Channel Capacities

(Notes for Course on Quantum Computation and Information Theory. Sec. 15)

Robert B. Griffiths

Version of 7 April 2003

References:
QCQI = *Quantum Computation and Quantum Information* by Nielsen and Chuang (Cambridge, 2000), Secs. 11.3, 12.1 to 12.4.

15 Channel Capacities

15.1 Introduction

★ The theory of quantum data compression and quantum channel capacities is basically an attempt to generalize a number of classical concepts and results to quantum systems. While a lot of progress has been made (at least in the opinion of those who have been publishing papers on the subject!), at present this generalization is far from complete. There are many outstanding problems which are the topics of ongoing research.

• The purpose of these notes is to present an overview of some aspects of the subject, with an emphasis on the general ideas, not formal definitions or proofs. We will focus on the simplest cases: memoryless 1 bit classical channels and their 1 qubit quantum counterparts.

★ Von Neumann entropy of a density operator. It is defined to be

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho) = -\sum_j p_j \log p_j, \tag{15.1}$$

where $\{p_j\}$ are the eigenvalues of ρ.

○ The units of $S(\rho)$, as for other entropies, are determined by the base used for logarithms. We shall (unless stated otherwise) use base 2, so S is measured in bits.

○ Remark. A typical way of defining a function $f(A)$, where f is a numerical function and A operator which can be written in diagonal form using an orthonormal basis $\{|a_j\rangle\}$, i.e., as $A = \sum_j \alpha_j [a_j]$, is $f(A) = \sum_j f(\alpha_j)[a_j]$, assuming that f is defined for each eigenvalue α_j. The function $f(x) = x \log x$ is defined for $x \geq 0$ if we let $f(0) = 0$.

• Note that the von Neumann entropy is equal to the Shannon entropy for a probability distribution $\{p_j\}$.

- A number of useful properties of $S(\rho)$ are noted in QCQI Sec. 11.3.

15.2 Classical Error Correction and Channel Capacity

★ An important problem in classical information theory is that of *reliable* transmission of information from one place to another. Given that the world is not perfect, ultimate reliability is not achievable, but one would like systems which will transmit a message with a suitably small probability of error, less than some ϵ.

- Consider a one bit memoryless channel, and suppose that the probability of error is p if 0 is transmitted, and q if 1 is transmitted. That is, if a 0 enters the channel, the probability of its exiting as a 1 is p, and as a 0 (no error) is $1 - p$.

□ Exercise. Suppose $p = q = p$ and that the tolerable error rate for an n-bit message is 0.1 in the sense that 90% of the time the entire message should arrive error-free. What must p be if (i) n=1, (ii) n=5, (iii) n=25?

★ In order to send long messages reliably over a noisy channel it is necessary to correct errors. Error-correcting codes are used to transmit a message in a redundant form so that the original can be recovered, provided not too many errors are introduced in the process of transmission. A simple example is the 3-bit code in which to transmit a 0 one sends three 0's in succession, 000, and to transmit a 1, three 1's, 111. If at most one of the bits is corrupted during the transmission process, the original 0 or 1 can be recovered by the *majority rule* decoding scheme: at the receiving end, a 101 is interpreted as a 1, etc.

□ Exercise. Suppose the 3-bit code is used with a channel with $p = q = p$. What is the probability p' of an error in the output bit? How small must p be in order for this scheme to show an improvement? What is the degree of improvement when p is quite small?

- The basic process we are interested in is shown schematically in Fig. 15.1. An initial message M on the left is *encoded* by the device $\mathcal{C}$ and sent over the channel to where it is *decoded* by the device $\mathcal{D}$ to yield a message M'. If $M' = M$ the transmission was successful, while $M' \neq M$ means that there was an error which was not caught by the detection process. The probability of an error depends on the message M, as well as on the encoding and decoding procedures, and the noise present in the channel (i.e., p and q for our one bit channel). Let ϵ denote the maximum probability of error over all possible messages M.

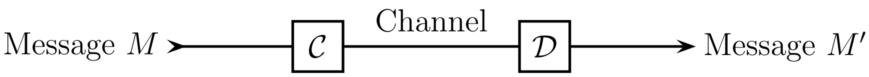

Figure 15.1: Coding and decoding.

★ The *capacity* of a noisy 1-bit channel can be defined in the following way (Cover and Thomas,[1] Sec. 8.5, somewhat modified). Let $[k, n]$ denote a code in which k successive bits

[1] *Elements of Information Theory* (Wiley, 1991).

of the message are encoded in n bits to form a codeword. The message is then transmitted by sending successive codewords, one bit at a time, through the channel. Each codeword is decoded at the receiving end in order to recover the original k bits. If these are identical to the original k-bit message, we say that it has been transmitted successfully; otherwise, if even one of the bits is wrong, there has been an error. Let us define the *uncertainty* of such a scheme as the maximum probability of error over *all* of the 2^k possible messages one can construct from k bits. The capacity of the channel C is then the upper limit, as k goes to infinity, of k/n for sequences of codes for which the uncertainty goes to zero.

• A more precise definition is the following. A number R is said to be an *achievable rate* for the channel if there is a sequence of codes $[k, n]$, along with an appropriate decoding method for each code, such that for all k sufficiently large in a sequence tending to infinity (e.g., $k = 4, 8, 20, \ldots$), with n is a function of k, it is the case that $k/n \geq R$, and the uncertainty (as defined above) goes to zero as $k \rightarrow \infty$. The capacity C is then the supremum over all achievable rates.

★ This type of definition has its utility, but actually using it to compute the capacity of a noisy channel is rather impractical. Shannon showed that one can define (or compute) the capacity using a far simpler procedure. Recall (Sec. 11.4) that for a noisy channel one can define a mutual information $I(X{:}Y)$ whose interpretation is that it is the average amount of information Bob learns, by observing the output, about what Alice inserted in the channel. This depends on the joint probability distribution $r(x, y)$ of the input bit x and the output bit y in the case of interest to us, which in turn is determined by the channel channel noise characteristics p and q, and also by the initial probability distribution $p(x)$ for the input. What Shannon showed is that

$$C = \max_{p(x)} I(X{:}Y). \tag{15.2}$$

While it is not (at least in general) a simple task to carry out the maximzation in closed form, it is not difficult to do it numerically.

• Note that (15.2) makes good intuitive sense, in that the average amount of information flowing through the channel every time a bit is transmitted is $I(X{:}Y)$ when successive bits are statistically independent. Finding useful codes which make good use of the channel is, of course, a nontrivial matter.

★ While sending the encoded message through the channel one bit at a time is the typical way of viewing channel capacity, there is an alternative which will be particularly helpful when we come to quantum channels. Rather than using one channel n times, we can think of n channels in parallel used just once to transmit a single codeword, as in Fig. 15.2.

• The capacity C is defined in the same way as before: the upper limit, as k goes to infinity, of k/n for sequences of codes for which the uncertainty goes to zero. Or use the achievable rate to obtain a more precise statement.

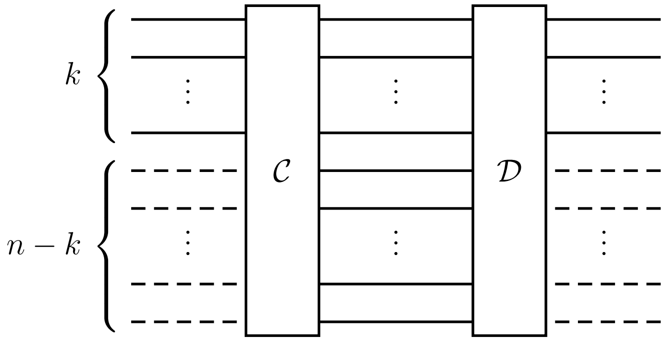

Figure 15.2: Coding and decoding using n parallel channels

15.3 Ideal and Noisy Quantum Channels

★ The distinction between quantum and classical channels is somewhat arbitrary, since in the real world all physical channels are, fundamentally, quantum objects. However, for many of them the process of transmitting information can be described quite satisfactorily in terms of classical physics. This is in particular the case for optical fibers in which the information resides in pulses of light containing large numbers of photons. We use the term "quantum channel" for cases in which classical concepts are not adequate for describing the transmission of information.

★ The simplest example of a quantum channel is the ideal 1-qubit channel in which if the qubit enters the channel in any state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ of its two-dimensional Hilbert space, it emerges from the channel in precisely the same state.

• The same definition will work for an ideal quantum channel associated with a Hilbert space of dimension d greater than 2. Usually we will be thinking of channels of dimension $d = 2^k$ corresonding to k qubits "in parallel".

• We will also use the term "ideal" for a channel in which the initial $|\psi\rangle$ is mapped to an output $U|\psi\rangle$, where U is a unitary map independent of $|\psi\rangle$. The reason is that the "noise" in this channel can be eliminated by simply placing a unitary transformation U^{-1} at the output of the channel, or at the input of the channel. The situation is analogous to that of a classical 1-bit channel which always maps 0 to 1 and 1 to 0: it is not really "noisy," because there is a simple way of processing the output in order to obtain a perfect channel.

• When a distinctive name is needed for the ideal channel in which $U = I$ (up to some phase factor), we shall refer to it as an *identity* channel, since the corresponding superoperator $\mathcal{Q}$ for the channel is the identity $\mathcal{I}$, which maps any operator to itself.

★ The noise of a noisy 1-qubit channel can be characterized in the following way. Pure states which enter the channel correspond to points on the surface of the Bloch sphere. Each such point is mapped into a density operator represented by a point lying inside the Bloch sphere (or possibly on its surface). The points which are images of the surface of the input sphere form an ellipsoid with three semi-major axes which are in general different.

◦ For some examples of such ellipsoids, see QCQI pp. 376ff.

• Since unitary transformations correspond to (proper) rotations of the Bloch sphere, a little thought will show that if we ignore (as in the ideal channel) changes produced by applying a unitary transformation to the output signal, a channel is characterized by 6 real parameters, which can be thought of as the semi-major axes of the ellipsoid and the vector from the center of the Bloch sphere to the center of the ellipsoid in a coordinate system in which the axes of the ellipsoid are lined up with the x, y, and z axes. By contrast, a 1-bit classical channel is characterized by two real numbers: the probability p that a 0 flips to a 1, and the probability q that a 1 flips to a 0.

○ The analog of a *symmetrical* classical channel in which $q = p$ is a 1-qubit channel in which the center of the ellipsoid is at the center of the Bloch sphere. There are but 3 parameters, the ellipsoid semiaxes, needed to characterize the noise in this case, in contrast to 1 parameter for a symmetrical classical channel.

★ In mathematical terms the channel is described by a *superoperator*. We will use $\mathcal{Q}$, rather than $\mathcal{E}$ as in QCQI, to avoid confusion with $\mathcal{E}$ as a symbol for an ensemble. If ρ is a density operator on the quantum Hilbert space $\mathcal{A}$ representing the entrance of the channel, then $\rho' = \mathcal{Q}(\rho)$ is the output density operator on the channel output $\mathcal{B}$ (which need not have the same dimension as $\mathcal{A}$.) The linear map $\mathcal{Q}$ has various properties discussed in QCQI, Sec. 8.2.

• The channel can also be thought of as produced by a single unitary transformation acting on the tensor product of the channel input Hilbert space $\mathcal{A}$, and an "environment" $\mathcal{C}$ in a particular pure state, and mapping $\mathcal{A} \otimes \mathcal{C}$ to $\mathcal{B} \otimes \mathcal{D}$. The partial trace of the resulting state over $\mathcal{D}$ yields ρ'.

15.4 Classical Capacity of a Quantum Channel

★ A topic much discussed in quantum information theory is the *classical capacity* C_C of a quantum channel for "classical information." While this may seem like (and often is) a purely theoretical enterprise, there is a reason why it might one day be of some interest. All channels used for high-speed communication of classical data, such as optical fibers, are, in fact, quantum mechanical in nature. Most of the time one can ignore this, because classical physics suffices to describe the process of information transfer: an optical pulse contains a large number of photons. But as technology improves, there will be (one suspects) a tendency for the engineers to use fewer photons per pulse in order to send more messages down the fiber, and eventually one reaches a limit in which quantum effects will be important.

• The basic idea of sending classical information over a quantum channel is the following. At one end big, macroscopic "classical" signals, are converted into states of individual carriers of quantum information, which are then sent over the quantum channel. At the other end these carriers are further processed or measured in order to again produce big, macroscopic signals. Both the beginning and the end signals are of a sort which can by analyzed using standard (classical) information theory, so the classical capacity can be computed using well-known formulas.

• The trouble is that there are a large number of different ways of converting classical signals into quantum states, and transforming such states back into classical signals. Presumably the capacity should be based on the optimum procedures for doing this. But what are they?

★ Here is one procedure. Pick an orthonormal basis $A = \{|a_0\rangle, |a_1\rangle\}$ at the input and a second orthonormal basis $B = \{|b_0\rangle, |b_1\rangle\}$ at the output of the noisy 1 qubit channel. If the initial state is one of the A states and we measure in the B basis, then the channel properties are determined by a set of conditional probabilities $\Pr(b_k \,|\, a_j)$, just as in a one-bit classical channel, and we can use them to calculate a capacity $C(A, B)$ as in (15.2). Thus we might define a classical capacity

$$C_a = \max_{A,B} C(A, B). \tag{15.3}$$

• However, there are other possibilities. Rather than measure the output in an orthonormal basis, we might carry out a POVM measurement $\mathcal{P}$, which can be thought of as a processing step in which the output qubit interacts with ancillary qubits through a specific unitary interaction, followed by a measurement (of the ordinary or projective sort) on this collection of qubits. As measurements in a particular basis are included among possible POVM's, this approach cannot decrease the maximum mutual information, and might even increase it.

• Similarly, rather than restricting the input to states of a given orthonormal basis, we might allow them to be drawn from a general ensemble. It is convenient to introduce two types of ensembles. An *ensemble of mixed states* will be denoted by

$$\mathcal{E} = \{p_x, \rho_x\}, \tag{15.4}$$

where x is just a label identifying different elements in the ensemble (you can think of it as an integer), $\{p_x\}$ is a collection of probabilities (each $p_x > 0$, $\sum_x p_x = 1$), and the ρ_x are density operators, all on the same Hilbert space. In an *ensemble of pure states*, each ρ_x is a projector $[\psi_x]$ on some $|\psi_x\rangle$, and for such an ensemble one can use a notation employing kets rather than the projectors,

$$\mathcal{F} = \{p_x, |\psi_x\rangle\}. \tag{15.5}$$

★ We now define a second capacity C_b of a quantum channel for classical information,

$$C_b = \sup_{\mathcal{F},\mathcal{P}} I(\mathcal{F}{:}\mathcal{P}), \tag{15.6}$$

where the maximum or suprememum is taken over all pure-state ensembles at the input, and all POVM measurements $\mathcal{P}$ at the output, with I the corresponding Shannon mutual information between Alice and Bob.

◦ We could have used mixed-state ensembles $\mathcal{E}$ in (15.6), for these include pure states as a special case, and one knows that it suffices to consider pure states.

◦ The quantity $I(\mathcal{F}{:}\mathcal{P})$ is to be thought of in the following sense. Alice uses a random number generator to generate a label x with probability p_x, records the value of x in her notebook, and then prepares a state $|\psi_x\rangle$ and sends it to Bob. Bob carries out his POVM measurement on the state sent by Alice, and this allows him to guess a value x' for x, which he records in his notebook. From the conditional probability $c(x'|x)$ of x' given x together with the probability p_x, one can calculate the mutual information between Alice and Bob using ordinary information theory. Then C_b is the maximum of this quantity over all possible input ensembles and output measuring strategies. Note that this includes varying p_x, the only sort of optimization one worries about in the case of a classical channel.

★ Both C_a and C_b are lower bounds, though in some cases they might be equal to, the "real" classical capacity C_r of a quantum channel, defined as follows. Recall that in the case of a classical channel we allow messages of k bits to be encoded in some fashion in strings of n bits which are then sent successively, or in parallel in the manner indicated in Fig. 15.2, through the noisy channel, and decoded at the other end. The obvious analog if one replaces the n noisy classical channels in the middle part of the figure with n noisy quantum channels is to let let $\mathcal{C}$ in Fig. 15.2 be a device which for each k-bit message produces some (in general entangled) quantum state on the tensor product of the n Hilbert spaces representing the inputs to the channnels. Let $\mathcal{D}$ represent an arbitrary POVM measurement on the collection, regarded as a single quantum system, of n qubits emerging from these channels after the noisy transmission, together with some algorithm which uses the output of the POVM to guess the value of the original k-bit string. Let ϵ be the uncertainty, defined (as earlier) as the maximum, over all initial k-bit strings, of the probability of an error: the probability that the output string differs from the input string. Then, in analogy with the classical case, C_r should be the upper limit of k/n as k goes to infinity, for sequences of encoding and decoding devices $\mathcal{C}$ and $\mathcal{D}$ chosen so that ϵ tends to zero in the same limit. Once again, a more precise definition can be formulated using the achievable rate, as in Sec. 15.2.

• The trouble with C_r is that no one knows how to calculate it, even numerically, for a general one-qubit quantum channel. In the quantum case there is nothing like the nice, simple expression (15.2) for classical channels. The best one can do is to place certain bounds on C_r. Thus it is obvious that $C_a \leq C_b \leq C_r$, and there is a still better lower bound C_p to C_r discussed at the end of the next section.

15.5 Holevo Function and Bound; Classical Product Capacity

★ Following (in part) the notation of Preskill's notes, Sec. 5.3, we let $\mathcal{E} = \{p_x, \rho_x\}$ be an ensemble, where the $\{p_x\}$ are probabilities (positive, sum to 1) and the $\{\rho_x\}$ are density operators on the same Hilbert space. The *Holevo function* χ is defined by

$$\chi(\mathcal{E}) = S(\sum_x p_x\rho_x) - \sum_x p_x S(\rho_x) = S(\rho) - \sum_x p_x S(\rho_x), \tag{15.7}$$

where

$$\rho = \sum_x \rho_x \tag{15.8}$$

is the density operator for the ensemble.

• As emphasized by Preskill, $\chi(\mathcal{E})$ is a function of the *ensemble* $\mathcal{E}$ and not a function of (just) its density operator ρ. If the ensemble consists of pure states, $\rho_x = [\psi_x]$ for every j, then $\chi(\mathcal{E}) = S(\rho)$ is the same as the von Neumann entropy of ρ, but in other cases it is smaller. Recall that, except for a pure state, there are many different ensembles which can be associated with the same density operator ρ in the sense of (15.8).

★ The Holevo function occurs in the *Holevo bound*

$$I(A{:}B) \leq \chi(\mathcal{E}), \tag{15.9}$$

interpreted in the following way. Alice chooses at random with probability p_x from the ensemble $\mathcal{E}$ a system described by the density operator ρ_x, notes down value of x, and gives the system (but not the value of x) to Bob. His task is to carry out whatever measurements he wants to, including POVM's, on this system in order to determine the value of x. A particular strategy will result in Bob making an estimate x' for x. Imagine that the same protocol is carried out a number of times, with x_j and x'_j the values for trial j. Then $I(A{:}B)$ is the mutual information per pair x_j, x'_j between the string $\{x_j\}$ in Alice's notebook and the string $\{x'_j\}$ in Bob's notbook.

◦ Both Bob and Alice are assumed to know which ensemble $\mathcal{E} = \{\{p_x, \rho_x\}$ is in use, and Bob can develop his measurement strategy accordingly.

◦ Note that Bob's task is to estimate the value of x, which is not the same thing as "measuring ρ_x." Unless ρ_x and $\rho_{x'}$ are orthogonal, i.e., $\rho_x\rho_{x'} = 0$, there is no way to reliably distinguish the two by means of a measurement, just as there is in general no way to distinguish two classical probability distributions in terms of a single trial unless they do not overlap.

• Alice's task is reasonably clear when $\mathcal{E}$ is an ensemble of pure states $\{\{p_x, |\psi_x\rangle\}$. But what if the some element is a mixed state? How is Alice to prepare *that*? The simple answer is that every mixed state ρ_x is itself associated with an ensemble of pure states occurring with certain probabilities, and Alice can pick one of these with the appropriate probability. So after running the first random number generator to determine x, she runs a second random number generator to produce a label y for one of the elements of the ensemble associated with ρ_x, and chooses the corresponding $|\psi_y\rangle$.

• An alternative way to think of the ρ_x mixed states is as follows, A machine has been programmed to produce a state $|\psi_x\rangle$ when supplied with the classical signal x, and signals are fed to it at random, x with probability p_x. The result is the pure state ensemble $\{\{p_x, |\psi_x\rangle\}$. But the machine is getting old and unreliable, and measurements show that the x signal does not always result in $|\psi_x\rangle$. Instead, the output must be described by a mixed state density operator ρ_x. Consequently, $\chi(\mathcal{E})$ is reduced by the negative terms on the right side of (15.8), reflecting the fact that Bob's task of identifying the correct x has become more difficult, due to this additional source of noise.

• Proving the Holevo bound is not trivial. Proofs are given in QCQI Sec. 12.1.1, and in Preskill's notes, Sec. 5.4.1.

★ In the case of an ensemble of pure states, the Holevo bound (15.9) states that $NS(\rho)$ is the most information which Bob can extract about the x values when Alice sends randomly chosen states of the ensemble a large number of times. This is closely connected with Schumacher's result on quantum data compression discussed below.

★ The Holevo function also enters into yet one more (in addition to those discussed in Sec. 15.4) measure of the classical capacity of a quantum channel. It is what QCQI, Sec. 12.3.2, call the *product capacity*, and which we will denote by the symbol C_p. It is similar to C_r except that instead of letting $\mathcal{C}$ in Fig. 15.2 represent an encoding of a k-bit string as an arbitrary quantum state on the n input channels, one only allows *product* states of the general form $|\phi_1\rangle \otimes |\phi_2\rangle \otimes \cdots \otimes |\phi_n\rangle$, where the subscript j refers to the j'th channel.

• A result of Holevo, Schumacher, and Westmoreland, QCQI Sec. 12.3.2, is that C_p is given by the formula

$$C_p = \max_{\mathcal{F}} \chi(\mathcal{Q}(\mathcal{F})), \tag{15.10}$$

where χ is the Holevo function defined in (15.7), and the maximization is over all ensembles of the form (15.5). By $\mathcal{Q}(\mathcal{F})$ we mean the ensemble obtained by sending each of the states in $\mathcal{F}$ through the quantum channel described by the superoperator $\mathcal{Q}$:

$$\mathcal{Q}(\mathcal{F}) = \{p_x, \mathcal{Q}([\psi_x])\}. \tag{15.11}$$

(As usual, $[\psi_x]$ denotes the projector onto $|\psi_x\rangle$.) Note that the new ensemble will (in general) be an ensemble of mixed states, of the type (15.4).

○ In QCQI the maximum in (15.10) is taken over all ensembles of mixed states, including pure states as special cases. Since the maximum is actually achieved on the pure states, it suffices to restrict consideration to the latter.

★ We have now introduced four definitions of the classical capacity of a quantum channel. They are ordered by the inequalities

$$C_a \leq C_b \leq C_p \leq C_r. \tag{15.12}$$

According to QCQI the experts think C_p may be equal to C_r. But as they have not been able to prove it, some scepticism is in order. What is not in doubt is that for a particular channel, for which $\mathcal{Q}$ is known, finding C_p using (15.10) by some numerical method, even though it looks rather formidable, is an easier task than finding C_b, and is very much simpler than trying to evaluate C_r directly in terms of the definition given in Sec. 15.4

15.6 Quantum Data Compression

• Reference: QCQI Sec. 12.2. (Also see Sec. 11.2 of these notes.)

★ A classical data compression result due to Shannon is stated and proved in QCQI Sec. 12.1. The essential idea was discussed in Sec. 11.2 of these notes. Given a random variable X with a probability distribution p_x, thought of as a message, then the minimum number of bits required to store or transmit n such messages, when n is large, is given approximately by $nH(X)$, where $H(X) = -\sum_x p_x \log p_x$ is the Shannon entropy associated with this probability distribution. For the meaning of "approximately," see the discussion in QCQI.

★ A quantum counterpart of this classical result is Schumacher's noiseless channel coding theorem, QCQI Sec. 12.2.2.[2] Here the idea is that a source produces a quantum state $|x\rangle$ in a Hilbert space $\mathcal{H}$ with probability p_x, and one wishes to store the information in a long string of n such states, i.e., in

$$|\Psi\rangle = |x_1\rangle \otimes |x_2\rangle \otimes \cdots |x_n\rangle \in \mathcal{H}^{\otimes n}, \tag{15.13}$$

in an efficient fashion. The idea is to be able with high probability to store messages of this sort on a smaller Hilbert space of dimension d in such a way that they can be recovered with high accuracy. High accuracy is interpreted to mean that the fidelity of the recovered message compared with the original is with high probability very close to 1. The result of Schumacher is that this can be done if d is of the order of 2^{nS}, or larger, where S is the von Neumann entropy, (15.1), of the density operator

$$\rho = \sum_x p_x[x] \tag{15.14}$$

corresponding to the ensemble of states in question.

○ The rigorous statement requires that n be sufficiently large in a sense determined by small constants $\epsilon > 0$, $\delta > 0$; see QCQI.

• In the case in which the different elements of the ensemble $\{|x\rangle\}$ are orthogonal to each other, Schumacher's result reduces in an obvious way to the classical result. One simply measures each element on the right side of (15.13) to see what state it is in, writes down the answer, compresses it using a classical algorithm to get a string of bits, creates a corresponding set of qubits so that one is storing the information in a proper quantum Hilbert space, etc.

• So the interesting case is the one in which at least some of the states are *not* orthogonal to each other. Here measurements (at least of the one-by-one type) will not work. Instead, the compression scheme is based on the following idea. There is a subspace (in QCQI the subspace is denoted by $T(n,\epsilon)$ and the corresponding projector is $P(n,\epsilon)$) of $\mathcal{H}^{\otimes n}$ with the property that with high probability a state of the form $|\Psi\rangle$ lies inside it, but its dimension d is of order 2^{nS}. Thus the Hilbert space constituted by this subspace contains the information required to reproduce $|\Psi\rangle$.

• One can imagine a unitary map which carries the subspace of interest into some other quantum system, say a collection of qubits. One will need approximately nS qubits for the storage if S is measured in bits.

★ What may at first seem surprising is that a state of the form $|\Psi\rangle$ contain a lot less "quantum" information than one might suppose by analogy with a set of classical symbols.

[2] Their discussion is somewhat more general, as they are interested in processes of compression where the quantum carriers of information are entangled with others not participating in the compression process.

For example, suppose that the quantum ensemble consists of the six states $|x^+\rangle$, $|x^-\rangle$, $|y^+\rangle$, $|y^-\rangle$, $|z^+\rangle$, $|z^-\rangle$ in spin-half or Bloch sphere notation, and each occurs with probability 1/6. The von Neumann entropy is 1 bit, whereas the entropy associated with six distinct classical symbols occurring with equal probability is $\log 6 = 2.58$ bits — more than twice as much.

○ The point is that when viewed as quantum objects, distinct classical objects (such as a magnetic domain oriented in one direction rather than its opposite) always correspond to orthogonal quantum states. But in the ensemble under discussion various pairs, such as $|x^+\rangle$ and $|y^+\rangle$, are not orthogonal.

15.7 Quantum Capacity of a Quantum Channel

★ To understand what is meant by the *quantum* capacity of a quantum channel, it is helpful to start with the notion of an *ideal* channel as defined in Sec. 15.3, in which states which enter the channel are mapped to the output by a unitary transformation U which is fixed, in the sense that it does not depend upon the input state. For purposes of exposition it is simplest to focus on the identity channel, $U = I$.

• While one usually thinks of sending n quantum signals successively through a single quantum channel, it is for conceptual purposes quite convenient to imagine identical channels placed in parallel, as in the middle section of Fig. 15.2.

★ It will be convenient to define the quantum capacity Q of an *ideal* channel to be

$$Q = \log d, \tag{15.15}$$

where d is the dimension of the Hilbert space representing the input of the channel, or its output, since the two are the same. If the logarithm is to base 2, Q is measured in *bits*. (Calling the unit a "qubit" would hinder useful comparisons between quantum capacities and classical capacities.)

• If two ideal channels are placed in parallel to form a single ideal channel, the input space of the latter is simply the tensor product, call it $\mathcal{A}_1 \otimes \mathcal{A}_2$ of the two input spaces, and the logarithmic measure in (15.15) means the total capacity will be the sum of the two capacities, since the dimension of a tensor product is the product of the dimensions. It is the same for three or more channels in parallel. In particular, n 1-qubit ideal channels in parallel have a capacity of n bits.

□ Exercise. Suppose that U and V are unitary operators (not necessarily the identity) mapping input to output for two ideal channels. If the two channels are in parallel the corresponding map on the tensor product is $U \otimes V$. Check that the combined channel is ideal in that not only product states but also entangled states in the input Hilbert space are mapped to the output in such a way that a suitable fixed unitary W applied to the output results in an identity channel.

★ To define the quantum capacity of a noisy one-qubit quantum channel, we again think of n identical (i.e., the noise characteristics are the same) channels in parallel, as in Fig. 15.2. However, the incoming and outgoing lines in this figure now represent qubits. On the left side, the first k qubits carry the quantum signal, and the last $n-k$ are ancillary qubits in some fixed state, which we assume to be $|0\rangle$. The encoding transformation $\mathcal{C}$ is a unitary operator on the Hilbert space of n qubits, as is the decoding transformation $\mathcal{D}$. The first k qubits on the right carry the output of the quantum signal, while the last $n-k$ are simply thrown away.

• Suppose the input signal is some (normalized) state $|\Psi\rangle$ in the 2^k-dimensional Hilbert space of the first k qubits on the left, and R_Ψ is the density operator for the corresponding output signal carried by the first k qubits on the right. The fidelity is defined as:

$$F_\Psi = \langle\Psi|R_\Psi|\Psi\rangle. \tag{15.16}$$

Let us say that the uncertainty is

$$\epsilon = \max_\Psi(1 - F_\Psi). \tag{15.17}$$

• We then define the quantum capacity Q to be the upper limit of k/n as k goes to infinity, for sequences of encoding and decoding devices chosen so that the uncertainty ϵ tends to zero as k becomes infinite.

◦ QCQI prefer to use the entanglement fidelity of a channel rather than the fidelity as defined in (15.16). However, for present purposes this makes no difference, as it has been shown[3] that if F in (15.17) is replaced by the entanglement fidelity, the resulting ϵ can be larger (it cannot be smaller) by at most a factor of 3/2.

◦ This is basically the same as the definition used in Sec. 15.4 for C_r, the "real" classical capacity, if we replace incoming and outgoing classical signals with quantum signals. The main difference is in the definition of the uncertainty. A good quantum coding scheme has to work reasonably well for *any* input state in the 2^k-dimensional Hilbert space.

◦ The same sort of definition will work for the quantum capacity of a channel with a Hilbert space of dimension $d > 2$. In this case the number n of channels in the center part of Fig. 15.2 will be smaller than the total number of qubits entering on the left, or leaving on the right. The capacity is again given by the upper limit of k/n as k becomes infinite, for a sequence or coding and decoding operations chosen so that the uncertainty goes to zero.

□ Exercise. Show that $Q \leq C_r$.

• Relating Q to the specific noise properties of a given quantum channel is difficult. In QCQI Sec. 12.4 a quantity called the *coherent information* is introduced as a sort of quantum analog of Shannon's mutual information. However, the identification is at best tentative, and its relation to the quantum capacity is not clear.

• Quantum capacities can also be defined for quantum channels assisted by classical communication, in one or both directions, between Alice and Bob. A discussion of these lies outside the scope of these notes.

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