ARTICLES

Capacity of the noisy quantum channel

Seth Lloyd*
D'Arbeloff Laboratory for Information Systems and Technology, Department of Mechanical Engineering, Massachusetts Institute of Technology, MIT 3-160, Cambridge, Massachusetts 02139
(Received 22 July 1996)

An upper limit is given to the amount of quantum information that can be transmitted reliably down a noisy, decoherent quantum channel using the high-probability states of quantum sources. A class of quantum error-correcting codes is presented that allows the information transmitted to attain this limit. The result is a quantum analog of Shannon's bound and code for the noisy classical channel [C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Chicago, 1948)].
[S1050-2947(97)05702-8]

PACS number(s): 03.65.Bz

I. INTRODUCTION

The ''quantum'' in quantum mechanics means ''how much''—in quantum mechanics, classically continuous variables such as energy, angular momentum and charge come in discrete units called quanta. This discrete character of quantum-mechanical systems such as photons, atoms, and spins allows them to register ordinary digital information. A left-circularly polarized photon can encode a 0, for example, while a right-circularly polarized photon can encode a 1. Quantum systems can also register information in ways that classical digital systems cannot: a transversely polarized photon is in a quantum superposition of left and right polarization and in some sense encodes both 0 and 1 at the same time. Even more surprising from the classical perspective are so-called entangled states, in which two or more quantum systems are in superpositions of correlated states, so that two photons can encode, for example, 00 and 11 at once. Such entangled states behave in ways that apparently violate classical intuitions about locality and causality (without, of course, actually violating physical laws).

Information stored on quantum systems that can exist in superpositions and entangled states is called quantum information. The unit of quantum information is the quantum bit, or qubit (pronounced ''*Q* bit'') [1], the amount of quantum information that can be registered on a single two-state variable such as a photon's polarization or a neutron's spin. This paper puts fundamental limits on the amount of quantum information that can be transmitted reliably along a noisy communication channel such as an optical fiber. Theorems are presented that limit the rate at which arbitrary superpositions of qubits can be sent down a channel with given noise characteristics, and encoding schemes are presented that attain that limit.

It is important to compare the results presented here—the use of a quantum channel to transmit quantum information—with schemes that use quantum channels to transmit classical information, as in Caves and Drummond's comprehensive review of quantum limits on bosonic communication rates [2]. The limit to the rate at which arbitrary sequences of ordinary classical bits, suitably encoded as quantum states, can be transmitted down a quantum channel such as an optical fiber is given by Holevo's theorem. In contrast, the results presented here limit the rate at which arbitrary *superpositions* of sequences of quantum bits can be sent reliably down a noisy, decoherent quantum channel. As such, the theorems presented in this paper are complementary to the results of Schumacher [1] and Josza and Schumacher [3] on the noiseless quantum channel. Any channel that can transmit quantum information can be used to transmit classical information as well. It is possible, however, for a channel to be able to transmit classical information without being able to transmit quantum information: examples of such completely decoherent channels will be discussed below.

The difference between quantum and classical information does not arise from a fundamental physical distinction between the systems that register, process, and transmit that information. As just noted, quantum channels can be used to transmit classical information. And after all, classical information-registering systems such as capacitors and neurons are at bottom quantum mechanical. The difference arises from the conditions under which such systems operate. When properly isolated from their environment, photons and atoms can exist in superpositions and entangled states for long periods of time, with experimentally measurably results. Capacitors and neurons, in contrast, interact strongly with a thermal environment, which prevents them from exhibiting coherent quantum effects. As a result, quantum information can be used to perform tasks that classical information cannot.

A full theory of quantum information and its properties does not yet exist. However, the ability to transmit and process quantum information reliably provides the solution to

*Electronic address: slloyd@mit.edu

 © 1997 The American Physical Society

problems for which no classical solution is known: if entangled quantum bits can be transmitted and received, quantum cryptographic techniques can be used to create probably secure shared keywords for unbreakable codes [4], while the ability to process quantum information allows quantum computers efficiently to factorize large numbers and to simulate local quantum systems [5].

For quantum information to prove useful, it must be transmitted and processed *reliably*. Quantum superpositions and entangled states tend to be easily disrupted by noise and by interactions with their environment, a process called decoherence [6,7]. Until recently, decoherence and noise seemed insurmountable obstacles to reliable quantum information transmission and processing. However, in 1995, Shor exhibited a quantum error-correcting routine [8]. Since then, several such routines have been proposed [9–13]. These routines have the feature, common to many classical error-correcting codes as well, that the rate of transmission of quantum information goes to zero as the reliability of transmission goes to 1. This paper shows that arbitrarily complicated quantum states can in principle be encoded, subjected to high levels of noise and decoherence, then decoded to give a state arbitrarily close to the original state, all with a finite rate of transmission of quantum information. The paper states and gives the proof of theorems that put an upper bound to the capacity of noisy, decoherent quantum channels to transmit quantum information reliably, and exhibits a class of quantum codes that attain that bound. As with Shannon's theory of the noisy classical channel [14], the quantum theorems bound the amount of information that can be sent by using as codewords the ''typical'' or ''high-probability'' states emitted by a source. In particular, theorems bound the amount of quantum information that can be sent using code words of finite length. Unlike the classical case, however, in the quantum case it may still be possible to circumvent these bounds by using an ''atypical'' or measure-zero set of infinite-length codewords.

II. QUANTUM SOURCES

A quantum channel has a *source* that emits systems in quantum states (the signal) to the channel and a receiver that receives the noisy, decohered signal emitted by the channel. For example, the source could be a highly attenuated laser that emits individual monochromatic photons, the channel could be an optical fiber, and the receiver could be a photocell. Or the source could be a set of ions in an ion-trap quantum computer [15] that have been prepared by a sequence of laser pulses in an entangled state, the channel could be the ion trap in which the ions evolve over time, and the receiver could be a microscope to read out the states of the ions via laser-induced fluorescence. This second example indicates that a quantum channel can transmit quantum information from one time to another as well as from one place to another. As Shannon emphasized, a computer memory is a communications channel.

A more complete picture of a quantum channel is shown in Fig. 1: the input signal is some unknown quantum state; the input is fed into an encoder that transforms it into a redundant form, the encoded signal is sent down the channel, subjected to noise and decoherence; the noisy, decohered signal is then fed into a decoder that attempts to restore the original signal. Quantum encoding and decoding requires the ability to manipulate quantum states in a systematic fashion, for example, by using Turchette *et al.*'s [16] photonic quantum logic gates or Monroe *et al.*'s realization [15] of the ion-trap quantum computer proposed by Cirac and Zoller [17]. From a practical point of view, such decoding and encoding may prove the most difficult part of reliable quantum information transmission and processing. This paper will simply exhibit coding and decoding schemes that attain the channel capacity: it will not address how such schemes can be carried out in practice.

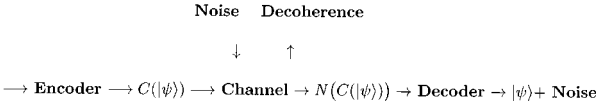

FIG. 1. Diagram of the noisy, decoherent quantum channel. To send an arbitrary quantum state $|\psi\rangle$ down the channel, first encode it in a redundant from $C(|\psi\rangle)$. The encoded state is sent down the channel, where it is subjected to noise and decoherence. The arrows indicate that noise is added to the signal, while decoherence arises from the environment getting information about the signal. The noisy, decoherent signal $N(C(|\psi\rangle))$ is then fed through a decoder that recreates the original state together with extra random information that depends on what errors occurred.

The quantum analog of Shannon's noisy coding theorem [14] is outlined in Fig. 2. In order to demonstrate this analog, it is helpful to set up a quantum formalism that corresponds closely to the classical picture of a noisy channel. Quantum systems and quantum signals are described by states $|\psi\rangle$ in a Hilbert space $\mathcal{H}$ or, more generally, by density matrices $\rho \in \mathcal{H}^* \otimes \mathcal{H}$. A quantum ensemble $\mathcal{E}=\{(|\psi_i\rangle, p_i)\}$ is a set of quantum states $|\psi_i\rangle$ belonging to the same Hilbert space $\mathcal{H}$, together with their probabilities p_i. The expectation value of a measurement on the ensemble corresponding to a Hermitian operator M is $\langle M\rangle_{\mathcal{E}}=\Sigma_i p_i \langle\psi_i|M|\psi_i\rangle = \mathrm{tr} M\rho_{\mathcal{E}}$, where $\rho_{\mathcal{E}}=\Sigma_i p_i |\psi_i\rangle\langle\psi_i|$ is the density matrix corresponding to the ensemble. The states $|\psi_i\rangle$ need not be orthonormal. The ensemble will be said to be normalized when $\Sigma_i p_i \langle\psi_i|\psi_i\rangle = \mathrm{tr}\rho_{\mathcal{E}} = 1$. That is, a quantum ensemble is just the quantum analog of a classical ensemble, where care has been taken to take into account the inherently statistical nature of quantum mechanics.

Two ensembles that have the same density matrix are statistically indistinguishable: no set of measurements can distinguish whether a sequence of states is drawn from one ensemble rather than the other. An example of statistically indistinguishable ensembles is

$$\mathcal{E}_1=\{(|\uparrow\rangle,\tfrac{1}{2}),(|\downarrow\rangle,\tfrac{1}{2})\}$$

and

$$\mathcal{E}_2=\{(|\uparrow\rangle,\tfrac{1}{3}),(\tfrac{1}{2}|\uparrow\rangle+\sqrt{3}/2|\downarrow\rangle,\tfrac{1}{3}),(\tfrac{1}{2}|\uparrow\rangle-\sqrt{3}/2|\downarrow\rangle,\tfrac{1}{3})\},$$

both with density matrices $\rho=\frac{1}{2}|\uparrow\rangle\langle\uparrow|+\frac{1}{2}|\downarrow\rangle\langle\downarrow|$. Note that an ensemble over a finite-dimensional Hilbert space can contain an infinite number of states, e.g., $\mathcal{E}=\{(e^{i\phi}|\uparrow\rangle, p(\phi)=1/2\pi)\}$, in which case each state is

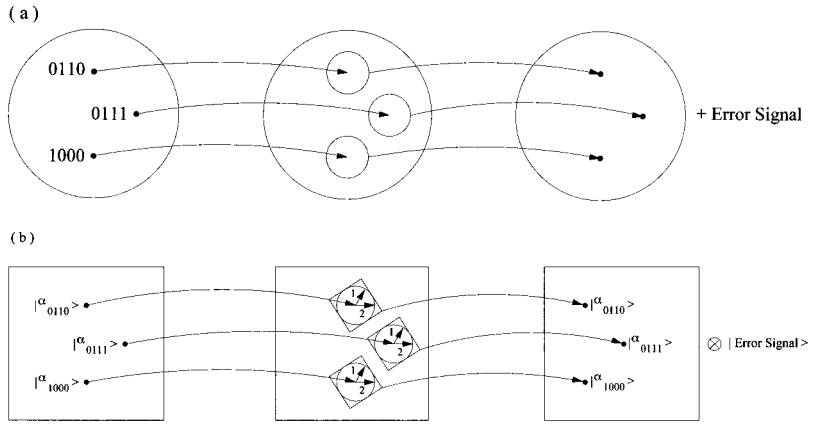

FIG. 2. Comparison of classical and quantum channels. (a) The noisy classical channel. The message is encoded as one of a set of codewords (0110, 0111, and 1000 are shown in the diagram) which are sparsely distributed in a large space of possible inputs. The codeword is sent through the channel: since the channel adds noise, each codeword input can result in many possible outputs, represented by the circles in the center of the diagram. These circles do not overlap, signifying that the code has been constructed so that different inputs almost never give the same output, in spite of the noise. As a result, the noisy outputs can be uniquely decoded and mapped back to the original input. The total number of messages that can be reliably encoded and decoded is given by the ratio between the volume of the output space, and the volume of the set of possible outputs corresponding to a typical codeword input: the channel capacity is the logarithm of this ratio. (b) The noisy quantum channel. The situation is much the same as the classical channel, with the nontrivial complication that quantum states are vectors, that quantum ''messages'' can be arbitrary superpositions of states, and that the coding and decoding procedure must accordingly preserve the phases and amplitudes of quantum states. The coding proceeds as follows: each basis vector of the quantum message space is encoded as a codeword ($|\alpha_{0110}\rangle$,$|\alpha_{0111}\rangle$, and $|\alpha_{1000}\rangle$ in the diagram) and sent down the channel. The channel adds noise and decoheres the encoded message, so that each input codeword can result in many possible output states, once again represented by circles in the diagram. Here the vector character of quantum mechanics comes in. As shown in the text, the outputs for a given input lie with high probability in some minimal subspace of the output Hilbert space; these subspaces are represented by the squares that almost contain the circles. As long as the minimal subspaces have vanishingly small overlap, the noisy outputs can be mapped back to the original codewords reliably. The maximum number of codewords that can be decoded reliably is given by the ratio of the dimension of the output Hilbert space to the dimension of the minimal subspace corresponding to a typical codeword, and the quantum channel capacity is the logarithm of this ratio. In the classical case, this would be the end of the story: in the quantum case it is not. Quantum coding must preserve phases and amplitudes. In the output subspace, arrows have been drawn to indicate that when one codeword is perturbed by the channel in direction 1, the other codewords are also perturbed in direction 1; when one codeword is perturbed in direction 2, the other codewords are perturbed in direction 2, etc. For the decoding procedure to recreate superpositions reliably, the amplitudes of the perturbations 1 and 2 and the angles between them must be the *same* in the subspaces corresponding to the different codewords. This is a necessary and sufficient condition for the quantum decoding procedure to recreate superpositions and entangled states reliably. The random coding of Theorem 3 obeys this condition and provides reliable transmission of quantum information.

paired with a continuous probability density, $p(\phi)$, and $\rho=\int_0^{2\pi}(1/2\pi)e^{i\phi}|\uparrow\rangle\langle\uparrow|e^{-i\phi}d\phi=|\uparrow\rangle\langle\uparrow|$: an ensemble with an infinite number of states can have the same density matrix as an ensemble with a single state. Because of the inherently statistical nature of quantum mechanics, different quantum ensembles can be statistically indistinguishable, while two classical ensembles are statistically indistinguishable if and only if they are identical. Further properties of quantum ensembles can be found in Appendix A.

A particularly interesting type of continuous quantum ensemble is the uniform ensemble over a Hilbert space $\mathcal{H}$, $\mathcal{E}_{\mathcal{H}}=\{(|\phi\rangle\in\mathcal{H},p_\phi=1/\mathrm{vol}\mathcal{H})\}$, where vol$\mathcal{H}$ is the volume of the unit sphere in $\mathcal{H}$. This ensemble contains every possible state and superposition of states is $\mathcal{H}$, all with equal probabilities. The corresponding density matrix is $\rho_{\mathcal{H}}=(1/d)\Sigma_{i=1}^{d}|\phi_i\rangle\langle\phi_i|$, where d is the dimension of $\mathcal{H}$ and $\{|\phi_i\rangle\}$ is an orthonormal basis for $\mathcal{H}$. If we wish to transmit arbitrary superpositions of states down quantum channels, the sources of interest are of the form $\mathcal{E}_{\mathcal{H}}$ for some $\mathcal{H}$.

Like Shannon, we will restrict our attention to a mathematically tractable subset of all possible sources [17]. Shannon concentrated on stationary, ergodic sources for the classical channel. A stationary source is one for which the probabilities for emitting states does not change over time; an ergodic source is one in which each subsequence of states appears in longer sequences with a frequency equal to its probability. Stationary, ergodic sources have a finite, though potentially arbitrarily large correlation time. We will concentrate on quantum sources with similar properties. First, we investigate ''memoryless'' ensembles, whose density matrix $\rho\otimes\rho\otimes\cdots\otimes\rho$ is the tensor product of N times its density matrix over a single time step.

There are many different quantum ensembles with density matrix $\rho\otimes\cdots\otimes\rho$. But as noted by Schumacher [1], and Josza and Schumacher [3], there is one ensemble in particular that effectively contains all such ensembles. Let $\rho\Sigma_i p_i|\phi_i\rangle\langle\phi_i|$, where the ϕ_i are orthonormal. Consider the subspace $\tilde{\mathcal{H}}^N$ spanned by the ''high-probability'' product states $|\phi_{i_1}\rangle\cdots|\phi_{i_N}\rangle$, where each $|\phi_i\rangle$ occurs in the product approximately p_iN times. These states are the analog of high-probability sequences of symbols for a classical source. The following theorem then follows as an immediate corollary to the noiseless quantum channel source theorem of Schumacher [1] and Josza and Schumacher [3].

Theorem 1. (Quantum source theorem.) Let $|\psi\rangle$ be selected from *any* ensemble with density matrix $\rho\otimes\cdots\otimes\rho$. Then, as $N\rightarrow\infty$, $|\psi\rangle$ is to be found in the high-probability subspace $\widetilde{\mathcal{H}}_N$ with probability 1. $\widetilde{\mathcal{H}}_N$ is a minimal subspace with this property, in the sense that any other such subspace contains $\widetilde{\mathcal{H}}_N$ asymptotically as $N\rightarrow\infty$.

That is, as $N\rightarrow\infty$, the ensemble $\mathcal{E}_{\widetilde{\mathcal{H}}^N}$ contains with probability one of the members of any ensemble with density matrix $\rho\otimes\cdots\otimes\rho$. A more precise statement of Theorem 1 is that as $N\rightarrow\infty$, $\Sigma_{|\psi\rangle}p_{|\psi\rangle}\langle\psi|P_{\widetilde{\mathcal{H}}_N}|\psi\rangle\rightarrow 1$, where $P_{\widetilde{\mathcal{H}}^N}$ is the projection operator onto $\widetilde{\mathcal{H}}^N$. The proof of Theorem 1 is found in Appendix B. By Shannon's source theorem, the dimension of $\mathcal{E}_{\widetilde{\mathcal{H}}^N}$ is approximately e^{NS}, where $S=-\mathrm{tr}\rho\ \ln\rho$.

We would like to find similar minimal subspaces for sources with memory. In fact, we can adequately approximate arbitrary stationary, ergodic sources with finite correlation length m by looking at ''block'' sources of the form $\rho^{\ell}\otimes\cdots\otimes\rho^{\ell}$, where ρ^{ℓ} is the reduced density matrix over ℓ qubits of the stationary, ergodic source, and $\ell\gg m$. As shown in Appendix B, the minimal subspace for such a block source asymptotically contains all but a set of measure zero of the states generated by the stationary, ergodic source. In addition, the entropy of the block source asymptotically approaches the entropy of the stationary, ergodic source as ℓ becomes much larger than m. As a result, by taking ℓ sufficiently large, the minimal subspace for the block source can be made arbitrarily close to a minimal subspace for the stationary, ergodic source.

As with Shannon's theorems for classical sources, which simplify the analysis of the classical noisy channel by focusing on high-probability inputs, and as with the use of high-probability subspaces in the noiseless quantum channel theorem in Refs. [1] and [3], the quantum source theorem simplifies the analysis of the noisy quantum channel by focusing on a particular subspace of inputs. A coding scheme that works for any ensemble with density matrix $\rho^{\ell}\otimes\cdots\otimes\rho^{\ell}$ works for the states in the high-probability subspace. Conversely, a coding scheme that works for the high-probability subspace works for any of the ensembles that it contains. Accordingly, from this point on, quantum sources will be taken to be ensembles over high-probability subspaces of block sources unless otherwise stated. It should be kept in mind, however, that the resulting theorems do not hold for coding schemes in which the codewords are taken from the measure zero set of states that do not fall in the high-probability subspace. This is not a strong restriction: for example, the theorems apply to all coding schemes that use codewords of finite length.

III. QUANTUM CHANNEL

A quantum communications channel takes quantum information as input and produces quantum information as output. An optical fiber is an example of a quantum channel: a photon in some quantum state goes in, suffers noise and distortion in passing through the fiber, and, if it is not absorbed and does not tunnel out, emerges in a transformed quantum state. In the normal formulation of quantum mechanics, the ingoing system that carries quantum information is described by a density matrix ρ_{in} and the outgoing system is described by a density matrix $\rho_{\mathrm{out}}=\mathcal{S}(\rho_{\mathrm{in}})$, where $\mathcal{S}$ is a trace-preserving linear operator called a superscattering operator. $\mathcal{S}$ plays the same role for the quantum channel as the Markoff matrix that gives the probability for outputs in terms of inputs plays for the classical channel. For simplicity, the quantum channel will be assumed to be stationary and memoryless, so that it has the same effect on each block of quantum bits that goes through.

An equivalent method of formulating the channel's dynamics specify its effect on each of an orthonormal basis $\{|\phi_i\rangle\}$ of input states: the output of the channel for input $|\phi_i\rangle$ is then given by the ensemble $\mathcal{E}_{|\phi_i\rangle}=\{(|\psi_{j(i)}\rangle,p_{j(i)})\}$ of output states into which $|\phi_i\rangle$ can evolve, together with the probabilities $p_{j(i)}$ that $|\phi_i\rangle$ evolves into the state $|\psi_{j(i)}\rangle$. The density matrix and ensemble pictures of the effect of the channel are related as follows: $\mathcal{S}(|\phi_i\rangle\langle\phi_{i'}|)=\Sigma_{j(i)}\sqrt{p_{j(i)}p_{j(i')}}|\psi_{j(i)}\rangle\langle\psi_{j(i')}|$, which for $i=i'$ gives $\mathcal{S}(|\phi_i\rangle\langle\phi_i|)=\Sigma_{j(i)}p_{j(i)}|\psi_{j(i)}\rangle\langle\psi_{j(i)}|$. The ensemble picture of the time evolution of open quantum systems is equivalent to the operator sum decomposition of the superscattering operator discussed in Ref. [18]. A further discussion of the properties of the ensemble picture can be found in Appendix A. For example, if the channel is noiseless and distortion free, then $\mathcal{S}$ is the identity operator and $\mathcal{E}_{|\phi_i\rangle}=\{(|\phi_i\rangle,1)\}$. This channel transmits both classical and quantum information perfectly. Another example is the completely decohering channel, which can be thought of as the channel that destroys off-diagonal terms in the density matrix: $\mathcal{S}(\Sigma_{ij}\alpha_{ij}|\phi_i\rangle\langle\phi_j|)=\Sigma_i\alpha_{ii}|\phi_i\rangle\langle\phi_i|$ or, equivalently, and perhaps more intuitively, as the channel that randomizes the phases of input states: $|\phi_i\rangle\rightarrow\mathcal{E}_{|\phi_i\rangle}=\{(e^{i\lambda}|\phi_i\rangle,p(\lambda)=1/2\pi)\}$. The completely decohering channel highlights the difference between the use of quantum channels to carry classical information and their use in carrying quantum information: it transmits classical information perfectly, but transmits no quantum information at all: no superpositions or entanglements survive transmission.

Most quantum channels are neither noiseless nor completely decohering. The next theorem quantifies just how much quantum information can be sent down a noisy, decohering channel. As above, we restrict our attention to block sources with density matrix $\rho_{\mathrm{in}}^{l}=\Sigma_i p_i|\phi_i\rangle^{l}\langle\phi_i|$, where $\{|\phi_i\rangle^{l}\}$ is a basis for blocks of ℓ qubits and i ranges from 1 to 2^{l}. The inputs to the channel are then described by a density matrix $\rho_{\mathrm{in}}^{lN}=\rho_{\mathrm{in}}^{l}\otimes\cdots\otimes\rho_{\mathrm{in}}^{l}$ and the output is described by a density matrix $\rho_{\mathrm{out}}^{lN}=\rho_{\mathrm{out}}^{l}\otimes\cdots\otimes\rho_{\mathrm{out}}^{l}$, where $\rho_{\mathrm{out}}^{l}=\mathcal{S}^{l}(\rho_{\mathrm{in}}^{l})=\Sigma_{i,j(i)}p_ip_{j(i)}|\psi_{j(i)}\rangle^{l}\langle\psi_{j(i)}|$ and $\mathcal{S}^{l}$ gives the effect of the channel on l qubits.

As $N\rightarrow\infty$, input states come from the subspace $\widetilde{\mathcal{H}}_{\mathrm{in}}^{lN}$ with probability 1, and output states lie in the subspace $\widetilde{\mathcal{H}}_{\mathrm{out}}^{lN}$ spanned by high-probability sequences of outputs, $|\psi_{j_1(i_1)}\rangle^{l}\ldots|\psi_{j_N(i_N)}\rangle^{l}$, where each $|\psi_{j(i)}\rangle^{l}$ appears in the sequence approximately $p_ip_{j(i)}N$ times. The dimension of $\widetilde{\mathcal{H}}_{\mathrm{out}}^{lN}$ is $\approx 2^{-N\ \mathrm{tr}\rho_{\mathrm{out}}^{l}\log_2\rho_{\mathrm{out}}^{l}}$. To gauge the quantity of quantum information sent down the channel, look at the effect of the channel on a typical input state $|\alpha_N^{l}\rangle=\Sigma_{i_1,\ldots,i_N}\alpha_{i_1,\ldots,i_N}|\phi_{i_1}\rangle^{l}\cdots|\phi_{i_N}\rangle^{l}\in\widetilde{\mathcal{H}}_{\mathrm{in}}^{lN}$, where the sum

is over high-probability intput sequences in which $|\phi_i\rangle^l$ appears approximately $p_i N$ times. We have the following.

Theorem 2. (Quantum channel theorem.) As $N\rightarrow\infty$, when $|\alpha_N^l\rangle$ is input to the channel, the output lies with probability 1 in a minimal subspace $\widetilde{\mathcal{H}}_\alpha^{lN}$ whose average dimension over α_N is the minimum of $e^{NS_{\text{out}}}$, $e^{NS_{\bar{\alpha}}}$, where $S_{\bar{\alpha}}=-\text{tr}\rho_{\bar{\alpha}}^l\ln\rho_{\bar{\alpha}}^l$ and $\rho_{\bar{\alpha}}^l=\Sigma_{i,i'}\sqrt{p_i p_i'}\mathcal{S}^l(|\phi_i\rangle)^l\langle\phi_{i'}|)\otimes|\phi_i\rangle^l\langle\phi_{i'}|$.

The proof of Theorem 2 is found in Appendix C. The proof is somewhat involved, but the form of $\rho_{\bar{\alpha}}$ can be understood simply. One of the primary uses of a quantum channel is the distribution of entangled quantum states for the purpose of quantum cryptography or teleportation. Take a two-variable entangled state of the form $\Sigma_i\sqrt{p_i}|\phi_i\rangle^l|\phi_i\rangle^l$, where $|\phi_i\rangle^l$ is a state of a block of l qubits. Like the state $(1/\sqrt{2})(|0\rangle|0\rangle+|1\rangle|1\rangle)$ described in Sec. I, this state is a maximally entangled state that registers all the states $|\phi_i\rangle^l|\phi_i\rangle^l$ at once; the factors of $\sqrt{p_i}$ ensure that each of the two quantum variables taken on its own is described by a density matrix ρ_{in}^l. Now send the first variable down the channel. The result is a partially entangled state for the two variables described by density matrix $\rho_{\bar{\alpha}}^l$. That is, $S_{\bar{\alpha}}^l$ is the entropy increase when one of two fully entangled variables is sent down the channel. A thorough treatment of the effect of noisy channels on entangled states can be found in Refs. [18] and [19]. The effect of the channel on an lN-qubit state $|\alpha_N^l\rangle$ can be understood as follows: almost all input states $|\alpha_N^l\rangle$ are fully entangled, with the density matrix ρ_{in}^l describing each block of l qubits on its own [20]. Sending n of the variables through the channel then increases the entropy by $nS_{\bar{\alpha}}$, which is in turn the logarithm of the dimension of the minimal subspace that can encompass the channel's possible outputs. If $S_{\bar{\alpha}}>S_{\text{out}}$, then sending all the variables through completely randomizes the output as $N\rightarrow\infty$ and no coherent quantum information survives the transmission through the channel.

Theorem 2 suggests that the amount of quantum information transmitted down the channel from a stationary, ergodic source with density matrix ρ_{in}^l be defined as $(1/l)I_Q(\rho_{\text{in}}^l)=(1/l)(-\text{tr}\rho_{\text{out}}^l\log_2\rho_{\text{out}}^l+\text{tr}\rho_{\bar{\alpha}}^l\log_2\rho_{\bar{\alpha}}^l=(1/l)(S_{\text{out}}-S_{\bar{\alpha}})$ if $S_{\text{out}}>S_{\bar{\alpha}}$, $=0$ otherwise. (Nielsen and Schumacher called this quantity I_Q, coherent information [19].) This definition of quantum information transmitted is the quantum analog of mutual information between channel inputs and outputs: when pure states are sent down the channel, I_Q tells how much information one obtains about which pure state $\in\widetilde{\mathcal{H}}_{\text{in}}^{lN}$ went in by looking at the noisy mixed state $\in\widetilde{\mathcal{H}}_{\text{out}}^{lN}$ that comes out.

Figure 2 provided a schematic comparison of noisy classical and quantum channels. The full justification of I_Q as the quantum information transmitted down a quantum channel will be presented in Sec. IV, in which quantum coding schemes will be presented that allow the reliable transmission of quantum information at a rate governed by I_Q, and in which it will be noted that no coding schemes exist for block sources that can surpass this rate. For the moment, consider three examples of quantum channels, each with source described by $\rho_{\text{in}}=\frac{1}{2}(|0\rangle\langle 0|+|1\rangle\langle 1|)$. (i) In the noiseless quantum channel, $-\text{tr}\rho_{\text{out}}\log_2\rho_{\text{out}}=1, -\text{tr}\rho_{\bar{\alpha}}\log_2\rho_{\bar{\alpha}}=0$, and $I_Q=1$ qubit, reflecting the fact that each qubit is received as sent. (ii) In the completely decohering or dephasing channel, $-\text{tr}\rho_{\text{out}}\log_2\rho_{\text{out}}=1$, $\rho_{\bar{\alpha}}=\frac{1}{2}(|0\rangle\langle 0|\otimes|0\rangle\langle 0|+|1\rangle\langle 1|\otimes|1\rangle\langle 1|)$, $-\text{tr}\rho_{\bar{\alpha}}\log_2\rho_{\bar{\alpha}}=1$, and $I_Q=0$ qubits, so that no quantum information is sent. (iii) Consider a partly dephasing channel in which $|0\rangle\langle 0|\rightarrow|0\rangle\langle 0|$, $|1\rangle\langle 1|\rightarrow|1\rangle\langle 1|$ and $|0\rangle\langle 1|\rightarrow(1-\epsilon)|0\rangle\langle 1|$, $|1\rangle\langle 0|\rightarrow(1-\epsilon)|1\rangle\langle 0|$. Here

$$\rho_{\bar{\alpha}}=\tfrac{1}{2}(|0\rangle\langle 0|\otimes|0\rangle\langle 0|+|1\rangle\langle 1|\otimes|1\rangle\langle 1|)+(1-\epsilon)(|1\rangle\langle 0|\otimes|1\rangle\langle 0|+|0\rangle\langle 1|\otimes|0\rangle\langle 1|),$$

and $-\text{tr}\rho_{\bar{\alpha}}\log_2\rho_{\bar{\alpha}}=-(1-\epsilon/2)\ln(1-\epsilon/2)-(\epsilon/2)\log_2(\epsilon/2)$, giving an I_Q that ranges continuously from 1 for $\epsilon=0$ (no decoherence) to 0 for $\epsilon=1$ (complete decoherence).

IV. OPTIMAL CODES FOR THE NOISY QUANTUM CHANNEL

Define the capacity of a quantum channel to carry quantum information encoded by sources with block length l to be $C_Q^l=\max_{\rho_{\text{in}}}(1/l)I_Q(\rho_{\text{in}})$. C_Q is the maximum over all sources with block length l of the coherent information I_Q per bit transmitted down the channel. We then have the following.

Theorem 3. (Noisy quantum channel coding theorem.) Consider a quantum channel with capacity C_Q^l. The output of a stationary, ergodic source with density matrix ρ^l over l bits can be encoded as high-probability states of a source with block length l, sent down the channel, and decoded with reliability $\rightarrow 1$ as $N\rightarrow\infty$ if and only if $(1/l)(-\text{tr}\,\rho^l\log_2\rho^l)\leqslant C_Q^l$.

Like Shannon's noisy coding theorem, Theorem 3 comes with the caveat that it applies to high-probability sources [21]. The proof of Theorem 3 is found in Appendix D. The idea behind the proof, as well as the theorem's meaning and implications can be understood as follows. The noisy, decohering quantum channel has two effects on the quantum information that it transmits. First, like the classical channel, it adds noise to the signal, flipping qubits and adding random information. Second, it decoheres the signal by randomizing phases and acquiring information about the quantum information transmitted. Decoherence is an effect with no classical analog: classical signals do not have phases, and acquiring information about a classical signal is harmless as long as the signal is not altered in the process. In quantum mechanics, however, acquiring information about the signal means effectively making a measurement on it, and quantum measurement unavoidably alters most quantum systems.

The problem of decoherence implies that signal must be encoded in such a way that any information the channel obtains about the encoded state reveals nothing about which state of the source was sent. Otherwise, the channel can effectively ''measure'' the output of the source, irretrievably disturbing it in the process. As noted by Shor, this may be accomplished by encoding the signal as an entangled state [8]. In fact, each encoded signal must have the same density matrix ρ_{in} as each other encoded signal for each qubit sent down the channel: otherwise the channel can distinguish between different signals and decohere them. If the signals are

encoded as entangled states in this fashion, the channel can decohere the codeword, but it cannot decohere the original signal.

Suppose someone hands you a quantum system in some unknown state generated by a stationary, ergodic source with density matrix ρ^l and asks you to transmit it reliably down a noisy, decoherent quantum channel. What do you do? (If someone hands you a system in a known quantum state, no quantum channel is necessary: you can just use a classical channel to transmit instructions for recreating the state using a quantum computer.) The following encoding attains the channel capacity. First, identify a source for the channel that attains the channel capacity, so that $I_Q(\rho^l_{\rm in})=lC^l_Q$. Next, encode the state to be transmitted by applying a transformation that maps an orthonormal basis for the input high-probability subspace to a *randomly chosen* set of orthogonal states taken from the high-probability subspace of the source that attains the channel capacity. Randomly chosen states have on average the maximum possible ''distance'' between them, and are the hardest for the channel to mix up [22]. In addition, such random states have the desired property that they are fully entangled, and each block of qubits in the encoded signal has density matrix $\rho_{\rm in}$ [20]. Now send the encoded signal down the channel. Because the states are fully entangled, the channel cannot obtain any information about the original preencoded state: all the channel can do to disrupt the encoded state is add entropy $S_{\rm out}-lC^l_Q$ per block transmitted. That is, the encoding protects the original state from decoherence, and as long as $(1/l)(-{\rm tr}\rho^l\log_2\rho^l)\leqslant C^l_Q$ there is enough redundancy in the encoded state to recreate the original state, just as in the classical case. This method works equally well if the initial state is pure, mixed, or entangled with some other system.

Examples. In the three cases discussed in Sec. III, the channel capacity is just I_Q, as calculated. The important fact to note is that even very high levels of decoherence ($\epsilon\rightarrow 1$) can be tolerated in principle. A case of considerable interest is that in which each qubit system sent down the channel has a probability η of being decohered and randomized (the ''depolarizing'' channel of Ref. [12]). In this case,

$$\rho_{\overline{\alpha}}=\sum_{ii'=0,1}((1-\eta)/2|i\rangle\langle i'|\otimes|i\rangle\langle i'| +(\eta/4)|i\rangle\langle i|\otimes|i'\rangle\langle i'|).$$

Here $S_{\overline{\alpha}}$ can be calculated for blocks of length 1 and is equal to $-(3\eta/4)\log_2(\eta/4)-(1-3\eta/4)\log_2(1-3\eta/4)$, which is equal to 1 for $\eta\approx 0.252$. The highest rate of errors that can be corrected by an optimal coding procedure for sources with block length 1 is just above 1/4 (see also Ref. [12]). This example contrasts with the classical channel, in which arbitrarily high levels of noise can be tolerated in principle: quantum coding can correct for arbitrarily high levels either of noise, or of decoherence, but not of both together.

V. METHODS FOR IMPROVING THE LIMITS

The examples discussed above have looked at memoryless sources with block length 1. The methods developed in Sec. IV can give higher transmission rates if qubits are blocked together in groups of ℓ as described for quantum sources above. Each block of ℓ qubits can be regarded as a composite quantum symbol, so that the block source has a density matrix $\rho^\ell\otimes\cdots\otimes\rho^\ell$, where ρ^ℓ is a density matrix over ℓ quantum symbols, as discussed in the section on quantum sources above.

Theorems 2 and 3 apply to block sources over the composite symbols. The coherent information $I_Q(\rho^\ell)$ can be defined as above to be the difference between the entropy of the output $-{\rm tr}\rho^\ell_{\rm out}\log_2\rho^\ell_{\rm out}$ and the entropy increase when the first ℓ bits (with reduced density matrix ρ^ℓ) of a pure, fully entangled state of 2ℓ bits are sent down the channel. The channel capacity for sources with block length ℓ can then be defined as in Theorem 3 to be $C^\ell_Q=(1/\ell)\max_{\rho^\ell}I_Q(\rho^\ell)$. The maximization procedure used for finding the quantum channel capacity in general yields a different, potentially higher channel capacity for codes composed of the composite symbols. (For the classical memoryless channel, in contrast, blocking symbols in groups yields the same channel capacity as for the ordinary stationary, ergodic channel [14].) The Shor-Smolin code of Ref. [21] blocks inputs together in groups of $\ell=5$: for the depolarizing channel discussed above, this code surpasses by a small amount the channel capacity C^1_Q for sources with blocks of length 1, but does not surpass (nor necessarily attain) the quantum channel capacity C^5_Q for stationary, ergodic sources over blocks of length 5. Theorem 3 gives a constructive procedure for reading the quantum channel capacity obtainable using codewords with a finite correlation length. The maximum error rate for which quantum information can be reliably transmitted down such depolarizing channels using codewords with an infinite correlation length is not known. Theorem 3 shows, however, that for any noisy quantum channel, the maximum rate of transmission of quantum information using sources with arbitrary block size is given by the maximum over all ℓ of C^ℓ_Q.

VI. DISCUSSION

In practice, even if the channel capacity is not exceeded, the amount of noise and decoherence that can be tolerated is limited by the ability to encode and decode: as $N\rightarrow\infty$, the error in the transmitted state goes to zero, but the amount of quantum information processing that must be done to encode and decode becomes large. The encoding and decoding itself must be performed reliably.

The usefulness of the classical noisy coding theorem is also limited by coding difficulties: in particular, random codes are hard to encode and decode. In this respect, however, the quantum theorem has a considerable advantage. As Shannon and Weaver noted, random codes are effective because the bits that make up the signal have no apparent order. In the classical case, this implies that sequences of bits must appear random. In the quantum case, however, as long as the encoded signal is fully entangled, each qubit in the signal taken on its own appears to be completely random. As a result, the code words themselves may be highly regular: the hashing procedure of Ref. [21] is an example of a straightforward computational procedure for constructing codewords that are sufficiently random to attain the limits of

Theorems 2 and 3. In the classical case, random codes are hard to construct. In the quantum case, codes that are sufficiently random to attain the channel capacity may be constructed by a brief quantum computation.

In conclusion, this paper has derived fundamental limits about the amount of quantum information that can be sent reliably down a quantum channel using high-probability states of block sources, and has exhibited codes that attain those limits. In fact, almost all codes attain those limits. The limits given for block sources apply to any coding scheme that was codewords of finite length and to stationary, ergodic sources with finite correlation length. As with Shannon's classical noisy coding theorem, the rate of transmission of quantum information remains finite as the probability of error goes to zero.

ACKNOWLEDGMENTS

This work was supported in part by Grant No. N00014-95-1-0975 from the Office of Naval Research and by the Quantum Information and Computation initiative under Grant No. DAAH04-96-1-0386 from DARPA, administrated by the Army Research Office. The author would like to acknowledge the benefit of discussions with many colleagues, particularly with C. Caves, M. Nielsen, J. Preskill, and B. Schumacher.

APPENDIX A

Here we discuss the properties of ensembles of states. The idea behind the ensemble picture of quantum mechanics is to deal with mixtures and superpositions in the same formalism. Accordingly, a primary purpose of the ensemble picture is to make an explicit distinction between quantum states that can interfere with each other and quantum states that cannot. The ensemble picture is constructed so that different members of an ensemble cannot interfere with each other, while corresponding members of different ensembles can interfere. The second purpose of the ensemble picture is to keep track explicitly of the normalization of states, so that high-probability sets of states can be identified correctly.

As noted in Sec. II above, a quantum ensemble $\mathcal{E}_\psi=\{(|\psi_j\rangle,p_j)\}$ is a set of quantum states together with their probabilities. Ensembles are collections of vectors and share many properties of vectors. For example, if $\mathcal{E}_\phi=\{(|\phi_j\rangle,q_j)\}$ we can define a scalar product $\mathcal{E}_\psi\cdot\mathcal{E}_\phi=\Sigma_j\sqrt{p_jq_j}\langle\psi_j|\phi_j\rangle$. If $\mathcal{E}$ is normalized, then $\mathcal{E}\cdot\mathcal{E}=\mathrm{tr}\,\rho_\mathcal{E}=1$. (Note that the rule for obtaining the proper statistics is to associate a factor of $\sqrt{p_j}$ with each occurrence of $|\psi_j\rangle$.) This vectorlike character of ensembles allows the straightforward characterization of properties of quantum operators. For example, the trace-preserving character of the superscattering operator (Sec. II) can be summarized by the requirement that $\mathcal{E}_{|\phi_j\rangle}\cdot\mathcal{E}_{|\phi_{j'}\rangle}=\delta_{jj'}$. The ensemble picture of open system time evolution declares that *either* $|\phi_1\rangle$ goes to $|\psi_{1(1)}\rangle$ with probability $p_{1(1)}$ (i.e., with probability amplitude $\sqrt{p_{1(1)}}$) and $|\phi_2\rangle$ goes to $|\psi_{1(2)}\rangle$ with probability $p_{1(2)}$, . . . or $|\phi_1\rangle$ goes to $|\psi_{2(1)}\rangle$ with probability $p_{2(1)}$ and $|\phi_2\rangle$ goes to $|\psi_{2(2)}\rangle$ with probability $p_{2(2)}$, etc., with $j=3,4,\ldots$.

A type of ensemble that will prove useful below is one that is obtained by superposing corresponding states from two ensembles. If corresponding states have the same probability, for example, if $p_j=q_j$ for the ensembles $\mathcal{E}_\phi,\mathcal{E}_\psi$ above, then the (not necessarily normalized) ensemble of superpositions of α times the states of $\mathcal{E}_\phi$ plus β times the corresponding states of $\mathcal{E}_\psi$ is just $\{(\alpha|\phi_j\rangle+\beta|\psi_j\rangle,p_j)\}$, with density matrix ρ as above. In fact, because we will work with ensembles of high-probability states, which have approximately equal probabilities, this is the type of ensemble that we will have occasion to use below. If the corresponding states from the different ensembles do not have the same probabilities, then we write the ensemble of superposed states as $\mathcal{E}_{\alpha\phi+\beta\psi}=\{(\alpha|\phi_j\rangle+\beta|\psi_j\rangle,p_jq_j)\}$ to indicate the ensemble obtained by superposing α times the states of $\mathcal{E}_\phi$ plus β times the corresponding states of $\mathcal{E}_\psi$, together with a list p_jq_j of the probabilities of the individual states in the superposition. The superposition ensemble $\mathcal{E}_{\alpha\phi+\beta\psi}$ is defined to be the ensemble of unnormalized states $\{(\alpha\sqrt{q_j}|\phi_j\rangle+\beta\sqrt{p_j}|\psi_j\rangle,1)\}$. (Note that this ensemble is normalized even though its states and probabilities are not.) We specify superposition ensembles in this fashion to keep track explicitly of the normalization of the individual states in the superposition. The proper overall normalization of such ensembles is obtained as above by associating a factor of $\sqrt{p_j}$ with each $|\psi_j\rangle$ and a factor of $\sqrt{q_j}$ with each $|\phi_j\rangle$, so that

$$\rho_{\mathcal{E}_{\alpha\phi+\beta\psi}}=\sum_j\ \alpha\overline{\alpha}q_j|\phi_j\rangle\langle\phi_j|+\alpha\overline{\beta}\sqrt{q_jp_j}|\phi_j\rangle\langle\psi_j|+\beta\overline{\alpha}\sqrt{p_jq_j}|\psi_j\rangle\langle\phi_j|+\beta\overline{\beta}p_j|\psi_j\rangle\langle\psi_j|.$$

If we wish to superpose many ensembles, $\mathcal{E}_i=\{(|\psi_{j(i)}\rangle,p_{j(i)})\}$, we will use i to index the ensembles, and j to index the different members of each ensemble: e.g., $\mathcal{E}_\beta=\{(\Sigma_i\beta_i|\psi_{j(i)}\rangle,p_{j(i)})\}$ is the ensemble obtained by superposing the jth members of each of the ensembles with probability $p_{j(i)}$ associated with the jth member of the ith ensemble. $\mathcal{E}_\beta$ has density matrix $\rho_\beta=\Sigma_{j(i),j(i')}\beta_i\overline{\beta_{i'}}\sqrt{p_{j(i)}p_{j(i')}}|\psi_{j(i')}\rangle\langle\psi_{j(i)}|$. In this notation, states with different j cannot interfere, but states with the same j but different i can interfere.

This definition of superpositions of ensembles allows us to complete the identification of ensembles with vectors by defining $\alpha\mathcal{E}_\phi+\beta\mathcal{E}_\psi=\mathcal{E}_{\alpha\phi+\beta\psi}$. In addition, this definition of superposition makes a self-consistent connection between the ensemble and superscattering pictures of time evolution, a fact that will prove useful below. The ensemble picture is equivalent to the operator sum representation of superscattering operators described, e.g., in Ref. [20], but makes explicit the normalization of the ''superscattered'' states, a feature which will be required below to identify high-probability subspaces.

APPENDIX B

Proof of Theorem 1. Theorem 1 follows directly from the results of Refs. [1] and [3], where a detailed treatment of high-probability subspaces may be found. The proof goes as follows. If $|\psi\rangle$ is selected from the ensemble with probability $p_{|\psi\rangle}$, then

$$\sum_{|\psi\rangle} p_{|\psi\rangle}\langle\psi|P_{\widetilde{\mathcal{H}}^N}|\psi\rangle = \mathrm{tr}\ P_{\widetilde{\mathcal{H}}^N}\rho^N$$

is just the total probability of the set of *classical* high-probability sequences, in the manner of Shannon [14], and $\rightarrow 1$ as $N\rightarrow\infty$. As a result, for any $\epsilon>0$, N can be picked sufficiently large so that a state picked from any stationary, ergodic ensemble with density matrix ρ has overlap $\geqslant 1-\epsilon$ with some state in $\widetilde{\mathcal{H}}^N$, with probability $\geqslant 1-\epsilon$. Minimality follows, since $\mathcal{E}_{\widetilde{\mathcal{H}}^N}$ is itself an ensemble with density matrix $P_{\widetilde{\mathcal{H}}^N}/\mathrm{tr}P_{\widetilde{\mathcal{H}}^N}\rightarrow\rho\otimes\cdots\otimes\rho$ as $N\rightarrow\infty$. As a result, any other subspace that obeys the requirements of Theorem 1 asymptotically contains all members of $\widetilde{\mathcal{H}}^N$ except a set of measure 0 as $N\rightarrow\infty$. Minimality is a relatively weak property: $\widetilde{\mathcal{H}}^N$ need not be the only minimal subspace. But all other such minimal subspaces $\hat{\mathcal{H}}^N$ have approximately the same dimension: $\log(\dim\hat{\mathcal{H}}^N)/\log(\dim\widetilde{\mathcal{H}}^N\rightarrow 1)$ as $N\rightarrow\infty$. In analogy to the classical case, for the purposes of rating the channel capacity the quantity of interest is the dependence of the logarithm of the dimension of the minimal subspaces on N.

To see that a stationary, ergodic source with finite correlation length m can be adequately approximated by block sources, divide up the sequences emitted by the source into blocks of length ℓ, where $\ell>m$. Because of the finite correlation length, the even blocks taken as a group have density matrix $\rho^\ell\otimes\cdots\otimes\rho^\ell$, where ρ^ℓ is the reduced density matrix for ℓ bits of the stationary, ergodic source. The odd blocks have the same density matrix. That is, the even blocks for the stationary, ergodic source taken on their own are in fact a block source, as are the odd blocks. The two taken together can be regarded as two correlated block sources and their states fall with probability one into the tensor product of the high-probability Hilbert spaces for the two block sources. But this tensor product space is just the high-probability Hilbert space for a single block source that includes both even and odd blocks. The entropy per qubit of the block source is $(1/\ell)(-\mathrm{tr}\rho^\ell\log_2\rho^\ell)$, which, as ℓ becomes large, converges to the entropy per qubit of the stationary, ergodic source: so as ℓ becomes $\gg m$, the dimension of the high-probability Hilbert space for the block source converges to the high-probability Hilbert space for the stationary, ergodic source. As a result, as ℓ becomes much larger than m, the minimal subspace for the block source can be made arbitrarily close to a minimal subspace for the stationary, ergodic source. Since Theorems 1, 2, and 3 are theorems about minimal subspaces for various sources, they hold not only for block sources, but for stationary, ergodic sources as well. The discussion in Appendixes C and D holds for sources with block length l as described in Secs. II–V above; in the interest of compactness of notation, the index l will be suppressed.

APPENDIX C

Proof of Theorem 2. There are several ways to prove the noisy channel theorem. One way is to follow along the lines suggested in the text and analyze the channel's effect on entangled states. The following method of proof is closer in spirit to the classical derivation of channel capacity.

In the density matrix picture of the channel, the channel has the effect

$$|\alpha\rangle\langle\alpha|\rightarrow\rho_\alpha = \sum_{i_1,\ldots,i_N,i'_1,\ldots,i'_N} \alpha_{i_1,\ldots,i_N}\overline{\alpha_{i'_1,\ldots,i'_N}} \times\mathcal{S}(|\phi_{i_1}\rangle\langle\phi_{i'_1}|)\otimes\cdots\otimes\mathcal{S}(|\phi_{i_N}\rangle\langle\phi_{i'_N}|), \quad \text{(C1)}$$

where the sum is taken over high-probability sequences in which i appears approximately p_iN times.

Equivalently, in the ensemble picture,

$$|\alpha\rangle\rightarrow\mathcal{E}_\alpha = \left\{\left(\sum_{i_1,\ldots,i_N}\alpha_{i_1,\ldots,i_N}|\psi_{j_1(i_1)}\rangle\cdots \times|\psi_{j_N(i_N)}\rangle, p_{j_1(i_1)}\cdots p_{j_N(i_N)}\right)\right\} \quad \text{(C2)}$$

$$\equiv\left\{\left(\sum_i \alpha_{\mathbf{i}}|\psi_{\mathbf{j(i)}}\rangle, p_{\mathbf{j(i)}}\right)\right\}, \quad \text{(C3)}$$

where the superposition ensemble is defined as in Appendix A and has density matrix ρ_α. [As in Appendix A, the ensemble of output states corresponding can also be written in unnormalized form $|\alpha\rangle\rightarrow\{(|\alpha^{\mathbf{j}}\rangle,1)\}$, where $|\alpha^{\mathbf{j}}\rangle=\Sigma_{\mathbf{i}}\alpha_{\mathbf{i}}\sqrt{p_{\mathbf{j(i)}}}|\psi_{j(\mathbf{i})}\rangle$, with $\rho_\alpha=\Sigma_{\mathbf{j}}|\alpha^{\mathbf{j}}\rangle\langle\alpha^{\mathbf{j}}|$.] Theorem 1 implies that as $N\rightarrow\infty$, then, with probability 1, the states of $\mathcal{E}_\alpha$ are to be found in the Hilbert space $\widetilde{\mathcal{H}}^N_\alpha$ spanned by high-probability states of the form $P_{\mathrm{HP}}|\alpha^{\mathbf{j}}\rangle\sim|\alpha^{\mathbf{j}}_{\mathrm{HP}}\rangle$, where P_{HP} projects onto the high-probability set of states $|\psi_{j_1(i_1)}\rangle\cdots|\psi_{j_N(i_N)}\rangle$ in which $|\psi_{j(i)}\rangle$ appears approximately $p_ip_{j(i)}N$ times. By the same reasoning as in Theorem 1, the projection of ρ_α onto $\widetilde{\mathcal{H}}^N_\alpha$, $\rho^{\mathrm{HP}}_\alpha=P_{\mathrm{HP}}\rho_\alpha P_{HP}$, is indistinguishable from ρ_α as $N\rightarrow\infty$. The minimality of $\widetilde{\mathcal{H}}^N_\alpha$ follows as in Theorem 1: since $\langle\alpha^{\mathbf{j}}|\alpha^{\mathbf{j}}_{\mathrm{HP}}\rangle=\langle\alpha^{\mathbf{j}}|P_{\mathrm{HP}}|\alpha^{\mathbf{j}}\rangle\rightarrow\langle\alpha^{\mathbf{j}}|\alpha^{\mathbf{j}}\rangle$ as $N\rightarrow\infty$, any Hilbert space that asymptotically contains the states $|\alpha^{\mathbf{j}}\rangle$ also asymptotically contains the states $P_{HP}|\alpha^{\mathbf{j}}\rangle$, and hence their span, which is $\widetilde{\mathcal{H}}^N_\alpha$. This proves the first part of Theorem 2.

For the purposes of Theorem 2, it is not necessary to calculate the dimension of the output Hilbert space $\widetilde{\mathcal{H}}^N_\alpha$ exactly, but only to identify the part of the logarithm of this dimension that grows linearly with N as $N\rightarrow\infty$. The dimension of the output Hilbert space $\widetilde{\mathcal{H}}^N_\alpha$ is equal to one over the average overlap of two members of that space: $\dim\widetilde{\mathcal{H}}^N_\alpha=(\mathrm{tr}_{\mathrm{HP}}\rho^2_\alpha)^{-1}$, where the trace $\mathrm{tr}_{\mathrm{HP}}$ is taken over high-probability sequences only. We wish to calculate the average dimension of the output Hilbert space over α. We have,

$$\mathrm{tr}_{\mathrm{HP}}\rho^2_\alpha = \sum_{\mathbf{j},\mathbf{l}}\langle\alpha^{\mathbf{l}}_{\mathrm{HP}}|\alpha^{\mathbf{j}}_{\mathrm{HP}}\rangle\langle\alpha^{\mathbf{j}}_{\mathrm{HP}}|\alpha^{\mathbf{l}}_{\mathrm{HP}}\rangle = \sum_{\mathbf{i},\mathbf{i}',\mathbf{j},\mathbf{k},\mathbf{k}',\mathbf{l}}\alpha_{\mathbf{i}}\overline{\alpha_{\mathbf{i}'}}\alpha_{\mathbf{k}}\overline{\alpha_{\mathbf{k}'}}\sqrt{p_{\mathbf{j(i)}}p_{\mathbf{j(i}')}p_{\mathbf{l(k)}}p_{\mathbf{l(k}')}} \times\langle\psi_{\mathbf{l(k}')}|\psi_{\mathbf{j(i)}}\rangle\langle\psi_{\mathbf{j(i}')}|\psi_{\mathbf{l(k)}}\rangle, \quad \text{(C4)}$$

where the sum is taken over high-probability states only. Now average over α: using the fact that

$$\langle \alpha_{i_1,\ldots,i_N}\overline{\alpha_{i_1',\ldots,i_N'}}\rangle_\alpha = \dim\widetilde{\mathcal{H}}_{\rm in}^N \delta_{i_1 i_1'}\cdots\delta_{i_N i_N'} \approx p_{i_1}\cdots p_{i_N}\delta_{i_1 i_1'}\cdots\delta_{i_N i_N'} \equiv p_{\mathbf{i}}\delta_{\mathbf{i}\cdot\mathbf{n}'}$$

for high-probability sequences $i_1,\ldots,i_N\equiv\mathbf{i}$ and that $\langle \alpha_{k_1,\ldots,k_N}\overline{\alpha_{l_1,\ldots,l_N}}\alpha_{m_1,\ldots,m_N}\overline{\alpha_{n_1,\ldots,n_N}}\rangle_\alpha$ is equal to zero unless either $k_r=l_r$, $m_r=n_r$, or $k_r=n_r$, $l_r=m_r$, we obtain

$$\langle {\rm tr}_{\rm HP}\rho_\alpha^2\rangle_\alpha = \sum_{\mathbf{i},\mathbf{j},\mathbf{k},\boldsymbol{l}} p_{\mathbf{i}}p_{\mathbf{k}}p_{\mathbf{j(i)}}p_{\boldsymbol{l}(\mathbf{k})}\langle\psi_{\boldsymbol{l}(\mathbf{k})}|\psi_{\mathbf{j(i)}}\rangle\langle\psi_{\mathbf{j(i)}}|\psi_{\boldsymbol{l}(\mathbf{k})}\rangle + \sum_{\mathbf{i},\mathbf{j},\mathbf{k},\boldsymbol{l}} p_{\mathbf{i}}p_{\mathbf{k}}\sqrt{p_{\mathbf{j(i)}}p_{\mathbf{j(k)}}p_{\boldsymbol{l}(\mathbf{k})}p_{\boldsymbol{l}(\mathbf{i})}}\langle\psi_{\boldsymbol{l}(\mathbf{i})}|\psi_{\mathbf{j(i)}}\rangle \times\langle\psi_{\mathbf{j(k)}}|\psi_{\boldsymbol{l}(\mathbf{k})}\rangle - \sum_{\mathbf{i},\mathbf{j},\mathbf{k}} p_{\mathbf{i}}^2 p_{\mathbf{j(i)}}p_{\boldsymbol{l}(\mathbf{i})}\langle\psi_{\boldsymbol{l}(\mathbf{i})}|\psi_{\mathbf{j(i)}}\rangle \times\langle\psi_{\mathbf{j(i)}}|\psi_{\boldsymbol{l}(\mathbf{i})}\rangle, \quad \text{(C5)}$$

where the sums are taken over high-probability sequences and states as before. Equation (C5) reduces to

$$\langle {\rm tr}_{\rm HP}\rho_\alpha^2\rangle_\alpha = {\rm tr}_{\rm HP}(\rho_{\rm out}^2)^N + {\rm tr}_{\rm HP}(\rho_{\overline{\alpha}}^2)^N - {\rm tr}_{\rm HP}(\rho_{i/o}^2)^N, \quad \text{(C6)}$$

where $\rho_{\rm out}$ and $\rho_{\overline{\alpha}}$ are defined as above, $\rho_{i/o}=\Sigma_i p_i \mathcal{S}(|\phi_i\rangle\times\langle\phi_i|)\otimes|\phi_i\rangle\langle\phi_i|$, and $(\rho^2)^N=\rho^2\otimes\cdots\otimes\rho^2$. We can now use the fact that ${\rm tr}_{\rm HP}(\rho^2)^N = 2^{N\ {\rm tr}\rho\ \log_2\rho}$, which can be simply verified in a basis in which ρ is diagonal. We then have

$${\rm tr}_{\rm HP}(\rho_{\rm out}^2)^N = 2^{N\ {\rm tr}\rho_{\rm out}\log_2\rho_{\rm out}} = 2^{-NS_{\rm out}}, \quad \text{(C7)}$$

$${\rm tr}_{\rm HP}(\rho_{\overline{\alpha}}^2)^N = 2^{N\ {\rm tr}\rho_{\overline{\alpha}}\log_2\rho_{\overline{\alpha}}} = 2^{-NS_{\overline{\alpha}}}, \quad \text{(C8)}$$

$${\rm tr}_{\rm HP}(\rho_{i/o}^2)^N = 2^{N\ {\rm tr}\rho_{i/o}\log_2\rho_{i/o}} = 2^{-N\left(\sum_i \mathcal{S}_{\rm out}(i)+\mathcal{S}_{\rm in}\right)}. \quad \text{(C9)}$$

As $N\to\infty$, $\langle(\dim\widetilde{\mathcal{H}}_\alpha^N)^{-1}\rangle_\alpha$ goes to the largest of these three terms, of which the third is less than or equal to either of the first two. We have actually calculated the average of the inverse of the dimension of the output subspace; however, the standard deviation $\sqrt{\langle({\rm tr}_{\rm HP}\rho_\alpha^2)^2\rangle_\alpha - \langle{\rm tr}_{\rm HP}\rho_\alpha^2\rangle_\alpha^2}$ can be shown to be proportional to $({\rm tr}_{\rm HP}\rho_{\rm out}^2 {\rm tr}_{\rm HP}\rho_{\overline{\alpha}}^2)^{N/2}$ and so goes to zero exponentially faster in N than $\langle{\rm tr}_{\rm HP}\rho_\alpha^2\rangle_\alpha$ except when $S_{\overline{\alpha}}=S_{\rm out}$, in which case $C_Q=0$. As a result, the average of the inverse is the inverse of the average and the average dimension of $\dim\widetilde{\mathcal{H}}_\alpha^N$ is the smaller of $2^{-N\ {\rm tr}\rho_{\overline{\alpha}}\log_2\rho_{\overline{\alpha}}}$ and $2^{-NP\ {\rm tr}\rho_{\rm out}\log_2\rho_{\rm out}}$, proving the second half of Theorem 2. Note also that the standard deviation of the dimension of $\widetilde{\mathcal{H}}_\alpha^N$ as a fraction of the average dimension also goes to zero as $N\to\infty$, showing that almost all α correspond to an output space of the same dimension.

APPENDIX D

Proof of Theorem 3. The high probability subspace for this source has dimension $2^{-N\ {\rm tr}\rho\ \log_2\rho}$. Encode the basis states $|\chi_i^N\rangle$ for the source as *randomly chosen* orthogonal states $|\alpha_i^N\rangle$ in the high-probability subspace of a source that attains the channel capacity. The channel takes each $|\alpha_i^N\rangle$ to some state in the ensemble $\mathcal{E}_{\alpha_1}$ with minimal subspace $\widetilde{\mathcal{H}}_{\alpha_1}^N$. The average over $\alpha_{\not/}$ of the overlap $|\langle\psi_{\alpha_1}|\psi_{\alpha_j}\rangle|$ of states $|\psi_{\alpha_i}\rangle\in\widetilde{\mathcal{H}}_{\alpha_i}^N$, $|\psi_{\alpha_j}\rangle\in\widetilde{\mathcal{H}}_{\alpha_j}^N$, for $i\neq j$ can be calculated as in Appendix C and is equal to $1/\dim\widetilde{\mathcal{H}}_{\rm out}^N = 2^{N\ {\rm tr}\rho_{\rm out}\log_2\rho_{\rm out}}$. If $P_{\alpha_1}^N$ is the projection operator onto $\widetilde{\mathcal{H}}_{\alpha_1}^N$, we have

$${\rm tr}P_{\alpha_i}^N P_{\alpha_j}^N = 2^{-N(-{\rm tr}\rho_{\rm out}\log_2\rho_{\rm out}+{\rm tr}\rho_{\overline{\alpha}}\log_2\rho_{\overline{\alpha}})} = 2^{-NC_Q}. \quad \text{(D1)}$$

That is, as $N\to\infty$, the overlap between any two individual output subspaces approaches 0 as long as the quantum channel capacity is not zero. The dimension of the direct sum of the output subspaces remains less than or equal to the dimension of $\mathcal{H}_{\rm out}^N$ if and only if $-{\rm tr}\ \rho\log_2\rho\leqslant C_Q$:

$$\dim\oplus\sum_i \widetilde{\mathcal{H}}_{\alpha_i}^N \to 2^{-N({\rm tr}\rho\ \log_2\rho - {\rm tr}\rho_{\overline{\alpha}}\log_2\rho_{\overline{\alpha}})} = 2^{N(-{\rm tr}\rho_{\rm out}\log_2\rho_{\rm out}-\zeta)}, \quad \text{(D2)}$$

where $\zeta = C_Q - (-{\rm tr}\rho\ \log_2\rho)$. So if $\zeta\geqslant 0$, the source entropy does not exceed the channel capacity and the output states corresponding to different input basis states all fall into distinct subspaces. The overlap of any one output subspace with the direct sum of all the remaining subspaces goes as $2^{-N\zeta}$. If $\zeta<0$, the output subspaces overlap and no unique decoding is possible. This proves that C_Q is an upper limit on the channel capacity for ''typical'' codewords belonging to the high-probability subspace (i.e., for a set of measure 1 as $N\to\infty$), but it does not rule out the possibility of the use of a set of codewords of measure 0.

In the case $\zeta\geqslant 0$, a unitary decoding transformation can now be applied to the output states to put each vector $|\psi_{\alpha_1}^N\rangle\in\widetilde{\mathcal{H}}_{\alpha_i}^N$ into the form $|\chi_i^N\rangle\otimes|\psi^N\rangle$, in which vectors in different output subspaces but with the same $|\psi_{\mathbf{j(i)}}\rangle$ in Eq. (C3) give the same $|\psi^N\rangle$. The decoding transformation can be constructed along the lines of a Schmidt orthogonalization procedure as follows: (1) Pick a $\mathbf{j}$ and map the unnormalized output states $|\alpha_i^{\mathbf{j}}\rangle$ to $|\chi_i^N\rangle|\psi_{\mathbf{j}}^N\rangle$, where $|\psi_{\mathbf{j}}^N\rangle$ is also not necessarily normalized. (2) Pick a $\mathbf{j}'$ and map the unnormalized states $|\alpha_i^{\mathbf{j}'}\rangle$ to $|\chi_i^N\rangle|\psi_{\mathbf{j}'}^N\rangle$ by mapping the part of $|\alpha_i^{\mathbf{j}'}\rangle$ that is orthogonal to $|\alpha_i^{\mathbf{j}}\rangle$ to the part of $|\chi_i^N\rangle|\psi_{\mathbf{j}'}^N\rangle$ that is orthogonal to $|\chi_i^N\rangle|\psi_{\mathbf{j}}^N\rangle$: as noted schematically in Fig. 2(b), this step relies crucially on the fact that $\langle\alpha_i^{\mathbf{j}}|\alpha_i^{\mathbf{j}'}\rangle$ is asymptotically equal to $\langle\alpha_{i'}^{\mathbf{j}}|\alpha_{i'}^{\mathbf{j}'}\rangle$ as $N\to\infty$. In fact, the same averaging techniques used in deriving the dimension of the output subspaces in the proof of Theorem 2 above give

$$\frac{\langle|\langle\alpha_i^{\mathbf{j}}|\alpha_i^{\mathbf{j}'}\rangle - \langle\alpha_{i'}^{\mathbf{j}}|\alpha_{i'}^{\mathbf{j}'}\rangle|^2\rangle_{\alpha_i\alpha_{i'}}}{(\langle|\langle\alpha_i^{\mathbf{j}}|\alpha_i^{\mathbf{j}'}\rangle|^2\rangle_{\alpha_i}\langle|\langle\alpha_{i'}^{\mathbf{j}}|\alpha_{i'}^{\mathbf{j}'}\rangle|^2\rangle_{\alpha_{i'}})^{1/2}} = 2^{N({\rm tr}\rho_{\rm out}\log_2\rho_{\rm out}-{\rm tr}\rho_{\overline{\alpha}}\log_2\rho_{\overline{\alpha}})} = 2^{-NC_Q}$$

which approaches 0 as $N\rightarrow\infty$ if and only if $C_Q>0$. (3) Pick a j'' and continue as before, mapping the part of $|\alpha_i^{j''}\rangle$ that is orthogonal to the subspace generated by $|\alpha_i^{\mathbf{j}}\rangle$ and $|\alpha_i^{\mathbf{j}''}\rangle$ to the part of $|\chi_i^N\rangle|\psi_{\mathbf{j}''}^N\rangle$ that is orthogonal to $|\chi_i^N\rangle|\psi_{\mathbf{j}}^N\rangle$ and $|\chi_i^N\rangle|\psi_{\mathbf{j}'}^N\rangle$. (4) Continue until all the $\mathbf{j}$'s have been mapped.

Because of the asymptotic orthogonality of the output spaces, this decoding recreates $|\chi_i^N\rangle$ with fidelity arbitrarily close to 1 as $N\rightarrow\infty$. The crucial point is that this decoding also recreates *superpositions* of input states with fidelity approaching 1 as $N\rightarrow\infty$: the decoding process is unitary and preserves the amplitudes and phases of the $|\chi_i^N\rangle$ so that $\Sigma_k\gamma_k|\chi_k^N\rangle$ is mapped to an ensemble $\{(\Sigma_k\gamma_k|\chi_k^N\rangle\otimes|\psi^N\rangle,p_{|\psi^N\rangle})\}$. The steps are as follows: first, encoding

$$\sum_k \gamma_k|\chi_k^N\rangle\rightarrow\sum_k \gamma_k \sum_{i_1,\ldots,i_N} \alpha^k_{i_1,\ldots,i_N}|\phi_{i_1}\rangle\cdots|\phi_{i_N}\rangle; \tag{D3a}$$

next, the effect of the channel

$$\sum_k \gamma_k|\chi_k^N\rangle\rightarrow\left\{\left(\sum_k \gamma_k \sum_{i_1,\ldots,i_N} \alpha^k_{i_1,\ldots,i_N}|\psi_{j_1(i_1)}\rangle\cdots\times|\psi_{j_N(i_N)}\rangle p_{j_1(i_1)}\cdots p_{j_N(i_N)}\right)\right\}; \tag{D3b}$$

and finally, decoding

$$\rightarrow\left\{\left(\sum_k \gamma_k|\chi_k\rangle\otimes\sum_{i_1,\ldots,i_N} \beta_{i_1,\ldots,i_N}|\psi_{j_1(i_1)}\rangle\cdots\times|\psi_{j_N(i_N)}\rangle p_{j_1(i_1)}\cdots p_{j_N(i_N)}\right)\right\}$$
$$=\left\{\left(\sum_k \gamma_k|\chi_k\rangle\otimes|\psi^N\rangle,p_{|\psi^N\rangle}\right)\right\}. \tag{D3c}$$

The fact that the decoding process faithfully recreates superpositions can also be verified in the density matrix picture by using the correspondence in Appendix C. Since the encoding and decoding preserves pure states with their phases, it also preserves mixed states and any entanglement between the input state and another quantum system.

This proves the if part of the theorem. The only if part for codewords from the high-probability subspace was proved above. This proves the theorem as stated.

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

[2] C. Caves and P. D. Drummond, Rev. Mod. Phys. **66**, ( ) (1994).

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

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

[5] P. W. Shor, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, edited by S. Goldwasser (IEEE Computer Society, City, 1994), pp. 124–134.

[6] W. H. Zurek, Phys. Today **44** (1991).

[7] R. Landauer, Philos. Trans. R. Soc. London Ser. A **353**, 367 (1995); in *Proceedings of the Drexel-4 Symposium on Quantum Nonintegrability—Quantum Classical Correspondence*, edited by D. H. Feng and B. L. Hu (International, City, 1996).

[8] P. W. Shor, Phys. Rev. A **52**, 2493 (1995).

[9] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996).

[10] A. Steane, Proc. R. Soc. London (to be published).

[11] R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, Phys. Rev. Lett. **77**, 198 (1996).

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

[13] A. Ekert and C. Machiavello (unpublished).

[14] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Chicago, 1948).

[15] C. Monroe, D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland, Phys. Rev. Lett. **75**, 4714 (1995).

[16] Q. A. Turchette, C. J. Hood, W. Lange, H. Mabuchi, and H. J. Kimble, Phys. Rev. Lett. **75**, 4710 (1995).

[17] J. I. Cirac and P. Zoller, Phys. Rev. Lett. **74**, 4091 (1995).

[18] B. Schumacher (unpublished); P. Hausladen, R. Josza, B. Schumacher, M. Westmoreland, and W. K. Wootters, Phys. Rev. A **54**, 1869 (1996).

[19] A similar expression for quantum channel capacity C_Q has been derived independently by M. A. Nielsen and B. Schumacher [Phys. Rev. Lett. (to be published)], who call the quantity I_Q ''coherent information.''

[20] S. Lloyd, Ph.D. thesis, Rockefeller University, 1988 (unpublished).

[21] A method for using measure-zero codes that surpasses the limits given by Theorems 2 and 3 for blocks of lengths $l=1$ has recently been suggested by P. W. Shor and J. A. Smolin (unpublished). These elegant codes rely on versions of the entanglement purification protocols described in Ref. [4], and like the codes described here give finite rates of reliable transmission as $N\rightarrow\infty$. As noted in the discussion of Theorem 3, the Shor-Smolin codes are an example of a code that uses a blocking procedure over blocks of five qubits. The rate of reliable transmission that they obtain is consistent with the channel capacity C_Q^5 given by Theorem 3 for such blocked codes.

[22] For a discussion of quantum analogs of Hamming distance, see A. Ekert and C. Macchiavello (unpublished), and D. Gottesman (unpublished).