General fidelity limit for quantum channels

Howard Barnum, Christopher A. Fuchs,[*] Richard Jozsa,[†] and Benjamin Schumacher[‡]
Center for Advanced Studies, Department of Physics and Astronomy,
University of New Mexico, Albuquerque, New Mexico 87131-1156
(Received 6 March 1996; revised manuscript received 15 July 1996)

We derive a general limit on the fidelity of a quantum channel conveying an ensemble of pure states. Unlike previous results, this limit applies to arbitrary coding and decoding schemes. This establishes in full generality the quantum noiseless coding theorem, formerly obtained only for the special case of unitary decoding schemes. [S1050-2947(96)06611-5]

PACS number(s): 03.65.Bz, 05.30.−d, 89.70.+c

I. QUANTUM ENCODING AND DECODING

One of the central problems in quantum information theory [1] is the transmission of pure quantum states from a sender to a receiver using the least possible channel resources. Suppose Alice generates the state $|a_i\rangle$ of the system Q with probability p_i. This is encoded by some (possibly mixed) state W_i of the channel system C (generally of smaller Hilbert-space dimension than Q) and delivered to Bob, who performs a decoding operation giving a state w_i of Q. We assume that no ''noise'' is present in the system except that introduced in the coding and decoding processes. Letting $\pi_i = |a_i\rangle\langle a_i|$, this may be represented by

$$\pi_i \to W_i \to w_i.$$

The decoded state w_i is not necessarily required to equal π_i exactly; it will suffice for Alice and Bob if the inputs and outputs are sufficiently close to each other. The ''closeness'' of the input and output states is measured by the *average fidelity* $\bar{F}$:

$$\bar{F} = \sum_i p_i F(\pi_i, w_i), \tag{1}$$

where $F(\pi_i, w_i) = \mathrm{Tr}\,\pi_i w_i$ is the probability that w_i will pass a test that checks its identity against π_i. Alice and Bob will succeed in their task if $\bar{F}$ is close to unity, and fail if it is not. Our problem is to characterize the minimal channel resources, i.e., the minimal dimension of the support of the states W_i, which are necessary and sufficient for high fidelity transmission [2,3].

This process of retrieving faithful copies of the input states from the states of the channel has applications in quan-

tum cryptography, where nonorthogonal states represent encrypted classical information [4,5], and in problems of efficient information storage and retrieval for quantum computers [6].

The decoding operation $W_i \to w_i$ must be accomplished without any ''side information''—i.e., the only information possessed by Bob about the input state is his knowledge of the message ensemble and the coding procedure that prepares the channel C. Bob's decoding procedure must be a dynamical evolution that is specified apart from the state on which it acts. On the other hand, we make no such assumption about Alice's encoding operation, so that the association $\pi_i \to W_i$ is completely arbitrary. Indeed we generally allow Alice to have knowledge of the identities of the specific input states and she is therefore able to effect arbitrary encodings. In contrast, Bob is unable to reliably identify the (generally nonorthogonal) channel states W_i [7,8] so his decoding procedure is restricted by the laws of quantum mechanics as described in Sec. III below.

Note that the encoding procedure here is more general than the scenario in which Alice is required to encode the input states *without* knowledge of their identities (knowing only their *a priori* distribution). In this situation the allowable encodings $\pi_i \to W_i$ are no longer arbitrary but subject to restrictions analogous to those on Bob's decoding procedures. (This is in contrast [8,9] to the corresponding situation with *classical* signals which may always be reliably identified without disturbance.) A remarkable consequence of the quantum noiseless coding theorem, and its converse described below, is that the minimal channel resources for high fidelity transmission in this situation are asymptotically the *same* as those for the case where Alice is able to apply arbitrary encoding processes, i.e., knowledge of the identity of the input states does not lead to any reduction of channel resources. Indeed in [2,3] an explicit encoding scheme is described which achieves (asymptotically) the minimal channel resources and this scheme operates without knowledge of the identity of the input states (being dependent only on their *a priori* distribution).

The quantum noiseless coding theorem proved in [2,3] relates the achievable average fidelity $\bar{F}$ to the size of the channel system. This size is given in terms of the number of two-level systems, or *qubits*, that comprise the channel when

coding is performed on large blocks of signals drawn identically from the original message ensemble.[1] Suppose we have input states π_i with probabilities p_i, as before, and let $\rho = \Sigma_i p_i \pi_i$ be the density operator describing the input ensemble. The von Neumann entropy of ρ is given by

$$S(\rho) = -\text{Tr}\rho\log\rho, \qquad (2)$$

where the base of the logarithm is 2. Then the quantum noiseless coding theorem states:

Let $\epsilon, \delta > 0$, and suppose $S(\rho) + \delta$ qubits are available in the channel per input state. Then for all sufficiently large N, there exists a coding and a decoding scheme which transmits blocks of N states with average fidelity $\overline{F} > 1 - \epsilon$.

In other words, the von Neumann entropy is a measure of the channel resources (in qubits) sufficient to transmit quantum states with arbitrarily high average fidelity. It is important to note that the fidelity of interest is the fidelity of the *entire block* of N input states. A converse to the theorem has also been given.

Let $\epsilon, \delta > 0$, and suppose $S(\rho) - \delta$ qubits are available in the channel per input state. Then for all sufficiently large N, for any coding and decoding scheme for blocks of N states, the average fidelity satisfies $\overline{F} < \epsilon$.

This converse states that the von Neumann entropy is a measure of the channel resources *necessary* to transmit quantum states with high average fidelity for the entire block.

In this formulation, the converse refers to all possible coding-decoding schemes. However, the proof given in [2] and [3] implicitly assumes that the decoding scheme is *unitary*—that is, that the map $W_i \to w_i$ is a unitary mapping from the channel's Hilbert space into the Hilbert space of the decoded signals. There are still other possibilities that must be considered. For example, the decoding scheme might involve a measurement, the discarding of an entangled subsystem, or any other process allowed within the the laws of physics. The converse of the quantum noiseless coding theorem cannot be established in full generality without considering all conceivable decoding schemes. Indeed in the Appendix we present a simple example containing all the salient features of this problem that shows for *particular* (nonoptimal) encodings it is possible for nonunitary decodings to provide higher fidelity than any unitary decoding scheme. Therefore, the issue of real concern for the converse is whether such nonunitary decoding schemes add any power to optimal encodings.

Our aim in this paper is to complete the general proof of the converse of the quantum noiseless coding theorem by establishing a lemma that links the average fidelity $\overline{F}$ of the decoded signal states to the size of the channel system and to

---

[1] Of course, the description of the channel in terms of qubits is mere convenience. Any channel described by a Hilbert space of dimension d is equivalent for our purposes to $\log d$ qubits.

properties of the density operator ρ of the ensemble of input states. This fidelity lemma may also prove useful in other contexts.

II. FIDELITY

Suppose ρ_1 and ρ_2 are density operators describing states of a quantum system Q. We can always imagine that these mixed states arise by a partial trace operation from pure states of an extended system QA. That is, there are states $|1\rangle$ and $|2\rangle$, called "purifications" of ρ_1 and ρ_2, for which

$$\rho_1 = \text{Tr}_A |1\rangle\langle 1|$$

$$\rho_2 = \text{Tr}_A |2\rangle\langle 2|.$$

We define (as in [10]) the fidelity $F(\rho_1, \rho_2)$ by

$$F(\rho_1, \rho_2) = \max|\langle 1|2\rangle|^2, \qquad (3)$$

where the maximum is taken over all purifications $|1\rangle$ of ρ_1 and $|2\rangle$ of ρ_2. Thus the fidelity is the largest squared inner product between purifications of two density operators. This definition provides a generalization to mixed states of the natural squared inner product measure of fidelity for pure states.

Basic properties of this notion of fidelity are described in detail in [10] and we note the following.

(i) $0 \leq F(\rho_1, \rho_2) \leq 1$ and $F(\rho_1, \rho_2) = 1$ if and only if $\rho_1 = \rho_2$.

(ii) $F(\rho_1, \rho_2) = F(\rho_2, \rho_1)$.

(iii) If one of the states ρ_1 is a projection π_1, i.e., a *pure* state, then we have the more direct expression

$$F(\pi_1, \rho_2) = \text{Tr}\pi_1\rho_2.$$

(A general expression for arbitrary mixed states is given in [10] but this is not required in the present work.)

(iv) In defining the fidelity for mixed states, it is sufficient to fix any one of the purifications $|1\rangle$ of ρ_1 and take the maximum of $|\langle 1|2\rangle|^2$ over arbitrary purifications $|2\rangle$ of ρ_2.

We can extend the definition of fidelity from normalized states to subnormalized states (in which $\text{Tr}\rho_1 < 1$) in an obvious way, by requiring that the purifications have the same normalization: $\langle 1|1\rangle = \text{Tr}\rho_1$.

We now establish a useful inequality for fidelity. Let ρ_1, ρ_2, and ρ_3 be states, and let $F_{12} = F(\rho_1, \rho_2)$, etc. We will require that $\text{Tr}\rho_3 = 1$, but ρ_1 and ρ_2 may be subnormalized. Then

$$F_{13} \leq F_{23} + 2(1 - \sqrt{F_{12}}) + 2\sqrt{2}\sqrt{F_{23}(1 - \sqrt{F_{12}})}. \qquad (4)$$

This implies that if F_{12} is close to unity and F_{23} is close to zero, then F_{13} must also be close to zero.

The proof is not difficult. We construct purifications for our states with these properties:

(i) The inner product $\langle 1|2\rangle$ is real and non-negative.

(ii) $F_{12} = \langle 1|2\rangle^2$,

(iii) $F_{13} = |\langle 1|3\rangle|^2$.

This can be done by the following procedure. We fix $|1\rangle$ and choose $|2\rangle$ and $|3\rangle$ so that $F_{12} = |\langle 1|2\rangle|^2$ and $F_{13} = |\langle 1|3\rangle|^2$. Next we adjust the phase of $|2\rangle$ to satisfy the first condition. Clearly, $F_{23} \geq |\langle 2|3\rangle|^2$.

Let $|x\rangle = |2\rangle - |1\rangle$. Then

$$\langle x|x \rangle = \langle 1|1 \rangle + \langle 2|2 \rangle - 2\langle 1|2 \rangle \leq 2(1-\sqrt{F_{12}}).$$

(Strict equality need not hold because ρ_1 and ρ_2 may be subnormalized.) Furthermore,

$$\sqrt{F_{13}} = |\langle 1|3 \rangle| = |\langle 2|3 \rangle - \langle x|3 \rangle| \leq |\langle 2|3 \rangle| + |\langle x|3 \rangle| \leq \sqrt{F_{23}}$$
$$+ \sqrt{\langle x|x \rangle} \leq \sqrt{F_{23}} + \sqrt{2(1-\sqrt{F_{12}})}.$$

Thus

$$F_{13} \leq F_{23} + 2(1-\sqrt{F_{12}}) + 2\sqrt{2}\sqrt{F_{23}(1-\sqrt{F_{12}})}$$

as we wished to prove.

We note in passing that, if we relax the condition that $\mathrm{Tr}\rho_3 = 1$, we arrive at a more general inequality for subnormalized states:

$$F_{13} \leq F_{23} + 2\mathrm{Tr}\rho_3(1-\sqrt{F_{12}}) + 2\sqrt{2\,\mathrm{Tr}\rho_3}\sqrt{F_{23}(1-\sqrt{F_{12}})}.$$

III. CHANNEL SIZE AND FIDELITY

The ''size'' of the channel system C is specified by the dimension d of the Hilbert space describing C. If C is composed of M qubits, then $d=2^M$. This means that in the process

$$\pi_i \to W_i \to w_i$$

the channel states W_i are operators on a d-dimensional Hilbert space. For convenience, we will imagine that the W_i actually act on a d-dimensional subspace of the n-dimensional Hilbert space describing the system Q. (We could always modify our decoding procedure so that the channel states were first unitarily moved into the output system Q and then subjected to a more general decoding process. The W_i states would then be the unitary images of the channel states in Q's Hilbert space.)

We are now ready to state our result. Imagine that an ensemble of pure states of Q (in which the state π_i appears with probability p_i) is described by a density operator $\rho = \Sigma_i p_i \pi_i$. Let λ_i be the eigenvalues of ρ, listed in descending order (so that $\lambda_1 \geq \cdots \geq \lambda_n$), and let $|\lambda_i\rangle$ be associated eigenvectors.

Fidelity lemma: Suppose the dimension of the Hilbert space for the channel is d, and write

$$\sum_{i=1}^{d} \lambda_i = \eta.$$

Then, for any encoding and decoding procedures, $\overline{F} < 6\eta$.

To prove this lemma, we first note that

$$\eta = \sum_{i=1}^{d} \lambda_i \geq d\lambda_{d+1},$$

so that $\lambda_{d+1} \leq \eta/d$. Now we construct a projection operator

$$\Lambda = \sum_{i=d+1}^{n} |\lambda_i\rangle\langle\lambda_i|,$$

which is the projection onto the subspace spanned by the eigenvectors corresponding to the $n-d$ smallest eigenvalues of ρ. We use Λ to project the input states π_i into (subnormalized) states $\widetilde{\pi}_i$

$$\widetilde{\pi}_i = \Lambda \pi_i \Lambda$$

$$\widetilde{\rho} = \sum_i p_i \widetilde{\pi}_i = \Lambda \rho \Lambda.$$

The largest eigenvalue of $\widetilde{\rho}$ is just λ_{d+1}.

Our plan is as follows. (For heuristic purposes and later application, we have in mind a situation with η small.) First, we will show that the original input states π_i are, on average, close to the projected states $\widetilde{\pi}_i$. Then we will show that the average of $F(\widetilde{\pi}_i, w_i)$ is small for all possible coding-decoding schemes. Using the fidelity inequality in Eq. (4) above, we will conclude that the average of $F(\pi_i, w_i)$ must therefore be small. The qualitative phrases ''close to'' and ''small'' will be quantified by the value of η.

Anticipating somewhat, we first find a lower bound for the average of the *square root* of $F(\pi_i, \widetilde{\pi}_i)$. Recall that $\pi_i = |a_i\rangle\langle a_i|$.

$$\sum_i p_i \sqrt{F(\pi_i, \widetilde{\pi}_i)} = \sum_i p_i \sqrt{\mathrm{Tr}\pi_i\Lambda\pi_i\Lambda}$$

$$= \sum_i p_i \sqrt{\langle a_i|\Lambda|a_i\rangle\langle a_i|\Lambda|a_i\rangle}$$

$$= \sum_i p_i \langle a_i|\Lambda|a_i\rangle = \mathrm{Tr}\rho\Lambda = 1 - \eta. \quad (5)$$

We wish the decoding procedure to be as general as possible. Therefore we only require that the procedure be specifiable independently of the state W_i to which it is applied, and that it be an allowable quantum dynamical evolution. The most general dynamical evolution possible in quantum mechanics is a trace-preserving completely positive map on the space of density operators [11]. Such a map can always be modeled by a unitary interaction between the system Q and an ancilla system A (initially in some standard pure state $|\phi_0\rangle$), after which A is discarded. We can therefore write

$$w_i = \mathrm{Tr}_A U(W_i \otimes |\phi_0\rangle\langle\phi_0|)U^\dagger \quad (6)$$

for some unspecified unitary U.

We can use this general form to find an upper bound for the average of $F(\widetilde{\pi}_i, w_i)$. Note that, although $\widetilde{\pi}_i$ is subnormalized, it is still an operator of rank 1, and thus we can write the fidelity as $\mathrm{Tr}\widetilde{\pi}_i w_i$. Let Γ_d be the projection onto the d-dimensional subspace occupied by the channel states W_i. Then, writing the trace over the Q Hilbert space as Tr_Q, etc.,

$$F(\widetilde{\pi}_i, w_i) = \sum_i p_i \mathrm{Tr}_Q \widetilde{\pi}_i (\mathrm{Tr}_A U(W_i \otimes |\phi_0\rangle\langle\phi_0|) U^\dagger)$$

$$= \sum_i p_i \mathrm{Tr}_{QA}(\widetilde{\pi}_i \otimes 1_A) U(W_i \otimes |\phi_0\rangle\langle\phi_0|) U^\dagger$$

$$\leq \sum_i p_i \mathrm{Tr}_{QA}(\widetilde{\pi}_i \otimes 1_A) U(\Gamma_d \otimes |\phi_0\rangle\langle\phi_0|) U^\dagger$$

$$= \mathrm{Tr}_{QA}(\widetilde{\rho} \otimes 1_A) U(\Gamma_d \otimes |\phi_0\rangle\langle\phi_0|) U^\dagger.$$

Now, every eigenvalue of $\widetilde{\rho} \otimes 1_A$ is an eigenvalue of $\widetilde{\rho}$. Furthermore, the operator $U(\Gamma_d \otimes |\phi_0\rangle\langle\phi_0|) U^\dagger$ is a projection onto a d-dimensional subspace. The trace will therefore be less than or equal to the sum of the d largest eigenvalues of $\widetilde{\rho} \otimes 1_A$, which in turn can be no larger than $d\lambda_{d+1}$

$$\sum_i p_i \mathrm{Tr}_Q \widetilde{\pi}_i w_i \leq d\lambda_{d+1} \leq d\left(\frac{\eta}{d}\right) = \eta. \tag{7}$$

We now find an upper bound for $\overline{F}$ by applying the fidelity inequality in Eq. (4) to each term in the average

$$F(\pi_i, w_i) \leq \underbrace{F(\widetilde{\pi}_i, w_i)}_{X_i} + \underbrace{2(1 - \sqrt{F(\pi_i, \widetilde{\pi}_i)})}_{Y_i}$$
$$+ \underbrace{2\sqrt{2F(\widetilde{\pi}_i, w_i)(1 - \sqrt{F(\pi_i, \widetilde{\pi}_i)})}}_{Z_i}.$$

We will bound the averages $\overline{X}$, $\overline{Y}$, and $\overline{Z}$ separately.

We have already bounded $\overline{X}$ in Eq. (7).

$$\overline{X} = \sum_i p_i X_i = \sum_i p_i \mathrm{Tr}_Q \widetilde{\pi}_i w_i \leq \eta.$$

Similarly, the bound for $\overline{Y}$ follows from Eq. (5).

$$\overline{Y} = \sum_i p_i Y_i = 2\left(1 - \sum_i p_i \sqrt{F(\pi_i, \widetilde{\pi}_i)}\right) = 2\eta.$$

To find an upper bound for $\overline{Z}$, we use these two results together with the Schwarz inequality

$$\overline{Z} = \sum_i p_i Z_i = 2\sum_i p_i \sqrt{X_i Y_i} \leq 2\sqrt{\sum_i p_i X_i}\sqrt{\sum_j p_j Y_j}$$

$$\leq 2\sqrt{2}\eta.$$

Therefore,

$$\overline{F} = \overline{X} + \overline{Y} + \overline{Z} \leq \eta + 2\eta + 2\sqrt{2}\eta < 6\eta,$$

which is what we wished to establish.

We point out once again that no assumption has been made about the encoding procedure $\pi_i \rightarrow W_i$. This may be completely arbitrary. We do not require that it be accomplished by a process that is "blind" to the input state π_i, that is, by a completely positive map. This means that we are allowing Alice to be completely cognizant of the identity of the input she is representing in the channel, even though it may be one of a nonorthogonal (and hence imperfectly distinguishable) set.

We note finally that the bound $\overline{F} < 6\eta$ is quite likely to be loose. For example, in [2] and [3], where the decoding scheme was assumed to be unitary, a bound of $\overline{F} \leq \eta$ was derived. This bound for unitary decoding is achieved by a very natural coding-decoding scheme—W_i is the renormalized projection of π_i into the subspace corresponding to ρ's largest d eigenvalues and the unitary decoding is just the identity. Denoting the projector onto this subspace by Γ_d, the fidelity may be written (taking the sum to exclude i such that π_i are orthogonal to Γ_d, which make zero contribution to average fidelity no matter how they are encoded)

$$\overline{F} = \sum_i p_i \mathrm{Tr}\left(\pi_i \frac{\Gamma_d \pi_i \Gamma_d}{\mathrm{Tr}\pi_i \Gamma_d}\right) = \sum_i p_i \frac{\langle a_i|\Gamma_d|a_i\rangle\langle a_i|\Gamma_d|a_i\rangle}{\langle a_i|\Gamma_d|a_i\rangle}$$

$$= \sum_i p_i \langle a_i|\Gamma_d|a_i\rangle = \mathrm{Tr}\rho\Gamma_d = \eta.$$

Nevertheless, the bound of 6η suffices for proving the converse of the quantum noiseless coding theorem.

IV. QUANTUM CODING

Suppose the input state π_i of Q occurs with probability p_i, so that the ensemble of inputs is described by $\rho = \Sigma_i p_i \pi_i$, as above. Further suppose that a long sequence of N such inputs, generated independently, is available. The ensemble of N sequences of input states is then described by

$$\rho^N = \overbrace{\rho \otimes \cdots \otimes \rho}^{N}.$$

For sufficiently large N, the structure of ρ^N is characterized by a *typical subspace* $\mathcal{T}_N$ [2,3].

The typical subspace may be described as follows. Fix $\epsilon, \delta > 0$. Then for sufficiently large N, there exists a subspace $\mathcal{T}_N$ spanned by eigenstates of ρ^N such that (i) if Π is the projection onto $\mathcal{T}_N$, then

$$\mathrm{Tr}\Pi\rho^N\Pi > 1 - \epsilon;$$

(ii) if $|\lambda\rangle$ is an eigenstate of ρ^N with eigenvalue λ, and $|\lambda\rangle \in \mathcal{T}_N$, then

$$2^{-N[S(\rho)+\delta]} < \lambda < 2^{-N[S(\rho)-\delta]}.$$

Now suppose that a sequence of N inputs is encoded somehow into a set of qubits, so that $S(\rho) - 2\delta$ qubits are used per input. The Hilbert space describing the channel of $N[S(\rho) - 2\delta]$ qubits will have dimension $d = 2^{N[S(\rho)-2\delta]}$. The channel states are used in some decoding procedure to produce an output state of N copies of Q.

According to our fidelity lemma, we can bound the fidelity of this process by calculating the sum of the largest d eigenvalues of ρ^N. We will denote this by Σ_d. This sum must certainly be smaller than the sum of all of the eigenvalues outside the typical subspace $\mathcal{T}_N$ plus d times the largest eigenvalue inside $\mathcal{T}_N$. That is,

$$\Sigma_d < \epsilon + d2^{-N[S(\rho)-\delta]} = \epsilon + 2^{N[S(\rho)-2\delta]} 2^{-N[S(\rho)-\delta]}$$

$$= \epsilon + 2^{-N\delta}.$$

For sufficiently large N, $\Sigma_d < 2\epsilon$. Thus, by our fidelity lemma, $\overline{F} < 12\epsilon$. Letting $\delta = \delta'/2$ and $\epsilon = \epsilon'/12$, we find that if $S(\rho) - \delta'$ qubits are available per input, then for sufficiently large N the average fidelity $\overline{F} < \epsilon'$. This establishes the converse to the quantum noiseless coding theorem for the most general sorts of coding and decoding schemes.

ACKNOWLEDGMENTS

APPENDIX

We demonstrate here, by explicit example, that decoding schemes more general than the set of unitary ones can be of some benefit in situations of nonoptimal coding.

Consider three signal states $|a_0\rangle$, $|a_1\rangle$, and $|a_2\rangle$ which are all real positive linear combinations of three fixed orthonormal vectors, so that we may picture them as vectors in the positive octant of $\mathbb{R}^3$. The states form three edges of a regular tetrahedron with the origin as their common vertex, and thus are all $60°$ apart. The states $|a_0\rangle$ and $|a_1\rangle$, in particular, are assumed to be in the positive quadrant of the x-y plane, each vector having an angle of $15°$ between itself and the nearest axis. The prior probabilities for the signal states are 0.49, 0.49, and 0.02, respectively. The encoding scheme associates the orthogonal projectors W_0 and W_1 onto the x and y axes, respectively, with the states $|a_0\rangle$ and $|a_1\rangle$. It associates the density matrix

$$W_2 = \tfrac{1}{2}|a_0\rangle\langle a_0| + \tfrac{1}{2}|a_1\rangle\langle a_1|,$$

corresponding to an equal mixture of $|a_0\rangle$ and $|a_1\rangle$, with the state $|a_2\rangle$. Note that the set of encoded states has a two-dimensional support, i.e., a support smaller than that containing the signal states.

Because the signal state $|a_2\rangle$ has such a small prior probability, the symmetry of this encoding should make it clear that the best unitary decoding scheme will be only slightly different from *not* decoding at all. (Actually, detailed calculation demonstrates that the optimal unitary decoding is to rotate the encoded states by $0.791°$ toward $|a_2\rangle$, but this only changes the average fidelity in the fourth significant figure.) Making this approximation, the average fidelity for this decoding scheme is

$$\overline{F} = 2 \times 0.49 \times \cos^2 15° + 0.02 \times \cos^2 60° = 0.919.$$

However there exists a simple nonunitary decoding scheme that achieves a better fidelity than this. Since some of the signals are encoded in orthogonal alternatives, it is plausible that a decoding device can use a measurement to gather information about the signal and use that information to produce decoded states that are closer, on average, to the originals. In particular, the decoding device can do the following. It first measures the observable corresponding to the x-y axis. If the outcome is x, it outputs the state $w_0 = \pi_0$; if the outcome is y, it outputs the state $w_1 = \pi_1$. Thus in the cases that Q was actually prepared in $|a_0\rangle$ or $|a_1\rangle$, the transmissions will have perfect fidelity. In the case that $|a_2\rangle$ was the actual signal state, the fidelity of the transmission will still be $\cos^2 60° = 0.25$. Therefore the average fidelity for this nonunitary decoding scheme is $\overline{F} = 0.985$, and this certainly beats the unitary scheme.

This simple example demonstrates that in some cases involving *particular nonoptimal* encoding schemes, it is possible for nonunitary decoding to increase the fidelity of a quantum channel. Nevertheless the converse of the quantum noiseless theorem implies that nonunitary decodings provide no asymptotic advantage over unitary decoding schemes in the problem of minimizing of channel resources over all possible coding-decoding schemes.

[1] C. H. Bennett, Phys. Today **48**, 24 (1995).

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

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

[4] C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), pp. 175–179.

[5] C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992).

[6] D. P. DiVincenzo, Science **279**, 255 (1995).

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

[8] A. Peres, Phys. Lett. A **128**, 19 (1988).

[9] C. A. Fuchs and A. Peres, Phys. Rev. A **53**, 2038 (1996).

[10] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).

[11] K. Hellwig and K. Kraus, Commun. Math. Phys. **16**, 142 (1970); K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983).