quantities are quite different. We can see this difference easily by considering a *quantum signal source*, which might be part of a quantum communication system. This is a device which codes each message a from the source A into a "signal state" $|\,a_M\rangle$ of a quantum system M. Then the ensemble of signals from the signal source will be represented by the density operator

$$\rho = \sum_a p(a)\pi_a$$

where the density operators π_a are the projections $\pi_a = |\,a_M\rangle\,\langle a_M\,|$. The von Neumann entropy of ρ will equal the Shannon entropy of the message source only in the special case when the signals $|\,a_M\rangle$ are orthogonal to one another, in which case the signal states are eigenstates of ρ. If the signals are not orthogonal, then $S(\rho) < H(A)$, and the eigenstates of ρ may have no simple relation to the signal states [4].

Of course, if the signal states are not orthogonal it will not be possible to distinguish between them perfectly. In other words, no "decoding observable" will be sufficient to recover the entire information content of the message in the quantum signal source. It might therefore be more appropriate to consider the *accessible information*, the maximum amount of information about the message that can be recovered in a measurement performed on M. The proper measure of recovered information is the *mutual information*, which for a pair of random variables X and Y is defined to be

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

In classical information theory, the mutual information is the amount of information about X that is acquired by determining the value of Y. Thus, if we denote by B the outcome of a measurement of an observable on M, the quantity $H(A:B)$ measures the information about the message source A that is acquired by measurement of the observable. The maximum accessible information is the maximum of $H(A:B)$ over all choices of decoding observable.

An important theorem of Kholevo [5] places an upper bound on $H(A:B)$ for a quantum channel. To state the theorem most generally, suppose the signal states are described by density operators ρ_a and are not necessarily pure states. Then the mutual information obtained by the measurement of any "positive operator" observable (of which ordinary quantum observables are a special case) is bounded by

$$H(A:B) \leq S(\rho) - \sum_a p(a)\, S(\rho_a). \tag{4}$$

In our case, the signals $\rho_a = \pi_a$ are pure states with zero entropy and Kholevo's theorem simply states that, if ρ is the density operator for the signal ensemble,

$$H(A:B) \leq S(\rho). \tag{5}$$

That is, although the Shannon entropy $H(A)$ of the message source is in general greater than the von Neumann entropy $S(\rho)$ of the signal ensemble, the accessible information is bounded by $S(\rho)$. Kholevo's theorem provides a connection between the von Neumann entropy of a quantum ensemble and the Shannon mutual information of a quantum communication channel [6].

This connection, however, is a fairly weak one. Kholevo's theorem is an inequality, and it is possible to construct simple quantum signal sources for which the mutual information $H(A:B)$ does not approach $S(\rho)$ closely for any choice of decoding observable [7]. Thus, although Kholevo's theorem gives an information-theoretic significance to $S(\rho)$, it does not provide an interpretation of $S(\rho)$ in terms of classical information theory. We could not use Kholevo's theorem, for example, to interpret the quantum entropy of some macrostate of a thermodynamic system as a measure of the resources necessary to represent information about the system's quantum microstate.

In this paper we will prove a "quantum coding theorem" that does allow exactly this sort of interpretation. This is accomplished by replacing the classical idea of a binary digit with a quantum two-state system, such as the spin of electron. These quantum bits, or "qubits", are the fundamental units of quantum information. We will show that the von Neumann entropy $S(\rho)$ of an ensemble is just the mean number of qubits necessary to encode the states in the ensemble in an ideal coding scheme. This theorem can be viewed as the kernel of a new approach to quantum information theory: instead of simply applying classical information theory to probabilities derived from quantum rules, we can adopt notions of coding and measures of information that are themselves distinctly quantum mechanical.

Section II provides some background about the Shannon entropy and classical ideas of "likely" and "unlikely" binary sequences. Section III distinguishes between the *copying* of quantum information, which is in general impossible, and the *transposition* of that information from one system to another. The fidelity of a transposition scheme is defined in Section IV, and two useful lemmas about fidelity are proven in Section V. These lemmas lead directly to the proof of the main theorem in Section VI. Section VII discusses issues related to entangled quantum states, and Section VIII presents some general remarks.

II. Entropy and Likely Sequences

It will be useful here to review basic concepts of probability, particularly those relating to the Shannon entropy $H(A)$. We will also outline a proof of the noiseless coding theorem of classical information theory [8].

Suppose $x_1, \ldots, x_N$ are N independent, identically distributed random variables, each with mean $\bar{x}$ and finite variance. Given $\delta, \epsilon > 0$ there exists N_0 such that, for $N \geq N_0$,

$$P\left(\left|\frac{1}{N}\sum_i x_i - \bar{x}\right| > \delta\right) < \epsilon. \tag{6}$$

This standard result is known as the Weak Law of Large Numbers. It tells that a sufficiently long sequence of independent, identically distributed random variables will, with a probability approaching unity, have an average that is close to the mean of each variable.

We can use the Weak Law to derive a relation between the Shannon entropy $H(A)$ and the number of "likely" sequences of N identical random variables. Suppose, as before, that a message source A produces the message a with probability $p(a)$. A sequence $\alpha = a_1 a_2 \ldots a_N$ of N independent messages from the same source will occur in an ensemble of all N-sequences with probability $P(\alpha) = p(a_1) \cdots p(a_N)$. We can now define a random variable for each message by $x = -\log p(a)$, so that $H(A) = \bar{x}$. It is easy to see that

$$-\log P(\alpha) = \sum_i x_i.$$

The Weak Law then tells us that, if $\epsilon, \delta > 0$ then for sufficiently large N

$$P\left(\left|-\frac{1}{N}\log P(\alpha) - H(A)\right| > \delta\right) < \epsilon \tag{7}$$

for N-sequences α. We can therefore partition the set of all N-sequences into two subsets:

- A set Λ of "likely" sequences, for which
$$\left|-\frac{1}{N}\log P(\alpha) - H(A)\right| \leq \delta.$$
- A set of "unlikely" sequences with total probability less than ϵ, for which this inequality fails.

In other words, with probability greater than $1 - \epsilon$ a sequence α is in Λ and thus satisfies

$$-\delta \leq -\frac{1}{N}\log P(\alpha) - H(A) \leq \delta$$

which in turn implies that

$$2^{-N(H(A)-\delta)} \geq P(\alpha) \geq 2^{-N(H(A)+\delta)}. \tag{8}$$

How many likely sequences are there? Let ν be the number of sequences in Λ. Then, using the right-hand inequality above,

$$\begin{aligned} 1 &\geq \sum_{\alpha\in\Lambda} P(\alpha) \\ &\geq \sum_{\alpha\in\Lambda} 2^{-N(H(A)+\delta)} \\ &= \nu\, 2^{-N(H(A)+\delta)}. \end{aligned}$$

Therefore, the number ν of likely sequences is bounded by

$$\nu \leq 2^{N(H(A)+\delta)}. \tag{9}$$

By a similar argument, it turns out that

$$\nu \geq (1-\epsilon)2^{N(H(A)-\delta)}. \tag{10}$$

Long sequences of independent messages from the message source A thus fall into two classes: a collection of approximately $2^{NH(A)}$ likely sequences, and a collection of unlikely sequences with total probability approaching zero. This suggests a strategy for coding. The likely sequences may be associated in a one-to-one fashion with binary sequences of length $NH(A)$; unlikely sequences, though perhaps very numerous, are in some sense negligible.

More formally, we can prove the following theorem about coding the output of A into binary sequences:

> **Theorem (Noiseless Coding Theorem):** Let A be a message source as described above, and let $\delta, \epsilon > 0$.
>
> **(i)** Suppose $H(A)+\delta$ bits are available per A message. Then for sufficiently large N, N-sequences of messages from A can be coded into binary sequences with probability of error less than ϵ.

(ii) Suppose $H(A)-\delta$ bits are available per A message. Then for sufficiently large N, if N-sequences of messages from A are coded into binary sequences, the probability of error will be greater than $1-\epsilon$.

Part (i) is proved as follows. From our previous discussion, we know that, for large enough N, the number of likely sequences $\nu \leq 2^{N(H(A)+\delta)}$. We can thus encode each of the likely sequences into a unique sequence of the $N(H(A)+\delta)$ available binary digits. The remaining unlikely sequences can be "erroneously" encoded in any way, say into the single binary sequence $000\ldots00$. The total probability of error is thus the total probability of the unlikely sequences, which can be made less than ϵ.

The proof of part (ii) is slightly more involved. Let $\eta, \zeta > 0$. For sufficiently large N, we can distinguish between unlikely sequences, with total probability less than η, and likely sequences, each one of which has probability

$$P(\alpha) \leq 2^{-N(H(A)-\zeta)}.$$

For coding we have $2^{N(H(A)-\delta)}$ available binary sequences. We assign each of these binary sequences to a sequence of A messages. Any leftover sequences of A messages will then have to be encoded in such a way that they will not be correctly decoded—they will be errors. Let P_0 be the probability that the message sequence is not decoded erroneously. The set of correctly coded sequences is certainly smaller than all unlikely sequences plus $2^{N(H(A)-\delta)}$ likely sequences. Thus,

$$\begin{aligned} P_0 &< \eta + \left(2^{N(H(A)-\delta)}\right)\left(2^{-N(H(A)-\zeta)}\right) \\ P_0 &< \eta + 2^{-N(\delta-\zeta)}. \end{aligned}$$

Now let $\eta = \epsilon/2$, $\zeta = \delta/2$, and choose N large enough so that $2^{-N\delta/2} < \epsilon/2$. Then $P_0 < \epsilon$, and the probability of error $1 - P_0 > 1 - \epsilon$, as we wished.

Consider the notion of the "probability of error" in more detail. In itself, a coding scheme is incomplete; we also require some prescription for decoding, for recovering the original message from the binary string. If the code is one-to-one, this can be done unambiguously. If two possible messages are represented by the same binary sequence, however, this sequence will sometimes be decoded incorrectly.

We can define the *fidelity* F of the coding/decoding arrangement as the probability that the decoded message is the same as the message before coding. The probability of error is thus $1 - F$. A high fidelity means a low probability of error, and vice versa. Thus, the noiseless coding theorem states that, if more than $H(A)$ bits per message are allowed, the fidelity can be made arbitrarily close to unity; and conversely, if fewer than $H(A)$ bits per message are allowed, the fidelity eventually approaches zero.

The fidelity F will have an analogue in the quantum domain.

III. Copying and Transposition

A quantum signal source generates the signal state $|\, a_M \rangle$ of a quantum system M with probability $p(a)$. The signal states are not in general orthogonal. The ensemble of possible signals is described by the density operator

$$\rho = \sum_a p(a)\, \pi_a, \tag{11}$$

where $\pi_a = |\, a_M \rangle \langle a_M \,|$, the density operator (for a pure state, a projection) associated with the signal state vector $|\, a_M \rangle$.

In quantum coding, we wish to represent the output of the signal source in another quantum system X. Quantum information theory, unlike its classical counterpart, requires us to draw a distinction between the *copying* and the *transposition* of information from M into X. In copying, the original signal state of M is undisturbed and X is brought into a state corresponding to the signal state; that is, the combined system evolves according to

$$|\, a_M, 0_X \rangle \rightarrow |\, a_M, a_X \rangle \, , \tag{12}$$

where $|\, 0_X \rangle$ is some standard "null" state of X and $|\, a_X \rangle$ is the representation of the signal $|\, a_M \rangle$ in X.

As shown by Wootters and Zurek [9], copying a quantum signal faithfully cannot be accomplished for all signal sources. The proof of this is elementary. Suppose a device existed that claimed to copy arbitrary states of M into states of X. That is, given two distinct signal states $|\, a_M \rangle$ and $|\, b_M \rangle$ of M, the action of the device would be

$$\begin{aligned} |\, a_M, 0_X \rangle &\rightarrow |\, a_M, a_X \rangle \\ |\, b_M, 0_X \rangle &\rightarrow |\, b_M, b_X \rangle \, . \end{aligned}$$

Consider now the signal state $|\, c_M \rangle = |\, a_M \rangle + |\, b_M \rangle$. (We do not need to consider the normalization of $|\, c \rangle$.) If the resulting state of X is to be a faithful copy, then $|\, c_X \rangle = |\, a_X \rangle + |\, b_X \rangle$. But from general considerations of quantum mechanics, we know that the dynamical evolution of the system is *linear*—in fact, a unitary transformation—so that

$$\begin{aligned} |\, c_M, 0_X \rangle &= |\, a_M, 0_X \rangle + |\, b_M, 0_X \rangle \\ &\rightarrow |\, a_M, a_X \rangle + |\, b_M, b_X \rangle \\ &\neq |\, c_M, c_X \rangle \, , \end{aligned}$$

since $|\, c_M, c_X \rangle \;=\; |\, a_M, a_X \rangle + |\, a_M, b_X \rangle + |\, b_M, a_X \rangle + |\, b_M, b_X \rangle$. That is, if two distinct states can be copied faithfully, a superposition of the two states cannot be.[1]

Copying can be accomplished if the possible states are mutually orthogonal—for example, we could measure an observable whose eigenstates are the signal states and then use the (classical) information about the outcome to manufacture as many perfect copies as desired. Quantum signal sources which have non-orthogonal signals, on the other hand, cannot be duplicated perfectly.

Transposition is a different matter. In transposition, the signal state of M is transferred to X without leaving a copy behind:

$$|\, a_M, 0_X \rangle \to |\, 0_M, a_X \rangle \tag{13}$$

where $|\, 0_X \rangle$ and $|\, 0_M \rangle$ are fixed "null" states for X and M. After transposition, the signal resides completely in the coding system X and the original signal in M has been erased. (The "quantum teleportation" discussed in [10] is a rather exotic example of a transposition process.)

Transposition is completely unitary for arbitrary input signal states of M, provided that the coded states in X have the same inner products as their precursors: $\langle a_X \mid b_X \rangle = \langle a_M \mid b_M \rangle$ for all signals $|\, a_M \rangle$ and $|\, b_M \rangle$. This can be accomplished if and only if the Hilbert space $\mathcal{H}_X$ has a dimension at

[1] To be complete, we should include in this discussion the change in state of the copying device, which in general may depend upon the input state. The process is actually

$$|\, a_M, 0_X, \psi_0 \rangle \to |\, a_M, a_X, \psi_a \rangle$$

where $|\, \psi_0 \rangle$ and $|\, \psi_a \rangle$ are states of the copier and its environment. This refinement does not modify the general argument.

least as large as the subspace of $\mathcal{H}_M$ spanned by the signal states. (We can without loss of generality suppose that this subspace is the entirety of $\mathcal{H}_M$.)

To specify the unitary evolution U that accomplishes transposition, we only need to specify how an orthogonal basis for $\mathcal{H}_M$ is mapped into an orthogonal basis for $\mathcal{H}_X$. The evolution of all other signals follows by linearity. Transposition is invertible, since the signal state can be transferred back from X to M via the unitary transformation U^{-1}. We can therefore imagine a communication scheme based upon transposition. At the coding end, the signal of a source system M is transposed via the unitary evolution U into the coding system X. The system X is conveyed from the transmitter to the receiver. At the decoding end, the unitary evolution U^{-1} is employed to recover the signal state from X into M', an identical copy of system M. Symbolically,

$$M \xrightarrow{U} X \xrightarrow{U^{-1}} M'.$$

The system X is the *quantum channel* in this communication scheme, and supports the transposition of the state of M into M'.

We are concerned here with the transposition of quantum information. This process requires that the quantum channel X be "large enough" (i.e., have a Hilbert space $\mathcal{H}_X$ of high enough dimension) to represent the signals in M. For perfect transposition, this means that $\dim \mathcal{H}_X \geq \dim \mathcal{H}_M$. It may be, however, that a perfect transposition of the signal is unnecessary, so that we only wish to perform an *approximate* transposition of quantum information from M to M' via X. Depending on the characteristics of the signal source, we may be able to make do with a smaller quantum channel and still have an adequately faithful representation of the signal. To explore this question, we need to do two things: first, describe what we mean by an approximate transposition; and second, define a measure of the *fidelity* of the process and relate the fidelity to the size of the quantum channel.

IV. Approximate Transposition and Fidelity

Consider the quantum communication channel outlined above. The signal state of M is unitarily transposed into X and can then be perfectly recovered into the system M'. However, let us suppose that we do not send *all* of the system X from the transmitter to the receiver. Instead, we will suppose that X is composed of two subsystems, which we will call C (for "channel") and E (for "extra"). Only the channel subsystem C is conveyed to the receiver to be used for decoding the signal into M'; the extra subsystem E is simply discarded. Clearly, the signal cannot in general be recovered exactly from C alone, since it may be that $\dim \mathcal{H}_C < \dim \mathcal{H}_M$. On the other hand, it may be possible to recover some approximation of the signal. We will call this quantum communication scheme *approximate transposition from M to M' via the limited channel C*.

To recover the signal from C into M', we will add to the channel system C an auxiliary system E' that is a copy of the discarded extra system E, and then perform a transposition from $C + E'$ to M via the unitary evolution operator U'. Symbolically, our scheme is:

$$\begin{array}{ccccccccc} M & \xrightarrow{U} & C+E & \longrightarrow & C & \longrightarrow & C+E' & \xrightarrow{U'} & M'. \\ & & & \downarrow & & \uparrow & & & \\ & & & E & & E' & & & \end{array}$$

(It may be convenient to choose $U' = U^{-1}$—in other words, to decode the signal from $C + E'$ into M' using the inverse of the coding transposition operator. We will not make this a general requirement.)

To determine the effectiveness of this transposition scheme, we need a measure of its fidelity. Suppose the original signal state of M is $|\, a_M \rangle$, represented by the density operator $\pi_a = |\, a_M \rangle \langle a_M \,|$. The final signal in M' will be a state represented by the density operator w_a. Because the system E has been discarded in the transfer process, the final signal state is not necessarily a pure state, and so w_a is not generally a projection operator.

To check how close the final signal w_a is to the original π_a, we can perform a "validation measurement" of the observable π_a. The measurement has two possible results: 1, indicating that the final signal matches the original; or 0, indicating that the final signal differs from the original. The probability that w_a passes this validation test is $\mathrm{Tr}\, \pi_a\, w_a$. Let us define the *fidelity* F to be the overall probability that a signal from the signal ensemble in M that is transmitted to M' passes a validation test comparing it to its original. That is,

$$F = \sum_a p(a)\, \mathrm{Tr}\, \pi_a\, w_a. \tag{14}$$

The fidelity F is clearly between 0 and 1, and equals unity only in the case of perfect transposition of all possible signals. F will be close to unity if (1) signals with large probability $p(a)$ are very little distorted in transmission, so that w_a nearly equals π_a; and (2) the set of signals which are greatly distorted, having w_a very different from π_a, has a small total probability.

It is instructive to trace out how the signal state changes through this communication scheme. The first stage of our communication scheme, the unitary coding transposition from M to $C + E$, is accomplished via the operator U. If the original signal state of M is π_a, then the signal state of $C + E$ can be written $\Pi_a = U\, \pi_a\, U^{-1}$. When we discard the extra system E, the remaining system C must be assigned a state $\mathrm{Tr}_E\, \Pi_a$, the partial trace of Π_a over E. After E' (which is in some state $|\,0_{E'}\rangle$) has been adjoined to the channel C, the combined system is in a state $W_a = \mathrm{Tr}_E\, \Pi_a \otimes |\,0_{E'}\rangle\,\langle 0_{E'}\,|$. Finally, the unitary decoding transposition occurs and $w_a = U'\,W_a\,(U')^{-1}$. Suppose we make the reasonable choice $U' = U^{-1}$, so that the decoding transposition is just the operator inverse of the coding transposition. Then the fidelity is

$$
\begin{aligned}
F &= \sum_a p(a)\,\mathrm{Tr}\,\pi_a\,w_a \\
&= \sum_a p(a)\,\mathrm{Tr}\,\left(U^{-1}\,\Pi_a\,U\right)\left(U^{-1}\,W_a\,U\right) \\
&= \sum_a p(a)\,\mathrm{Tr}\,\Pi_a\,W_a
\end{aligned}
\tag{15}
$$

so that we can calculate the fidelity of the transposition from M to M' by examining only the signal states of $C+E$ and $C+E'$.

V. Two Fidelity Lemmas

Intuitively, we can say that, if the channel system C is "too small" then the fidelity F must be "close to" 0. Conversely, if the channel system is "large enough" then we will be able to make the fidelity F "close to" 1. Making rigorous theorems out of the phrases "large enough", "too small", and "close to" is the task of this section. We will prove a pair of general lemmas that will in the next section be central to the proof of our main coding theorem.

We begin by considering a channel C that is "too small". That is, we will prove the following lemma:

> **Lemma 1.** Suppose $\dim \mathcal{H}_C = d$, and suppose that the ensemble of signals in M described by $\rho = \sum_a p(a)\,\pi_a$ has the property that,

for any projection Γ onto a d-dimensional subspace $\mathcal{H}_M$,

$$\mathrm{Tr}\,\rho\,\Gamma < \eta$$

for some fixed η. Then the fidelity $F < \eta$.

If the signal ensemble has small "weight" in every subspace of the same size as $\mathcal{H}_C$, then the fidelity of the transposition will be correspondingly small. The proof of this fact follows. Consider a signal state $|\,a_M\rangle$ that is transposed according to our general scheme. Assuming that the system E' is initially in some pure state $|\,0_{E'}\rangle$, the signal state W_a of $C + E'$ is supported only on a d-dimensional subspace of $\mathcal{H}_{C+E'} = \mathcal{H}_C \otimes \mathcal{H}_{E'}$, the subspace of states of the form $|\,\psi_C, 0_{E'}\rangle$. Therefore the final decoded signal state w_a of M' is supported only on a d-dimensional subspace of $\mathcal{H}_{M'}$. Call the projection onto this subspace Γ.

Let $|\,\phi_k\rangle$ for $k = 1, \ldots, d$ be the orthogonal basis for this subspace that is composed of eigenstates of w_a. We can write w_a as

$$w_a = \sum_k q_k \,|\,\phi_k\rangle\,\langle\phi_k\,|$$

where the q_k are eigenvalues of w_a, including all of the non-zero ones. Clearly, $q_k \leq 1$. The projection Γ is simply

$$\Gamma = \sum_k |\,\phi_k\rangle\,\langle\phi_k\,|\,.$$

Now consider the term $\mathrm{Tr}\,\pi_a w_a$, which appears in the expression for the fidelity.

$$\mathrm{Tr}\,\pi_a w_a \;=\; \mathrm{Tr}\,\pi_a \left(\sum_k q_k \,|\,\phi_k\rangle\,\langle\phi_k\,| \right)$$

$$
\begin{aligned}
&= \sum_k q_k \,\mathrm{Tr}\, \pi_a \,|\, \phi_k\rangle \,\langle \phi_k \,| \\
&\leq \sum_k \,\mathrm{Tr}\, \pi_a \,|\, \phi_k\rangle \,\langle \phi_k \,| \\
&= \mathrm{Tr}\, \pi_a \left(\sum_k |\, \phi_k\rangle \,\langle \phi_k \,| \right) \\
&= \mathrm{Tr}\, \pi_a \Gamma.
\end{aligned}
$$

The fidelity F is

$$
\begin{aligned}
F &= \sum_a p(a) \,\mathrm{Tr}\, \pi_a w_a \\
&\leq \sum_a p(a) \,\mathrm{Tr}\, \pi_a \Gamma \\
&= \mathrm{Tr}\, \left(\sum_a p(a) \pi_a \right) \Gamma \\
&= \mathrm{Tr}\, \rho\, \Gamma
\end{aligned}
$$

and therefore $F < \eta$.

If the system E' is not in a pure state, the final signal state w_a will be a mixture of states, each of which is supported on a d-dimensional subspace. The overall fidelity will be a weighted average of terms that are bounded in the above manner. Thus, $F < \eta$ in this more general situation as well, and the theorem holds.

It is worth remarking that the condition requiring $\mathrm{Tr}\, \rho \Gamma < \eta$ for all projections Γ can be rephrased in terms of the eigenvalues of ρ. Let P_n be the eigenvalues of ρ, and let $|\, n\rangle$ be the corresponding eigenstates. The quantities $Q_n = \langle n \,|\, \Gamma \,|\, n\rangle$ satisfy $0 \leq Q_n \leq 1$, and $\sum_n Q_n = \mathrm{Tr}\, \Gamma = d$. Then

$$
\mathrm{Tr}\, \rho\, \Gamma = \sum_n P_n \, Q_n.
$$

It is easy to see that this sum will be maximized if the Q_n are chosen to be 1 for the values of n corresponding to the d largest eigenvalues P_n, and zero for other values of n. We can actually achieve this largest value by choosing Γ to be the projection onto the subspace spanned by the eigenstates with the d largest eigenvalues of ρ. Thus, $\mathrm{Tr}\,\rho\,\Gamma < \eta$ if and only if the sum of any d eigenvalues of ρ is less than η.

We next turn to the case in which the channel C is "large enough" to allow transposition with high fidelity.

> **Lemma 2.** Suppose that $\dim \mathcal{H}_C = d$, and suppose that there exists a projection Γ onto a d-dimensional subspace of $\mathcal{H}_M$ such that
> $$\mathrm{Tr}\,\rho\,\Gamma > 1 - \eta.$$
> Then there exists a transposition scheme with fidelity $F > 1 - 2\eta$.

If the signal ensemble has sufficient "weight" on a subspace of the same size as $\mathcal{H}_C$, then it is possible to make a transposition with a fidelity that is correspondingly close to 1. To prove this, we will actually construct such a transposition scheme and find its fidelity.

We first note that, from the remark above, we lose no generality if we suppose that Γ is a projection onto a subspace Λ of $\mathcal{H}_M$ spanned by d eigenstates of ρ. Let us number the eigenstates of ρ in such a way that the eigenstates $|\,1\rangle\,,\ldots,|\,d\rangle$ span Λ, while $|\,d+1\rangle\,,\ldots,|\,D\rangle$ (where $D = \dim \mathcal{H}_M$) are orthogonal to Λ and span the orthogonal subspace $\Lambda^{\perp}$. Then we have

$$\begin{aligned}
\sum_{n=1}^{d} |\,n\rangle\,\langle n\,| &= \Gamma \\
\sum_{n=1}^{d} P_n &> 1 - \eta \\
\sum_{n=d+1}^{D} P_n &< \eta .
\end{aligned}$$

Our strategy is as follows. We transpose the eigenstates of ρ that are in Λ in such a way that they will be faithfully represented by states of the channel C and will be correctly reconstructed in M'. We can do this since $\dim \Lambda = \dim \mathcal{H}_C$. However, the eigenstates of ρ are not necessarily signal states, and we have no guarantee that *any* signal actually lies within Λ and is thus transposed without distortion. Nevertheless, since Λ includes most of the "weight" of the signal ensemble (except for a small piece of measure less than η), we will be able to show that enough of the signals are sufficiently close to the subspace Λ to achieve the required fidelity.

To specify the unitary transformation U that accomplishes the coding transposition, we specify how the orthogonal basis of ρ eigenstates is mapped into orthogonal states of $C + E$. Consider the following mapping:

$$|\, n\rangle \longrightarrow \begin{cases} |\, n_C, 0_E\rangle & n = 1, \ldots, d \\ |\, 0_C, n_E\rangle & n = d+1, \ldots, D \end{cases} \tag{16}$$

where the $|\, n_C\rangle$ and $|\, n_E\rangle$ are orthogonal sets of states of the systems C and E, respectively, and $|\, 0_C\rangle$ and $|\, 0_E\rangle$ are fixed "null" states. We require that the null state $|\, 0_E\rangle$ be orthogonal to each of the $|\, n_E\rangle$ for $n = d+1, \ldots, D$. Roughly speaking, states in Λ are mapped into states of C and states in $\Lambda^\perp$ are mapped into states of E. More precisely, the *distinction* between states in Λ is now made between states of C, and the distinction between states in $\Lambda^\perp$ is now made between states in E.

The extra system E is now discarded, and a new copy E' is joined to the system. We specify that E' be initially in the state $|\, 0_{E'}\rangle$, so that the ρ eigenstates are now mapped into states

$$|\, n\rangle \longrightarrow \begin{cases} |\, n_C, 0_{E'}\rangle & n = 1, \ldots, d \\ |\, 0_C, 0_{E'}\rangle & n = d+1, \ldots, D \end{cases} . \tag{17}$$

Finally, we decode the signal into M' by using the inverse U^{-1} of the coding transformation.

How does a particular signal state $|\, a_M\rangle$ of M fare in this approximate transposition scheme? We can write any state of M as a superposition of states

$$|\, a_M\rangle = \lambda_a \,|\, \lambda(a)_M\rangle + \mu_a \,|\, \mu(a)_M\rangle \,, \tag{18}$$

where $|\, \lambda(a)_M\rangle$ is in Λ and $|\, \mu(a)_M\rangle$ is in $\Lambda^\perp$, and $|\lambda_a|^2 + |\mu_a|^2 = 1$. The states $|\, \lambda(a)_M\rangle$ and $|\, \mu(a)_M\rangle$ can be expanded in terms of the basis eigenstates of ρ:

$$|\, a_M\rangle = \lambda_a \left(\sum_{n=1}^{d} \langle n \,|\, \lambda(a)_M\rangle \,|\, n\rangle \right) + \mu_a \left(\sum_{n=d+1}^{D} \langle n \,|\, \mu(a)_M\rangle \,|\, n\rangle \right).$$

This state is mapped by the coding transposition U into a state $|\, a_{C+E}\rangle$, where

$$\begin{aligned}
|\, a_M\rangle \;\; &\longrightarrow \;\; |\, a_{C+E}\rangle \\
&= \;\; \lambda_a \left(\sum_{n=1}^{d} \langle n \,|\, \lambda(a)_M\rangle \,|\, n_C, 0_E\rangle \right) + \mu_a \left(\sum_{n=d+1}^{D} \langle n \,|\, \mu(a)_M\rangle \,|\, 0_C, n_E\rangle \right) \\
&= \;\; \lambda_a \,|\, \lambda(a)_C, 0_E\rangle + \mu_a \,|\, 0_C, \mu(a)_E\rangle
\end{aligned}$$

where $|\, \lambda(a)_C\rangle$ and $|\, \mu(a)_E\rangle$ have the obvious definitions. Parenthetically, we note that $\langle \mu(a)_E \,|\, 0_E\rangle = 0$.

The signal state of $C + E$ can be written as a projection operator $\Pi_a = |\, a_{C+E}\rangle \langle a_{C+E} \,|$, which is

$$\begin{aligned}
\Pi_a \;\; = \;\; & |\lambda_a|^2 \;|\, \lambda(a)_C, 0_E\rangle \langle \lambda(a)_C, 0_E \,| \\
& + \lambda_a \mu_a^* \;|\, \lambda(a)_C, 0_E\rangle \langle 0_C, \mu(a)_E \,| \\
& + \lambda_a^* \mu_a \;|\, 0_C, \mu(a)_E\rangle \langle \lambda(a)_C, 0_E \,| \\
& + |\mu_a|^2 \;|\, 0_C, \mu(a)_E\rangle \langle 0_C, \mu(a)_E \,| \,.
\end{aligned} \tag{19}$$

When E is discarded, the state of the channel C is obtained by performing a partial trace on Π_a, yielding

$$\begin{aligned} \mathrm{Tr}_E \Pi_a &= |\lambda_a|^2 \; |\, \lambda(a)_C \rangle \, \langle \lambda(a)_C \,| \\ &\quad + |\mu_a|^2 \; |\, 0_C \rangle \, \langle 0_C \,| \,. \end{aligned}$$

Adjoining the system E' in the state $|\, 0_{E'} \rangle$ yields W_a:

$$\begin{aligned} W_a &= |\lambda_a|^2 \; |\, \lambda(a)_C, 0_{E'} \rangle \, \langle \lambda(a)_C, 0_{E'} \,| \\ &\quad + |\mu_a|^2 \; |\, 0_C, 0_{E'} \rangle \, \langle 0_C, 0_{E'} \,| \,. \end{aligned} \tag{20}$$

Since we decode this signal into M' using the inverse of the coding transposition, the overall fidelity is just $F = \sum_a p(a) \, \mathrm{Tr} \, \Pi_a \, W_a$. For a given signal,

$$\begin{aligned} \mathrm{Tr} \, \Pi_a \, W_a &= |\lambda_a|^4 + |\lambda_a|^2 \, |\mu_a|^2 \, |\, \langle \lambda(a)_C \,|\, 0_C \rangle \,|^2 \\ &\geq |\lambda_a|^4 \\ &= \left(1 - |\mu_a|^2\right)^2 \\ &\geq 1 - 2|\mu_a|^2. \end{aligned}$$

The fidelity is thus

$$F \geq 1 - 2 \left(\sum_a p(a) |\mu_a|^2 \right). \tag{21}$$

We required that $\mathrm{Tr} \, \rho\Gamma > 1 - \eta$. This is just

$$\begin{aligned} \mathrm{Tr} \, \rho\Gamma &= \sum_a p(a) \, \mathrm{Tr} \, \pi_a \Gamma \\ &= \sum_a p(a) |\lambda_a|^2 \\ &= \sum_a p(a) (1 - |\mu_a|^2) \\ &= 1 - \sum_a p(a) |\mu_a|^2. \end{aligned}$$

Therefore, our requirement on $\mathrm{Tr}\,\rho\Gamma$ amounts to requiring that $\sum_a p(a)|\mu_a|^2 < \eta$. This means that the fidelity of our coding scheme satisfies

$$F > 1 - 2\eta,$$

as we wished to prove.

To summarize, we have related the fidelity of our transposition scheme to the dimension d of the Hilbert state space of the channel system C. If d is small enough that the signal ensemble ρ has weight less than η on every d-dimensional subspace of $\mathcal{H}_M$, then the fidelity must satisfy $F < \eta$. On the other hand, if we can find a d-dimensional subspace of $\mathcal{H}_M$ on which ρ has a weight greater than $1 - \eta$, we can devise a transposition scheme with fidelity $F > 1 - 2\eta$. Furthermore, we can restrict our attention to subspaces spanned by eigenstates of ρ, establishing the existence or non-existence of suitable subspaces by considering sums of d distinct eigenvalues of ρ. This connection between the eigenvalues of ρ (which form a probability distribution) and the fidelity of approximate transposition through a limited channel C will be important in the next section.

VI. Qubits and Quantum Coding

In the classical noiseless coding theorem, there are three central features. First, we specified a single elementary coding system, the binary digit or "bit"; all messages were encoded using bits. Second, we allowed ourselves to encode, not individual messages, but entire sequences of N messages from independent, identical sources. Third, we did not require that the coding be completely error-free; it sufficed that we could make the classical fidelity (the probability of encoding the message in a correctly decodable way) arbitrarily close to unity.

Each of these three central ideas must be adapted to our quantum context. We have already done this with the third item, the fidelity criterion. In quantum coding, the fidelity F is the probability that a transposed signal state will pass a validation test comparing it to the original signal. The fidelity lemmas of the previous section give us bounds on F in various situations. It remains for us to consider the quantum generalizations of the first two features, the elementary coding system and the "block coding" of long sequences of messages.

For our elementary coding system we choose the two-level spin system, which we will call a "quantum bit" or *qubit*. The qubit will be our fundamental unit of quantum information, and all signals will be encoded into sequences of qubits. Let us denote a single qubit system by Q. Our quantum channel C will be composed of some (possibly large) number K of copies of Q (denoted Q^K), so that

$$\mathcal{H}_C = \underbrace{\mathcal{H}_Q \otimes \cdots \otimes \mathcal{H}_Q}_{K \text{ times}}.$$

The dimension of $\mathcal{H}_C$ is 2^K. The analogy between the bit and the qubit is obvious. The qubit is more general, since there are more possible coding states than just two (although there are only two orthogonal ones), and since a collection of qubits can exist in a non-classically entangled quantum state.

To discuss quantum block coding, we must consider an *extended* quantum signal source. This is a system consisting of N independent copies of the system M (which we will denote M^N). Each subsystem M_k is in a signal state $|a_k\rangle$, generated according to the signal probability distribution $p(a_k)$.

That is, the system M^N is in the state

$$|\,\alpha\rangle = |\,a_1, a_2, \ldots, a_N\rangle$$

with probability $p(\alpha) = p(a_1)p(a_2)\cdots p(a_N)$. Our job will be to transpose the signal state of the joint system M^N into an identical system $(M')^N$ using a channel composed of qubits.

The density operator ρ^N describing the signal ensemble of M^N is simply the direct product of the density operators for the signal ensembles of the individual subsystems: $\rho^N = \rho_1 \otimes \cdots \otimes \rho_N$. This means that the eigenstates of ρ^N will be products states $|\,n_1, \ldots, n_N\rangle$ of eigenstates of the ρ_i, and the eigenvalues of ρ^N will be products of eigenvalues of the ρ_i:

$$P_{n_1,\ldots,n_N} = P_{n_1}\cdots P_{n_N}.$$

Now, for the system M, the eigenvalues P_n ($n = 1, \ldots, D$) of the signal ensemble density operator ρ have all of the properties of a probability distribution over the integers 1 through D. (They are in fact the probability distribution for the outcomes of a complete measurement with eigenstates $|\,n\rangle$.) Furthermore, the von Neumann entropy of ρ is just the Shannon entropy of this distribution:

$$S(\rho) = -\sum_n P_n \log P_n. \tag{22}$$

An eigenstate of ρ^N corresponds to a sequence $n_1, \ldots, n_N$ of N integers, and the eigenvalue of this eigenstate is just the probability of this sequence if it had been generated by N independent trials using the probability distribution P_n.

As long as we are only interested in eigenvalues and eigenstates of the density operators, we can pretend that each quantum signal source is a classical message source that uses the integers 1 through D as its "alphabet" and has a message probability distribution P_n with Shannon entropy $S(\rho)$. The extended signal source is the extended message source of sequences of these integers. From our discussion above, we know that for large enough N the sequences can be divided into two sets:

- a set of about $2^{N\,S(\rho)}$ "likely" sequences, and
- a set of sequences with small total probability—i.e., whose corresponding eigenvalues have a small sum.

These two sets specify two orthogonal subspaces of the Hilbert space $\mathcal{H}_{M^N}$. One of these (which we can call the "likely" subspace Λ) has a dimension of about $2^{N\,S(\rho)}$, and could be faithfully transposed into the states of a collection of $NS(\rho)$ qubits. The other, $\Lambda^{\perp}$, has small "weight" with respect to ρ^N, and therefore will not affect the fidelity too much.

To be exact, we can now prove the following theorem:

> **Theorem (Quantum Noiseless Coding Theorem):** Let M be a quantum signal source with a signal ensemble described by the density operator ρ, and let $\delta, \epsilon > 0$.
>
> **(i)** Suppose that $S(\rho)+\delta$ qubits are available per M signal. Then for sufficiently large N, groups of N signals from the signal source M can be transposed via the available qubits with fidelity $F > 1 - \epsilon$.

(ii) Suppose that $S(\rho) - \delta$ qubits are available per M signal. Then for sufficiently large N, if groups of N signals from the signal source M are transposed via the available qubits then the fidelity $F < \epsilon$.

Part (i) is proved in this way. We first note that, if the quantum channel C is $N(S(\rho) + \delta)$ qubits Q, then $\dim \mathcal{H}_C = 2^{N(S(\rho)+\delta)}$. From our proof of the classical theorem and our probability-eigenvalue analogy, we know that, for large enough N, the number of "likely" eigenstate sequences $\nu \leq 2^{N(S(\rho)+\delta)}$ and the sum of the remaining eigenvalues of ρ^N can be made less than $\epsilon/2$. We can add a few additional eigenstate sequences if necessary to the "likely" set to bring the total to exactly $\nu = \dim \mathcal{H}_C$, and this will not increase the sum of the remaining eigenvalues. Let Γ be the projection onto the ν-dimensional subspace Λ spanned by the "likely" eigenstates. Then $\operatorname{Tr} \rho^N \Gamma > 1 - \epsilon/2$. By Lemma 2 there is a transposition scheme with fidelity $F > 1 - \epsilon$.

For part (ii), we simply note that our classical discussion tells us that, for large enough N, no $2^{N(S(\rho)-\delta)}$ eigenstate sequences for M^N have eigenvalues which have a sum as large as ϵ. Therefore, for every projection Γ onto a subspace of dimension $2^{N(S(\rho)-\delta)} = \dim \mathcal{H}_C$, we have $\operatorname{Tr} \rho^N \Gamma < \epsilon$. Then by Lemma 1, every transposition scheme has fidelity $F < \epsilon$. Both parts of the theorem are now proved.

VII. Entangled Systems

We have considered the situation in which various quantum states (the signal states of M) were generated probabilistically, so a density operator ρ is necessary for the description of the mixed state of the ensemble. Density operators also arise when the system M is only part of a larger system $M+Z$ which is in a pure but entangled quantum state $|\,\psi_{M+Z}\rangle$. Such a state can always be written in a *polar decomposition* (sometimes called the "Schmidt decomposition" [11])

$$|\,\psi_{M+Z}\rangle = \sum_n \sqrt{P_n}\,|\,n_M, n_Z\rangle\,, \tag{23}$$

where the $|\,n_M\rangle$ and the $|\,n_Z\rangle$ are orthogonal sets of states for M and Z, respectively. To write down a state of M alone, we must do a partial trace over Z, yielding the density operator

$$\rho = \sum_n P_n\;|\,n_M\rangle\,\langle n_M\,|\,, \tag{24}$$

that is, a density operator with eigenstates $|\,n_M\rangle$ and eigenvalues P_n. The von Neumann entropy $S(\rho)$ of this density operator is sometimes cited as an "information-theoretic" measure of the degree of entanglement between the quantum systems M and Z [12, 6].

Suppose we now perform an approximate transposition from M to M'. Does this transposition faithfully transpose the overall quantum state of $M+Z$ into a state of $M'+Z$? In other words, does our scheme also faithfully transfer the quantum entanglement of the system with the rest of the world from M to M'?

The answer is yes. The state $|\,\psi_{M+Z}\rangle$, which can be represented by the projection $\pi = |\,\psi_{M+Z}\rangle\,\langle\psi_{M+Z}\,|$, is transposed into some final state w of $M'+Z$. The fidelity F of this process is

$$F = \mathrm{Tr}\,\pi w.$$

In our transposition scheme, we perfectly transpose via the channel C those eigenstates of ρ that are in a "likely" subspace Λ of $\mathcal{H}_M$ with $\dim\Lambda =$

$\dim \mathcal{H}_C = d$. We can number our eigenstates n_M as before so that the first d of them lie in Λ. Now we can write our overall state as

$$|\,\psi_{M+Z}\rangle = \lambda\,|\,\lambda_{M+Z}\rangle + \mu\,|\,\mu_{M+Z}\rangle$$

where $|\,\lambda_{M+Z}\rangle$ and $|\,\mu_{M+Z}\rangle$ are orthogonal normalized states

$$\begin{aligned}
|\,\lambda_{M+Z}\rangle &= \sum_{n=1}^{d} c_n\,|\,n_M, n_Z\rangle \\
|\,\mu_{M+Z}\rangle &= \sum_{n=d+1}^{D} c'_n\,|\,n_M, n_Z\rangle
\end{aligned}$$

and $|\lambda|^2 + |\mu|^2 = 1$. We can represent this state by the projection $\pi = |\,\psi_{M+Z}\rangle\,\langle\psi_{M+Z}\,|$.

In our scheme, the state $|\,\lambda_{M+Z}\rangle$ is transposed perfectly and the state $|\,\mu_{M+Z}\rangle$ is transposed into some mixed state, yielding a mixed state w of the system $M' + Z$. This mixed state is

$$\begin{aligned}
w &= |\lambda|^2\,|\,\lambda_{M'+Z}\rangle\,\langle\lambda_{M'+Z}\,| \\
&\quad + |\mu|^2 \left(\sum_{n=d+1}^{D} |c'_n|^2\,|\,0_{M'}, n_Z\rangle\,\langle 0_{M'}, n_Z\,| \right).
\end{aligned}$$

From this it follows that the fidelity $F = \mathrm{Tr}\,\pi w \geq |\lambda|^4$, as before. If $|\lambda|^2 > 1 - \eta$, then $F > 1 - 2\eta$, which is the same result we obtained in Lemma 2.

Once we have this result, we can repeat the argument of our coding theorem for N copies of the entangled system $M + Z$, concluding that the entanglement between M and Z can be faithfully transposed from M to M' using channel C with $S(\rho)$ qubits—or more exactly, that we can faithfully transpose the entanglement of many such systems if we have at least $S(\rho)$ qubits per system. We can therefore interpret $S(\rho)$ as a measure of the physical resources necessary to faithfully move the quantum entanglement between M and the rest of the world (the system Z) from one system to another.

VIII. Remarks

The von Neumann entropy $S(\rho)$ of a signal ensemble of pure states can be interpreted as the number of qubits per signal necessary to transpose it with near-perfect fidelity. If more than $S(\rho)$ qubits are available per signal, we can in the long run make the fidelity F as close to unity as we like; but if fewer than $S(\rho)$ are available per signal, the fidelity F will eventually approach zero. Furthermore, $S(\rho)$ is also (in the same sense) the number of qubits needed to transpose a part of a entangled system while maintaining the fidelity of the overall state near to unity. Thus, the quantum entropy S is a measure of the physical resources necessary to represent the information content of a system in a mixed state, whether the mixed state arises from a stochastic process or by the tracing out of quantum entanglement with the external world. Quantum entropy is measured in qubits.

In the proof of this theorem, a great deal of the mathematical machinery developed for the classical theorem could be inherited with only minor changes. Instead of probability distributions, we considered sets of eigenvalues of density operators. The two fidelity lemmas that we proved allowed us to connect statements about these eigenvalues to statements about the fidelity of an approximate transposition scheme over a limited channel.

A simpler approach to the process of approximate transposition and the fidelity lemmas has been suggested by R. Jozsa [13]. This avoids the explicit use of the auxiliary systems E and E' by invoking a (non-unitary) measurement process [14].

We should note that the argument for the coding theorem given here is somewhat akin to the work of Neill Graham in his paper on the many-worlds interpretation of quantum mechanics [15]. Graham investigated how the probabilities of measurement results might arise in the many-worlds interpretation, given that the final entangled state of the system and measuring apparatus always contains all possible outcomes in the superposition. By considering the final state of a large collection of system-apparatus pairs, Graham showed that the Hilbert space can be decomposed into two subspaces: (1) a "typical" subspace, in which the statistical frequencies of the measurement results closely match the probabilities given by the quantum rules, and (2) an "atypical" subspace which contributes very little to the final superposition, even though its dimension might be very large.

Some unresolved questions arise from our work. We have considered here only *pure* signal states π_a from our quantum signal source M. Suppose instead that the signals are mixed states ρ_a, with $\rho = \sum_a p(a)\rho_a$. Then it is not clear how to proceed, for the natural generalization of the fidelity,

$$F \stackrel{?}{=} \sum_a p(a) \operatorname{Tr} \rho_a w_a$$

may not be close to unity even if $w_a = \rho_a$ for all signals.

Furthermore, we have proved a "noiseless" coding theorem. Shannon's more powerful results deal with the information capacity of channels with noise [1]. It would be very desirable to develop coding theorems for noisy quantum channels.

Nevertheless, the fidelity and coding results presented here may be a starting point for a new approach to quantum information theory, one with possible applications to quantum cryptography [16] and the theory of quantum computers [17]. They also provide an information-theoretic interpretation of the von Neumann entropy, with potential implications for the conceptual foundations of quantum statistical mechanics.

Acknowledgments

The term "qubit" was coined in jest during one of the author's many intriguing and valuable conversations with W. K. Wootters, and became the initial impetus for this work. The author is also grateful to C. H. Bennett and R. Jozsa for their helpful suggestions and for numerous words of encouragement.

References

[1] C. E. Shannon, Bell System Technical Journal **27**, 379 (1948).

[2] E. Jaynes, Phys. Rev. **106**, 620 (1957); and E. Jaynes, Phys. Rev. **108**, 171 (1957).

[3] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (English translation by R. T. Beyer), Princeton University Press (1955).

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

[5] A. S. Kholevo, *Problemy Peredachi Informatsii* **9**, 3 (1973). This paper has appeared in English translation in *Problems of Information Transmission* **9**, 177 (1973).

[6] B. W. Schumacher, *Communication, Correlation, and Complementarity*, Ph.D. thesis (The University of Texas at Austin, 1990).

[7] A. Peres and W. K. Wootters, *Phys. Rev. Lett.* **66**, 1119 (1991).

[8] See any textbook on information theory for more about entropy and coding; for example, M. Mansuripur, *Introduction to Information Theory*, Prentice-Hall (1987).

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

[10] C. H. Bennett, G. Brassard, C. Crèpeau, R. Jozsa, A. Peres, and W. K. Wootters, *Phys. Rev. Lett.* **70**, 1895 (1993).

[11] E. Schmidt, *Mathematishe Annalen* **63**, 433 (1906).

[12] H. Everett, in *The Many-Worlds Interpretation of Quantum Mechanics,*, edited by B. S. DeWitt and N. Graham, Princeton (1973), pp. 3–140.

[13] R. Jozsa, private communication.

[14] R. Jozsa and B. W. Schumacher, *J. Mod. Optics*, to appear.

[15] N. Graham, in *The Many-Worlds Interpretation of Quantum Mechanics,*, edited by B. S. DeWitt and N. Graham, Princeton (1973), pp. 229–253.

[16] S. Wiesner, *SIGACT News* **15:1**, 78 (1983); C. H. Bennett and G. Brassard, *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing*, Bangalore, India, December 1984, pp. 175–179; C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, *J. Cryptography* **5**, 3 (1992).

[17] D. Deutsch *Proc. Roy. Soc. Lond.* **A 400**, 97 (1985).