In *Geom. Funct. Anal. (GAFA)*, Special Volume—GAFA2000, 816–838 (2000).

Quantum Information Theory: Results and Open Problems[1]

Peter Shor
AT&T Labs—Research, Florham Park, NJ 07932

1 Introduction

The discipline of information theory was founded by Claude Shannon in a truly remarkable paper [28] which laid down the foundations of the subject. We begin with a quote from this paper which is an excellent summary of the main concern of information theory:

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.

Quantum information theory is motivated largely by the same problem, the difference being that either the method of reproduction or the message itself involves fundamentally quantum effects. For many years, information theorists either ignored quantum effects or approximated them to make them susceptible to classical analysis; it was only in the last decade or so that the systematic study of quantum information theory began. We next give a quote from John R. Pierce which shows roughly the state of quantum information theory a quarter century ago. In a 1973 retrospective [25], celebrating the 25th anniversary of Shannon's paper, Pierce says

> I think that I have never met a physicist who understood information theory. I wish that physicists would stop talking about reformulating information theory and would give us a general expression for the capacity of a channel with quantum effects taken into account rather than a number of special cases.

In retrospect, this quote seems both optimistic and pessimistic. It was certainly pessimistic in that there are now many physicists who understand information theory, and I believe that even when Pierce wrote this, there were several who did. Ironically, one of the first fundamental theorems of quantum information theory was proved in the same year [17]. On the other hand, Pierce was quite optimistic in that he seems to have believed that finding the capacity of a quantum channel would be fairly straightforward for a physicist with the right background. This has not proven to be the case; even now, we do not have a general formula for the capacity of a quantum channel. However, there have been several recent fundamental advances made in this direction, and I describe these in this paper.

[1] A large part of this paper is included in the paper "Quantum Shannon Theory," which will appear in the IEEE Information Theory Society Newsletter.

2 Shannon theory

Shannon's 1948 paper [28] contained two theorems for which we will be giving quantum analogs. The first of these is the *source coding* theorem, which gives a formula for how much a source emitting random signals can be compressed, while still permitting the original signals to be recovered with high probability. Shannon's source coding theorem states that n outputs of a source X can be compressed to length $nH(X) + o(n)$ bits, and restored to the original with high probability, where H is the entropy function. For a probability distribution with probabilities $p_1, p_2, \ldots, p_n$, the entropy H is

$$H(\{p_i\}) = \sum_{i=1}^{n} -p_i \log p_i, \tag{1}$$

where information theorists generally take the logarithm base 2 (thus obtaining bits as the unit of information).

The second of these theorems is the *channel coding* theorem, which states that with high probability, n uses of a noisy channel N can communicate $Cn - o(n)$ bits reliably, where C is the channel capacity given by

$$C = \max_{p(X)} I(X; N(X)) \tag{2}$$

Here the maximum is taken over all probability distributions on inputs X to the channel, and $N(X)$ is the output of the channel given input X. The *mutual information* I is defined as:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) &(3)\\ &= H(X) + H(Y) - H(X,Y), &(4)\end{aligned}$$

where $H(X,Y)$ is the entropy of the joint distribution of X and Y, and $H(Y|X)$ is the conditional entropy of Y, given X. That is, if the possible values of X are $\{X_i\}$, then the conditional entropy is

$$H(Y|X) = \sum_i \Pr(X = X_i)\, H(Y|X = X_i). \tag{5}$$

In this paper, I outline the progress that has been made in extending these formulae to quantum channels, while also taking a few side detours that address related problems and results in quantum information theory. I will keep this paper at a fairly low technical level, so I only sketch the proofs for some of the results I mention.

When the formula for mutual information is extended to the quantum case, two generalizations have been found that both give capacities of a quantum channel, although these capacities differ in both the resources that the sender and receiver have available and the operations they are permitted to carry out. One of these formulae generalizes the expression (3) and the other the expression (4); these expressions are equal in the classical case.

3 Quantum mechanics

Before we can start talking about quantum information theory, I need to give a brief description of some of the fundamental principles of quantum mechanics. The first of these principles that we present is the *superposition principle*. In its most basic form, this principle says that if a quantum system can be in one of two distinguishable states $|\,x\rangle$ and $|\,y\rangle$, it can be in any state of the form $\alpha\,|\,x\rangle + \beta\,|\,y\rangle$, where α and β are complex numbers with $|\alpha|^2 + |\beta|^2 = 1$. Here $|\,\cdot\rangle$ is the notation that physicists use for a quantum state; we will occasionally be using it in the rest of this paper. Recall we assumed that $|\,x\rangle$ and $|\,y\rangle$ were distinguishable, so there must conceptually be some physical experiment which distinguishes them (this experiment need not be performable in practice). The principle says further that if we perform this experiment, we will observe $|\,x\rangle$ with probability $|\alpha|^2$ and $|\,y\rangle$ with probability $|\beta|^2$. Furthermore, after this experiment is performed, if state $|\,x\rangle$ (or $|\,y\rangle$) is observed the system will thereafter behave in the same way as it would have had it originally been in state $|\,x\rangle$ (or $|\,y\rangle$).

Mathematically, the superposition principle says that the states of a quantum system are the unit vectors of a complex vector space, and that two orthogonal vectors are distinguishable. In accordance with physics usage, we will denote quantum states by column vectors. The Dirac *bra-ket* notation denotes a column vector by $|\,v\rangle$ (a *ket*) and its Hermitian transpose (i.e., complex conjugate transpose) by $\langle v\,|$ (a *bra*). The inner product between two vectors, v and w, is denoted $\langle w|v\rangle = w^\dagger v$, where $w^\dagger$ is the conjugate transpose of w. Multiplying a quantum state vector by a complex phase factor (a unit complex number) does not change any properties of the system, so mathematically the state of a quantum system is a point in projective complex space. Unless otherwise stated, however, we will denote quantum states by unit vectors in a complex vector space $\mathbb{C}^d$.

We will be dealing solely with finite dimensional vector spaces. Quantum information theory is already complicated enough in finite dimensions without introducing the additional complexity of infinite-dimensional vector spaces. Many of the theorems we will be discussing do indeed generalize naturally to infinite-dimensional spaces.

A *qubit* is a two-dimensional quantum system. Probably the most widely known qubit is the polarization of a photon, and we will thus be using this example in the remainder of the paper. For the polarization of a photon, there can only be two distinguishable states. If one sends a photon through a birefringent crystal, it will take one of two paths, depending on its polarization. By re-orienting this crystal, these two distinguishable polarization states can be chosen to be horizontal and vertical, or they can be chosen to be right diagonal and left diagonal. In accordance with the superposition principle, each of these states can be expressed as a complex combination of basis states in the other basis. For example,

$$\begin{aligned} |\,\swarrow\nearrow\rangle &= \frac{1}{\sqrt{2}}\,|\,\leftrightarrow\rangle + \frac{1}{\sqrt{2}}\,|\,\updownarrow\rangle \\ |\,\nwarrow\searrow\rangle &= \frac{1}{\sqrt{2}}\,|\,\leftrightarrow\rangle - \frac{1}{\sqrt{2}}\,|\,\updownarrow\rangle \end{aligned}$$

$$
\begin{aligned}
|\circlearrowright\rangle &= \frac{1}{\sqrt{2}}|\leftrightarrow\rangle + \frac{i}{\sqrt{2}}|\updownarrow\rangle \\
|\circlearrowleft\rangle &= \frac{1}{\sqrt{2}}|\leftrightarrow\rangle - \frac{i}{\sqrt{2}}|\updownarrow\rangle
\end{aligned}
$$

Here, $|\circlearrowright\rangle$ and $|\circlearrowleft\rangle$ stand for right and left circularly polarized light, respectively; these are another pair of basis states for the polarization of photons. For example, when diagonally polarized photons are put through a birefringent crystal oriented in the $\updownarrow, \leftrightarrow$ direction, half of them will behave like vertically polarized photons, and half like horizontally polarized photons.

If you have two quantum systems, their joint state space is the tensor product of their individual state spaces. For example, the state space of two qubits is $\mathbb{C}^4$ and of three qubits is $\mathbb{C}^8$. The high dimensionality of the space for n qubits, $\mathbb{C}^{2^n}$, is one of the places where quantum computation attains its power.

The polarization state space of two photons has as a basis the four states

$$
|\updownarrow\updownarrow\rangle, \quad |\updownarrow\leftrightarrow\rangle, \quad |\leftrightarrow\updownarrow\rangle, \quad |\leftrightarrow\leftrightarrow\rangle.
$$

This state space includes states such as an EPR (Einstein, Podolsky, Rosen) pair of photons

$$
\frac{1}{\sqrt{2}}(|\updownarrow\leftrightarrow\rangle - |\leftrightarrow\updownarrow\rangle) = \frac{1}{\sqrt{2}}(|\nearrow\nwarrow\rangle - |\nwarrow\nearrow\rangle), \tag{6}
$$

where neither qubit alone has a definite state, but which has a definite state when considered as a joint system of two qubits. In this state, the two photons have orthogonal polarizations in whichever basis they are measured in. Bell [3] showed that the outcomes of measurements on the photons of this state cannot be reproduced by joint probability distributions which give probabilities for the outcomes of all possible measurements, and in which each of the single photons has a definite probability distribution for the outcome of measurements on it, independent of the measurements which are made on the other photon. In other words, there cannot be any set of hidden variables associated with each photon that determines the probability distribution obtained when this photon is measured in any particular basis.

I will present here another demonstration of this impossibility of local hidden variables; namely, the proof involving the *GHZ state* (named for Greenburger, Horne and Zeilinger) [14]. Many fewer people have seen this than have seen Bell's inequalities, probability because it is much more recent; however, the demonstration for the GHZ state is in some ways simpler because it is deterministic. From now on, instead of using $|\updownarrow\rangle$ and $|\leftrightarrow\rangle$ for qubits, we will use $|0\rangle$ and $|1\rangle$, as these are equivalent and probably more familiar to our audience. The GHZ state is

$$
\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle). \tag{7}
$$

The thought experiment demonstrating the impossibility of hidden variables involves measuring each of the qubits in either the C basis $\frac{1}{\sqrt{2}}(|0\rangle \pm i|1\rangle)$ or in the D basis $\frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$. For photon polarization, the C basis corresponds to circularly polarized light and the D basis to diagonally polarized light. We will first suppose that each of

the qubits is measured in the D basis. This projects the joint state of our three qubits onto one of the eight mutually orthogonal vectors

$$\frac{1}{\sqrt{8}}(|\,0\rangle \pm |\,1\rangle)(|\,0\rangle \pm |\,1\rangle)(|\,0\rangle \pm |\,1\rangle). \tag{8}$$

Let us consider the state formed by taking all plus signs in the superpositions above. This is equivalently

$$\frac{1}{\sqrt{8}}(|\,000\rangle + |\,001\rangle + |\,010\rangle + |\,011\rangle + |\,100\rangle + |\,101\rangle + |\,110\rangle + |\,111\rangle) \tag{9}$$

The inner product of this state with the GHZ state (7) is $\frac{1}{2}$, so the probability of observing the state (9) when measuring all three qubits in the D basis is $(\frac{1}{2})^2 = \frac{1}{4}$. It is easy to check that similarly, the probability of observing any of the states of (8) with an even number of $-$'s is $\frac{1}{4}$ and that the probability of observing any state of (8) with an odd number of $-$'s is 0.

We now consider measuring two of the qubits in the C basis and one (say the third) in the D basis. This measurement projects onto the eight basis states

$$\frac{1}{\sqrt{8}}(|\,0\rangle \pm i\,|\,1\rangle)(|\,0\rangle \pm i\,|\,1\rangle)(|\,0\rangle \pm |\,1\rangle). \tag{10}$$

Here, it is easy to check that if we measure the GHZ state (7) in this basis, we will always observe an odd number of $-$'s.

We can now show that it is impossible to assign measurement outcomes to each of the qubits independent of the basis that the other qubits are measured in, and remain consistent with the predictions of quantum mechanics. Consider the following table

qubit 1	qubit 2	qubit 3	parity
D	D	D	even
D	C	C	odd
C	D	C	odd
C	C	D	odd

(11)

The last entry in each row gives the parity of the number of $-$'s if the three qubits are measured in the bases given by the first three entries of the row. Suppose there is a definite outcome assigned to each qubit for each of the two possible measurement bases. Since each basis appears for each qubit exactly twice in the table, the total number of $-$'s in the table would thus have to be even. However, the results predicted by quantum mechanics (the fourth column) are that the total number of $-$'s in the table is odd. This implies that the outcome of at least one measurement on one qubit must depend on the measurements which are made on the other qubits, and that this must hold even if the qubits are spatially separated. It can be shown, however, that this correlation cannot be used to transmit any information between people holding these various qubits; for example, the probability that a qubit is found to be $+$ $(-)$ is one-half independent of the measurements on the other qubits, so which measurements are chosen for the other qubits do not affect this probability (although the outcomes of these measurements may).

The next fundamental principle of quantum mechanics we discuss is the *linearity principle.* This principle states that an isolated quantum system undergoes linear evolution. Because the quantum systems we are considering are finite dimensional vector spaces, a linear evolution of these can be described by multiplication by a matrix. It is fairly easy to check that in order to make the probabilities sum to one, we must restrict these matrices to be unitary (a matrix U is unitary if $U^\dagger = U^{-1}$; unitary matrices are the complex matrices which take unit vectors to unit vectors).

Although many explanations of quantum mechanics restrict themselves to pure states (unit vectors), for quantum information theory we need to treat probability distributions over quantum states. These naturally give rise to objects called density matrices. For an n-dimensional quantum state space, a *density matrix* is an $n \times n$ Hermitian trace-one positive semidefinite matrix.

A rank one density matrix ρ corresponds to the pure state $|\, v\rangle$ where $\rho = |\, v\rangle \langle v\,|$. Recall $\langle v\,|$ was the complex conjugate transpose of $|\, v\rangle$, and for most of this paper we denote $\langle v\,|$ by $v^\dagger$. Density matrices arise naturally from quantum states in two ways.

The first way in which density matrices arise is from probability distributions over quantum states. Suppose that we have a system which is in state v_i with probability p_i. The corresponding density matrix is

$$\rho = \sum_i p_i v_i v_i^\dagger. \tag{12}$$

An important fact about density matrices is that the density matrix for a system gives as much information as possible about experiments performed on the system. That is, any two systems with the same density matrix ρ cannot be distinguished by experiments, provided that no extra side information is given about these systems.

The other way in which density matrices arise is through disregarding part of an entangled quantum state. Recall that two systems in an entangled (pure) state have a definite quantum state when considered jointly, but each of the two systems individually cannot be said to have a definite state. Suppose that we have a pure state ρ_{AB} on a tensor product system $\mathcal{H}_A \otimes \mathcal{H}_B$. If we can only see the first part of the system, this part behaves as though it is in the state $\rho_A = \mathrm{Tr}_B \rho_{AB}$. Here, Tr_B is the partial trace operator. Consider a joint system in the state

$$\rho_{AB} = \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}. \tag{13}$$

In this example, the dimension of $\mathcal{H}_A$ is 3 and the dimension of $\mathcal{H}_B$ is the size of the matrices B_{ij}. The partial trace of ρ_{AB}, tracing over $\mathcal{H}_A$, is

$$\mathrm{Tr}_A\, \rho_{AB} = B_{11} + B_{22} + B_{33} \tag{14}$$

Although the above formula also determines the partial trace when we trace over $\mathcal{H}_B$, through a change of coordinates, it is instructive to give this explicitly:

$$\mathrm{Tr}_B\, \rho_{AB} = \begin{pmatrix} \mathrm{Tr}\, B_{11} & \mathrm{Tr}\, B_{12} & \mathrm{Tr}\, B_{13} \\ \mathrm{Tr}\, B_{21} & \mathrm{Tr}\, B_{22} & \mathrm{Tr}\, B_{23} \\ \mathrm{Tr}\, B_{31} & \mathrm{Tr}\, B_{32} & \mathrm{Tr}\, B_{33} \end{pmatrix}. \tag{15}$$

The final ingredient we need before we can start explaining quantum information theory is a *von Neumann measurement*. We have seen examples of this process before, while explaining the superposition principle and the GHZ non-locality proof; however, we have not yet given the general mathematical formulation of a von Neumann measurement. Suppose that we have an n-dimensional quantum system $\mathcal{H}$. A von Neumann measurement corresponds to a complete set of orthogonal subspaces S_1, S_2, ..., S_k of $\mathcal{H}$. Here, complete means that the subspaces S_i span the space $\mathcal{H}$, so that $\sum_i \dim S_i = n$. Let Π_i be the projection matrix onto the subspace S_i. If we start with a density matrix ρ, the von Neumann measurement corresponding to the set $\{S_i\}$ projects ρ into one of the subspaces S_i. Specifically, it projects ρ onto the i'th subspace with probability $\mathrm{Tr}\,\Pi_i\rho$, the state after the projection being $\Pi_i\rho\Pi_i$, renormalized to be a unit vector. A special case that is often encountered is when the S_i are all one-dimensional, so that $S_i = w_i w_i^\dagger$, and the vectors w_i form an orthogonal basis of $\mathcal{H}$. Then, a vector v is taken to w_i with probability $|w_i^\dagger v|^2$, and a density matrix ρ is taken to w_i with probability $w_i^\dagger \rho w_i$.

4 Von Neumann entropy

We are now ready to consider quantum information theory. We will start by defining the entropy of a quantum system. To give some intuition for this definition, we first consider some special cases. Consider n photons, each being in the state $|\updownarrow\rangle$ or $|\leftrightarrow\rangle$ with probability $\frac{1}{2}$. Any two of these states are completely distinguishable. There are thus 2^n equally probable states of the system, and the entropy is n bits. This is essentially a classical system.

Consider now n photons, each being in the state $|\updownarrow\rangle$ or $|\swarrow\!\!\nearrow\rangle$ with probability $\frac{1}{2}$. These states are not completely distinguishable, so there are effectively considerably less than 2^n states, and the entropy should intuitively be less than n bits.

By thermodynamic arguments involving the increase in entropy associated with the work extracted from a system, von Neumann deduced that the *(von Neumann) entropy* of a quantum system with density matrix ρ should be

$$H_{\mathrm{vN}}(\rho) = -\mathrm{Tr}\rho\log\rho. \tag{16}$$

Recall that ρ is positive semidefinite, so that $-\mathrm{Tr}\rho\log\rho$ is well defined. If ρ is expressed in coordinates in which it is diagonal with eigenvalues λ_i, then in these coordinates $-\rho\log\rho$ is diagonal with eigenvalues $-\lambda_i\log\lambda_i$. We thus see that

$$H_{\mathrm{vN}}(\rho) = H_{\mathrm{Shan}}(\lambda_i), \tag{17}$$

so that the von Neumann entropy of a density matrix is the Shannon entropy of the eigenvalues. (Recall $\mathrm{Tr}\rho = 1$, so that $\sum_i \lambda_i = 1$.) This definition is easily seen to agree with the Shannon entropy in the classical case, where all the states are distinguishable.

5 Source coding

Von Neumann developed the above definition of entropy for thermodynamics. One can ask whether this is also the correct definition of entropy for information theory. We will first give the example of quantum source coding [20, 26], also called *Schumacher compression*, for which we will see that it is indeed the right definition. We consider a memoryless quantum source that at each time step emits the pure state v_i with probability p_i. We would like to encode this signal in as few qubits as possible, and send them to a receiver who will then be able to reconstruct the original state. Naturally, we will not be able to transmit the original state flawlessly. In fact, the receiver cannot even reconstruct the original state perfectly most of the time, which is the situation that is possible in classical communication theory. Unlike classical signals, however, quantum states are not completely distinguishable theoretically, so reconstructing the original state most of the time is too stringent a requirement. What we will require is that the receiver be able to reconstruct a state which is almost completely indistinguishable from the original state nearly all the time. For this we need a measure of indistinguishability; we will use a measure called *fidelity*. Suppose that the original signal is a vector

$$u = v_1 \otimes v_2 \otimes \ldots \otimes v_n.$$

Then the fidelity between the signal u and the output ρ (which is in general a mixed state, i.e. a density matrix, on n qubits) is $F = u^\dagger \rho u$ and the average fidelity is this fidelity F averaged over u. If the output is a pure state v, the fidelity $F = u^\dagger v v^\dagger u = |u^\dagger v|^2$. The fidelity measures the probability of success of a test which determines whether the output is the same as the input.

Before I can continue to sketch the proof of the quantum source coding theorem, I need to review the proof of the classical source coding theorem. Suppose we have a memoryless source, i.e., a source X that at each time step emits the i'th signal type, S_i, with probability p_i, and where the probability distribution for each signal is independent of the previously emitted signals. The idea behind classical source coding is to show that with high probability, the source emits a *typical sequence*, where a sequence of length n is typical if it contains approximately np_i copies of the signal S_i for every i. The number of typical sequences is only $2^{nH(X)+o(n)}$. These can thus be coded in $nH(X) + o(n)$ bits.

The tool that we use to perform Schumacher compression is that of *typical subspaces*. Suppose that we have a density matrix $\rho \in \mathcal{H}$, where $\mathcal{H} = \mathbb{C}^k$, and we take the tensor product of n copies of ρ in the space $\mathcal{H}^n$, i.e., we take $\rho^{\otimes n} \in \mathbb{C}^{nk}$. There is a typical subspace associated with $\rho^{\otimes n}$. Let $\hat{v}_1$, $\hat{v}_2$, ..., $\hat{v}_k$ be the eigenvectors of ρ with associated eigenvalues λ_1, λ_2, ..., λ_k. Since $\mathrm{Tr}\rho = 1$, these λ_i form a probability distribution. Consider typical sequences of the eigenvectors $\hat{v}_i$, where λ_i is the probability of choosing $\hat{v}_i$. A typical sequences can be turned into a quantum state in $\mathcal{H}^{\otimes n}$ by taking the tensor products of its elements. That is, if a typical sequence is $\hat{v}_{i_1}$, $\hat{v}_{i_2}$, ..., $\hat{v}_{i_n}$, the corresponding quantum state is $w = \hat{v}_{i_1} \otimes \hat{v}_{i_2} \otimes \ldots \otimes \hat{v}_{i_n}$. The typical subspace $\mathcal{T}$ is the subspace spanned by typical sequences of the eigenvectors. The subspace $\mathcal{T}$ has dimension equal to the number of typical sequences, or $2^{H_{\mathrm{vN}}(\rho)n+o(n)}$.

We can now explain how to do Schumacher compression. Suppose we wish to

compress a source emitting v_i with probability p_i. Let the typical subspace corresponding to $\rho^{\otimes n}$ be $\mathcal{T}$, where $\rho = \sum_i p_i v_i v_i^\dagger$ is the density matrix for the source, and where we are using a block length n for our compression scheme. We take the vector $u = v_{i_1} \otimes v_{i_2} \otimes \ldots \otimes v_{i_n}$ and make the von Neumann measurement that projects it into either $\mathcal{T}$ or $\mathcal{T}^\perp$. If u is projected onto $\mathcal{T}$, we send the results of this projection; this can be done with $\log \dim \mathcal{T} = nH_{\mathrm{vN}}(\rho) + o(n)$ qubits. If u is projected onto $\mathcal{T}^\perp$, our compression algorithm has failed and we can send anything; this does not degrade the fidelity of our transmission greatly, because this is a low probability event.

Why did this work? The main element of the proof is to show that the probability that we project u onto $\mathcal{T}$ approaches 1 as n goes to ∞. This probability is $u^\dagger \Pi_{\mathcal{T}} u$. If this probability were exactly 1, then u would necessarily be in $\mathcal{T}$, and we would have noiseless compression. If the probability that the state u is projected onto $\mathcal{T}$ is $1-\epsilon$, then $u^\dagger \Pi_{\mathcal{T}} u = 1-\epsilon$, and when u is projected onto $\mathcal{T}$, the fidelity between the original state u and the final state $\Pi_{\mathcal{T}} u$ is thus $|\langle u|\Pi_{\mathcal{T}} u\rangle|^2 = (1-\epsilon)^2$.

Now, recall that if two density matrices are equal, the outcomes of any experiments performed on them have the same probabilities. Thus, the probability that the source v_i with probabilities p_i projects onto the typical subspace is the same as for the source $\hat{v}_i$ with probabilities λ_i, where $\hat{v}_i$ and λ_i are the eigenvalues and eigenvectors of $\rho = \sum_i p_i v_i v_i^\dagger$. We know from the classical theory of typical sequences that $w = \hat{v}_{i_1} \otimes \hat{v}_{i_2} \otimes \ldots \otimes \hat{v}_{i_k}$ is in the typical subspace at least $1-\epsilon$ of the time; because the $\hat{v}_i$ are distinguishable, this is essentially the classical case, and w is in the typical subspace exactly when the sequence of $\hat{v}_i$ is a typical sequence.

6 Accessible information

The next concept is that of *accessible information*. Here, we again have a source emitting state ρ_i with probability p_i. Note that now, the states ρ emitted may be density matrices rather than pure states. We will ask a different question this time. We now want to obtain as much information as possible about the sequence of signals emitted by the source. That is, we wish to maximize the mutual information $I(X;Y)$ where X is the variable telling which signal ρ_i was emitted, and Y is the variable giving the outcome of a measurement on X. This gives the capacity of a channel where at each time step the sender must choose one of the states ρ_i to send, and must furthermore choose ρ_i a fraction p_i of the time; and where the receiver makes a separate measurement on each signal sent.

To find the accessible information, we need to maximize over all measurements. For this, we need to be able to characterize all possible quantum measurements. It turns out that von Neumann measurements are not the most general class of quantum measurements; the most general measurements are the *positive operator valued measurements*, or *POVM's*. One way to describe these is as von Neumann measurements on a quantum space larger than the original space; that is, by supplementing the quantum state space by an *ancilla* space and taking a von Neumann measurement on the joint state space.

For a POVM, we are given a set of positive semidefinite matrices E_i satisfying

$\sum_i E_i = I$. The probability of the i'th outcome is then

$$p_i = \mathrm{Tr}(E_i \rho) \tag{18}$$

For a von Neumann measurement, we take $E_i = \Pi_{S_i}$, the projection matrix onto the i'th orthogonal subspace S_i. The condition $\sum_i \Pi_{S_i} = I$ is equivalent to the requirement that the S_i are orthogonal and span the whole state space. To obtain the maximum information from a POVM, we can assume that the E_i's are pure states; if there is an E_i that is not rank one, then we can always achieve at least as much accessible information by refining that E_i into a sum $E_i = \sum_j E_{ij}$ where the E_{ij} are rank one.

We now give some examples of the measurements maximizing accessible information. The first is one of the simplest examples. Suppose that we have just two pure states in our ensemble, with probability $\frac{1}{2}$ each. For example, we could take the states $|\updownarrow\rangle$ and $|\swarrow\!\!\nearrow\rangle$. Let us take $v_1 = (1, 0)$ and $v_2 = (\cos\theta, \sin\theta)$. We will not prove it here, but the optimal measurement for these is the von Neumann measurement with two orthogonal vectors symmetric around v_1 and v_2. That is, the measurement with projectors

$$w_1 = \left(\cos(\tfrac{\pi}{2} + \tfrac{\theta}{2}), \sin(\tfrac{\pi}{2} + \tfrac{\theta}{2})\right) \tag{19}$$

$$w_2 = \left(\cos(-\tfrac{\pi}{2} + \tfrac{\theta}{2}), \sin(-\tfrac{\pi}{2} + \tfrac{\theta}{2})\right) \tag{20}$$

This measurement is symmetric with respect to interchanging v_0 and v_1, and it leads to a binary symmetric channel with error probability

$$\cos^2\left(\frac{\pi}{2} + \frac{\theta}{2}\right) = \frac{1}{2} - \frac{\sin\theta}{2}. \tag{21}$$

The accessible information is thus $1 - H(\frac{1}{2} - \frac{\sin\theta}{2})$.

For the ensemble containing v_1 and v_2 with probability $\frac{1}{2}$ each, the density matrix is

$$\rho = \frac{1}{2}\begin{pmatrix} 1 + \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & 1 - \cos^2\theta \end{pmatrix}, \tag{22}$$

which has eigenvalues $\frac{1}{2} \pm \cos\theta$, so the von Neumann entropy of the density matrix is $H(\frac{1}{2} - \frac{\cos\theta}{2})$. The values of I_{acc} and H_{vN} are plotted in Figure 1. One can see that the von Neumann entropy is larger than the accessible information.

Note that in our first example, the optimum measurement was a von Neumann measurement. If there are only two states in an ensemble, it has been conjectured that the measurement optimizing accessible information is always a von Neumann measurement, mainly because extensive computer experiments have not found a counterexample [11]. This conjecture has been proven for quantum states in two dimensions [22]. Our next example shows that this conjecture does not hold for ensembles composed of three or more states.

Our second example is three photons with polarizations that differ by 60° each. These are represented by the vectors

$$\begin{aligned} v_0 &= (1, 0) \\ v_1 &= (-\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}) \\ v_2 &= (-\tfrac{1}{2}, -\tfrac{\sqrt{3}}{2}) \end{aligned}$$

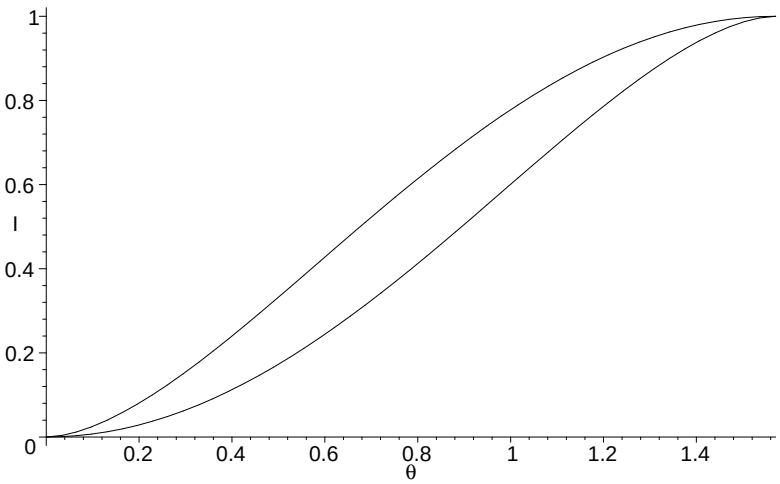

Figure 1: A plot of the von Neumann entropy of the density matrix and the accessible information for the ensemble of two pure quantum states with equal probabilities and that differ by an angle of θ, for $0 \leq \theta \leq \pi/2$. The top curve is the von Neumann entropy and the bottom the accessible information.

The optimal measurement for these states is the POVM corresponding to the vectors w_i where $w_i \perp v_i$. We take $E_i = \frac{2}{3} w_i w_i^\dagger$, in order for $\sum_i E_i = I$. If we start with vector v_i, it is easy to see that we never obtain w_i, but do obtain the other two possible outcomes with probability $\frac{1}{2}$ each. This gives $I_{\rm acc} = \log 3 - 1$. For these three signal states, it is also easy to check that the density matrix $\rho = \frac{1}{2} I$, so $H_{\rm vN} = 1$. Again, we have $I_{\rm acc} < H_{\rm vN}$.

This leads to a conjecture: that $I_{\rm acc} \leq H_{\rm vN}$. The correct theorem is somewhat stronger, and we will shortly state it. The first published proof of this theorem was given by Holevo [17]. It was earlier conjectured by Gordon [12] and stated by Levitin with no proof [21].

Theorem (Holevo): *Suppose that we have a source emitting a (possibly mixed) state ρ_i with probability p_i. Let*

$$\chi = H_{\rm vN}(\sum_i p_i \rho_i) - \sum_i p_i H_{\rm vN}(\rho_i). \tag{23}$$

Then

$$I_{\rm acc} \leq \chi. \tag{24}$$

The conditions for equality in this result are known. If all the ρ_i commute, then they are simultaneously diagonalizable, and the situation is essentially classical. In this case, $I_{\rm acc} = \chi$; otherwise $I_{\rm acc} < \chi$.

7 The classical capacity of a quantum channel

One can ask the question: is this quantity $I_{\rm acc}$ the most information that one can send using the three states of our second example? The answer is, surprisingly, "no". Suppose that we use the three length-two codewords $v_0 \otimes v_0$, $v_1 \otimes v_1$, and $v_2 \otimes v_2$. These are three pure states in the four-dimensional quantum space of two qubits. However, since there are only three vectors, they lie in a three-dimensional subspace. The inner product between any two of these states is $\frac{1}{4}$. One can show that the optimal measurement is attained by the von Neumann measurement having three basis vectors obtained by "pulling" the three vectors $v_i \otimes v_i$ apart until they are all orthogonal. This measurement gives $I_{\rm acc} = 1.369$ bits, which is larger than $2(\log 3 - 1)$ bits = 1.170 bits. In fact, 1.369 bits is larger than twice the maximum accessible information attainable by varying both the probability distribution and the measurement on the three states v_0, v_1 and v_2. This maximum is attained using just two of these states, and is $1 - H(\frac{1}{2} - \frac{\sin(\pi/3)}{2}) = .6454$. We thus find that block coding lets us achieve a better information transmission rate than $I_{\rm acc}$.

Having found that length two codewords work better than length one codewords, the natural question becomes: as the lengths of our codewords go to infinity, how well can we do. The answer is:

Theorem (Holevo[18], Schumacher–Westmoreland[27]): *The classical capacity obtainable using codewords composed of signal states ρ_i, where the probability of using ρ_i is p_i, is*

$$\chi = H_{\rm vN}(\sum_i p_i \rho_i) - \sum_i p_i H_{\rm vN}(\rho_i). \tag{25}$$

We will later give a sketch of the proof of this formula in the special case where the ρ_i are pure states. We will first ask: Does this formula give the capacity of a quantum channel $\mathcal{N}$?

Before we address this question (we will not be able to answer it) we should give the general formulation of a quantum channel. If $\mathcal{N}$ is a memoryless quantum communication channel, then it must take density matrices to density matrices. This means $\mathcal{N}$ must be a trace preserving positive map. Here, trace preserving is required since it must preserve trace 1 matrices, and positive means it takes positive semidefinite matrices to positive semidefinite matrices. For $\mathcal{N}$ to be a valid quantum map, it must have one more property: namely, it must be completely positive. This means that $\mathcal{N}$ is positive even when it is tensored with the identity map. There is a theorem [16] that any such map can be expressed as

$$\mathcal{N}(\rho) = \sum_i A_i \rho A_i^\dagger \tag{26}$$

where A_i are matrices such that $\sum_i A_i^\dagger A_i = I$.

A natural guess at the capacity of a quantum channel $\mathcal{N}$ would be the maximum of χ over all possible distributions of channel outputs, that is,

$$\chi_{\max}(\mathcal{N}) = \max_{\{p_i\},\{\rho_i\}} \chi\Big(\{(\mathcal{N}(\rho_i), p_i)\}\Big), \tag{27}$$

since the sender can effectively communicate to the receiver any of the states $\mathcal{N}(\rho_i)$. We do not know whether this is the capacity of a quantum channel; if the use of entanglement between separate inputs to the channel helps to increase channel capacity, it might be possible to exceed this $\chi_{\max}$. This can be addressed by answering a question that is simple to state: Is $\chi_{\max}$ additive [1]? That is, if we have two quantum channels $\mathcal{N}_1$ and $\mathcal{N}_2$, is

$$\chi_{\max}(\mathcal{N}_1 \otimes \mathcal{N}_2) = \chi_{\max}(\mathcal{N}_1) + \chi_{\max}(\mathcal{N}_2). \tag{28}$$

Proving subadditivity of this quantity is easy. The question is whether strictly more capacity can be attained by using the tensor product of two channels jointly than by using them separately.

We now return to the discussion of the proof of the Holevo-Schumacher-Westmoreland theorem in the special case where the ρ_i are pure states. The proof of this case in fact appeared before the general theorem was proved [15]. The proof uses three ingredients. These are

1. random codes,
2. typical subspaces,
3. the square root measurement.

The square root measurement is also called the "pretty good" measurement, and we have already seen an example of it. Recall our second example for accessible information, where we took the three vectors $v_i \otimes v_i$, where $v_i = (\cos\frac{2\pi i}{3}, \sin\frac{2\pi i}{3})$ for $i = 0, 1, 2$. The optimal measurement for I_{acc} on these vectors was the von Neumann measurement obtained by "pulling" them farther apart until they were orthogonal. This is, in fact, an example of the square root measurement.

Suppose that we are trying to distinguish between vectors u_1, u_2, ..., u_n, which appear with equal probability (the square root measurement can also be defined for vectors having unequal probabilities, but we do not need this case). Let $\phi = \sum_i v_i v_i^\dagger$. The square root measurement has POVM elements $E_i = \phi^{-1/2} v_i v_i^\dagger \phi^{-1/2}$. We have

$$\sum_i E_i = \phi^{-1/2} \left(\sum_i v_i v_i^\dagger \right) \phi^{-1/2} = I, \tag{29}$$

so these E_i do indeed form a POVM.

We can now give the coding algorithm for the capacity theorem for pure states. We choose N codewords $u_j = v_{i_1} \otimes v_{i_2} \otimes \cdots \otimes v_{i_n}$, where the v_i are chosen at random with probability p_i. We then use the codewords u_j to send information; we need to show that each codeword is can be identified with high probability.

To decode, we perform the following steps:

1. Project into the typical subspace $\mathcal{T}$. Most of the time, this projection works, and we obtain $\tilde{u}_j = \Pi_{\mathcal{T}} u_j$, where $\Pi_{\mathcal{T}}$ is the projection matrix onto the subspace $\mathcal{T}$.
2. Use the square root measurement on the $\tilde{u}_j$.

The probability of error is

$$1 - \frac{1}{N}\sum_{j=1}^{N} |\tilde{u}_j \phi^{-1/2} \tilde{u}_j|^2. \tag{30}$$

The intuition for why this procedure works (this intuition is not even close to being rigorous; the proof works along substantially different lines) is that for this probability of error to be small, we need that $\phi^{-1/2}\tilde{u}_j$ is close to $\tilde{u}_j$ for most j. However, the $\tilde{u}_j$ are distributed more or less randomly in the typical subspace $\mathcal{T}$, so $\phi = \sum_j \tilde{u}_j \tilde{u}_j^\dagger$ is moderately close to the identity matrix on its support, and thus $\phi^{-1/2}\tilde{u}_j$ is close to $\tilde{u}_j$. Note that we need that the number N of u_j is less than $\dim \mathcal{T}$, or otherwise it would be impossible to distinguish the $\tilde{u}_j$; by Holevo's bound (24) a d-dimensional quantum state space can carry at most d bits of information.

8 Quantum teleportation and superdense coding

In this section, we will first describe *quantum teleportation*, a surprising phenomenon which is an unusual means of transmitting a quantum state. It is impossible to send a quantum state over a classical channel. Quantum teleportation lets a sender and a receiver who share an EPR pair of qubits send two classical bits and use this EPR pair in order to communicate one qubit [5]. (See Figure 2.)

To perform teleportation, the sender starts with a qubit in an unknown state, which we take to be $\alpha\,|\,0\rangle + \beta\,|\,1\rangle$, and a shared EPR pair, which we assume is in the state $\frac{1}{\sqrt{2}}(|\,00\rangle + |\,11\rangle)$, with the sender holding the first qubit of the EPR pair and the receiver holding the second qubit. The joint system is thus in the tensor product of these two states, which is

$$\frac{1}{\sqrt{2}}(\alpha\,|\,0\rangle + \beta\,|\,1\rangle)(|\,00\rangle + |\,11\rangle). \tag{31}$$

Note that the sender has posession of the first two qubits, and the receiver posession of the third one. Using the distributive law, we can rewrite the above state (31 as

$$\frac{1}{\sqrt{8}}\Big[\begin{array}{ll} (|\,00\rangle + |\,11\rangle) & (\alpha\,|\,0\rangle + \beta\,|\,1\rangle) \\ +(|\,00\rangle - |\,11\rangle) & (\alpha\,|\,0\rangle - \beta\,|\,1\rangle) \\ +(|\,10\rangle + |\,01\rangle) & (\beta\,|\,0\rangle + \alpha\,|\,1\rangle) \\ +(|\,10\rangle - |\,01\rangle) & (\beta\,|\,0\rangle - \alpha\,|\,1\rangle)\end{array}\Big] \tag{32}$$

The sender can now perform the von Neumann measurement that projects the state onto one of the four lines of Eq. (32), as the four states

$$\tfrac{1}{\sqrt{2}}(|\,00\rangle \pm |\,11\rangle), \quad \tfrac{1}{\sqrt{2}}(|\,10\rangle \pm |\,01\rangle)$$

are all orthogonal. This leaves the receiver with one of the four states

$$\alpha\,|\,0\rangle \pm \beta\,|\,1\rangle\,, \quad \beta\,|\,0\rangle \pm \alpha\,|\,1\rangle\,,$$

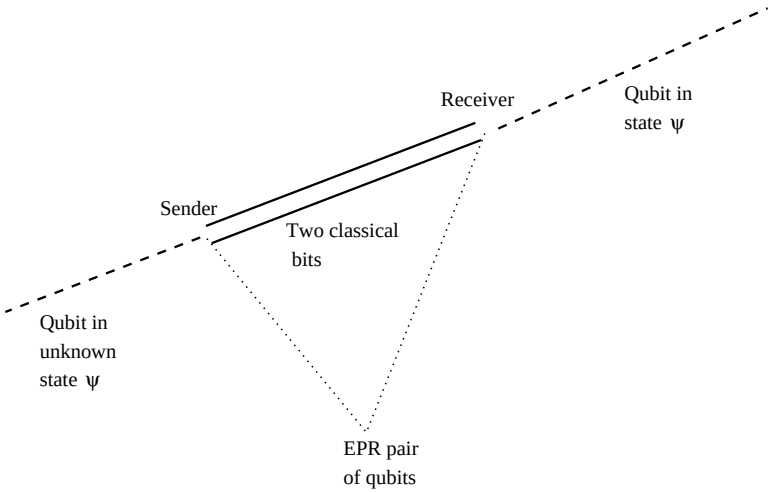

Figure 2: A schematic drawing of quantum teleportation. The sender has a qubit in an unknown state ψ that he wishes to send to the receiver. He also has half of an EPR state which he shares with the receiver. The sender makes a joint measurement on the unknown qubit and half of his EPR state, and communicates the results (2 classical bits) to the receiver. The receiver then makes one of four unitary transformations (depending on the two classical bits he received) on his half of the EPR state to obtain the state ψ.

all of which can be transformed into $\alpha \mid 0\rangle + \beta \mid 1\rangle$ by the appropriate unitary transform. The sender needs to communicate to the receiver which of the four measurement outcomes was obtained (using two bits), and the receiver can then perform the appropriate unitary transform to obtain the original quantum state.

Quantum teleportation is a counterintuitive process, which at first sight seems to violate certain laws of physics; however, upon closer inspection one discovers that no actual paradoxes arise from teleportation. Teleportation cannot be used for superluminal communication, because the classical bits must travel at or slower than the speed of light. While a continuous quantum state appears to have been transported using two discrete bits, by Holevo's bound (24) one qubit can be used to transport at most one classical bit of information, so it is not possible to increase the capacity of a classical channel by encoding information in the teleported qubit. Finally, there is a theorem of quantum mechanics that an unknown quantum state cannot be duplicated [30]. However, the original state is necessarily destroyed by the measurement, teleportation cannot be used to clone a quantum state.

There is a converse process to teleportation, *superdense coding*, which uses a shared EPR pair and a single qubit to encode two classical bits [8]. In this protocol, the sender and receiver use the same operations as teleportation, but reverse their roles; the sender performs the unitary transformation and the receiver performs the measurement. (See Figure 3.)

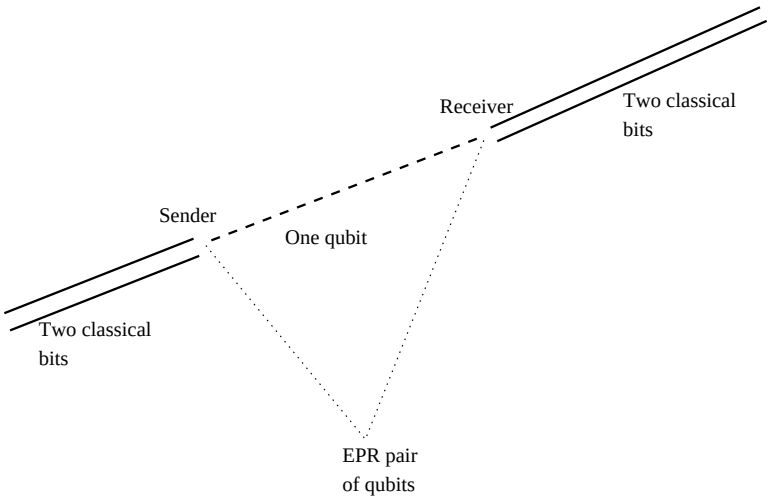

Figure 3: A schematic drawing of superdense coding. The sender can communicate two classical bits to the receiver using one qubit and a shared EPR pair. Here, the sender makes the same unitary transformation that the receiver would make in quantum teleportation, and the receiver makes the joint measurement that the sender would make in quantum teleportation.

9 Other results from quantum information theory

In this final section, I briefly survey some other results of quantum information theory which were unjustly neglected by the previous sections of this paper.

Using teleportation, the sender can send the receiver qubits over a classical channel if they possess shared EPR pairs. Thus, *shared EPR pairs* (an instance of quantum entanglement) can be seen as a resource that lets these two parties send quantum information over a classical channel, a task that would otherwise be impossible. This leads to the question: how do you quantify entanglement? If two parties have n copies of an entangled state ρ, how many EPR pairs does this let them share? We will let the two parties use classical communication and perform local quantum operations on their own states, but no quantum communication and no quantum operations on the joint state space will be allowed.

If ρ is a pure state, then the answer is known and quite nice [4]. Let the two parties' quantum state spaces be A and B. Then if $\rho \in A \otimes B$ is a pure state, n copies of ρ can be made into

$$nH_{\text{vN}}(\text{Tr}_A \rho) + o(n) = nH_{\text{vN}}(\text{Tr}_B \rho) + o(n) \tag{33}$$

nearly perfect EPR pairs, and vice versa, where the fidelity of the actual state with the desired state goes to 1 as the block length n goes to infinity.

If ρ is not a pure state, the situation becomes much more complicated. In this case, we can define entanglement of formation (E_F), which is asympotitically the number of EPR pairs that we need to form ρ; and distillable entanglement (E_D), which is

asymptotically the number of EPR pairs which can be created from ρ. If ρ is pure, then these two quantities are equal, but this does not appear to be true if ρ is mixed.

Much like the classical capacity of a quantum channel, there is a nice expression which would be equal to the entanglement of formation if it could be proved to be additive. We call it the one-shot entanglement of formation, and it is the minimum average entanglement over ensembles of pure states whose density matrix is ρ. That is,

$$E_{F,1}(\rho) = \min_{\sum_i p_i \rho_i = \rho} \sum_i p_i H_{\mathrm{vN}}(\mathrm{Tr}_A \rho_i). \tag{34}$$

We now give another capacity for quantum channels, one which has a capacity formula which can actually be proven. Suppose that we have a quantum channel $\mathcal{N}$. Recall that if $\mathcal{N}$ is a noiseless quantum channel, and if the sender and receiver possess shared EPR pairs, they can use superdense coding to double the classical information capacity of $\mathcal{N}$. If $\mathcal{N}$ is a noisy quantum channel, using shared EPR pairs can also increase the classical capacity of $\mathcal{N}$. We define the entanglement assisted capacity, C_E, as the quantity of classical information that can asymptotically be sent per channel use if the sender and receiver have access to a sufficient quantity of shared entanglement.

Theorem (Bennett, Shor, Smolin Thapliyal [6, 7]): *The entanglement assisted capacity is*

$$C_E(\mathcal{N}) = \max_{\rho \in \mathcal{H} \otimes \mathcal{S}} H_{vN}(\mathrm{Tr}_{\mathcal{R}}(\mathcal{N} \otimes \mathcal{I})\rho) + H_{vN}(\mathrm{Tr}_{\mathcal{S}}(\mathcal{N} \otimes \mathcal{I})\rho) - H_{vN}((\mathcal{N} \otimes \mathcal{I})\rho) \tag{35}$$

where $\mathcal{R}$ and $\mathcal{S}$ stand for receiver and sender, respectively. Here ρ is maximized over pure states on the tensor product of the input state space $\mathcal{H} = \mathbb{C}^d$ of the channel and a quantum space $\mathcal{S}$ (which may be assumed also to be of dimension d) that the sender keeps.

The quantity being minimized in the above formula (35) is called quantum mutual information, and it is a generalization of the expression for mutual information in the form of Eq. (4). The proof of this result uses typical subspaces, superdense coding, the Holevo-Schumacher-Westmoreland theorem on the classical capacity of a quantum channel, and the strong subadditivity property of von Neumann entropy.

Finally, we briefly mention the problem of sending quantum information (i.e., a quantum state) over a noisy quantum channel. In this scenario, several of the theorems that make classical channel capacity behave so nicely are not true. Here, a back channel from the receiver to the sender increases the quantum channel capacity, leading to two quantum capacities, Q_2 where the receiver has a classical back channel from himself to the sender, and $Q \leq Q_2$, where all communication is from the sender to the receiver over the noisy quantum channel $\mathcal{N}$. There is a conjectured capacity formula for Q. It is essentially the last two terms of the expression (35) for entanglement-assisted capacity

$$Q(\mathcal{N}) = \lim_{n \to \infty} \max_{\rho \in (\mathcal{H} \otimes \mathcal{S})^n} H_{vN}(\mathrm{Tr}_{\mathcal{S}}(\mathcal{N} \otimes \mathcal{I})\rho) - H_{vN}((\mathcal{N} \otimes \mathcal{I})\rho) \tag{36}$$

where ρ, $\mathcal{H}$ and $\mathcal{S}$ are defined as in (35). The quantity being maximized is called the *coherent information*. We now need to take the maximum over the tensor product of n uses of the channel, and let n go to infinity, because unlike the classical (or the quantum) mutual information, the coherent information is not additive [10]. The quantity

(36) is an upper bound for the quantum capacity of a noisy quantum channel $\mathcal{N}$ [2], and is conjectured to be equal to this capacity [19].

There are many more results in quantum information theory, including several large areas that I have not discussed at all. I have not mentioned *quantum error-correcting codes*, which are the tools one needs to send quantum information over a noisy channel [13]. I have also not mentioned quantum cryptography, in connection with which there exist several recent security proofs [9, 23, 24, 29], and associated results on tradeoffs between disturbing a quantum state and extracting information from it. Finally, I have not mentioned a large literature on entangled quantum states shared among more than two parties. I hope that this paper stimulates some readers to learn more about quantum information theory.

References

[1] G. G. Amosov, A. S. Holevo, R. F. Werner, "On some additivity problems in quantum information theory," LANL e-print math-ph/0003002, available at `http://xxx.lanl.gov`.

[2] H. Barnum, J. A. Smolin and B. M. Terhal, "Quantum capacity is properly defined without encodings," *Phys. Rev. A,* vol. 58, pp. 3496–3501, 1998.

[3] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics,* vol. 1, pp. 195–200, 1964.

[4] C. H. Bennett, H. J. Bernstein, S. Popescu and B. Schumacher, "Concentrating partial entanglement by local operations," *Phys. Rev. A,* vol. 53, pp. 2046–2052, 1996.

[5] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Phys. Rev. Lett.,* vol. 70, pp. 1895–1899, 1993.

[6] C. H. Bennett, P. W. Shor, J. A. Smolin and A. V. Thapliyal, "Entanglement-assisted classical capacity of noisy quantum channels," *Phys. Rev. Lett.,* vol. 83, pp. 3081–3084, 1999.

[7] C. H. Bennett, P. W. Shor, J. A. Smolin and A. V. Thapliyal, manuscript in preparation.

[8] C. H. Bennett and S. J. Wiesner, "Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states," *Phys. Rev. Lett.,* vol. 69, pp. 2881–2884, 1992.

[9] E. Biham, M. Boyer, P. O. Boykin, T. Mor and V. Roychowdhury, "A proof of the security of quantum key distribution," in *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing,* (ACM Press, New York, 2000), pp. 715–724; longer version LANL e-print quant-ph/9912053, available at `http://xxx.lanl.gov`.

[10] D. DiVincenzo, J. A. Smolin and P. W. Shor, "Quantum-channel capacity of very noisy channels, *Phys. Rev. A,* vol. 57, pp. 830–839, 1998.

[11] C. A. Fuchs and A. Peres, personal communication.

[12] J. P. Gordon, "Noise at optical frequencies; information theory," in *Quantum Electronics and Coherent Light; Proceedings of the International School of Physics*

Enrico Fermi, Course XXXI, (P. A. Miles, ed., Academic Press New York, 1964) pp. 156–181.

[13] D. Gottesman, "An introduction to quantum error correction," LANL e-print quant-ph/0004072, available at `http://xxx.lanl.gov`.

[14] D. M. Greenburger, M. A. Horne, A. Shimony and A. Zeilinger, "Bell's theorem without inequalities," *Am. J. Phys.*, vol. 58, pp. 1131-1143, 1990.

[15] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, "Classical information capacity of a quantum channel," *Phys. Rev. A*, vol. 54, pp. 1869–1876, 1996.

[16] K. Hellwig and K. Krauss, "Operations and measurements II," *Communications in Mathematical Physics*, vol. 16, pp. 142-147, 1970.

[17] A. S. Holevo, "Information theoretical aspects of quantum measurements," *Probl. Info. Transm. (USSR)*, vol. 9, no. 2, pp. 31–42, 1973 (in Russian); [translation: A. S. Kholevo, *Probl. Info. Transm.*, vol. 9, pp. 177–183, 1973].

[18] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Info. Theory*, vol. 44, pp. 269–273, 1998.

[19] M. Horodecki, P. Horodecki and R. Horodecki, "Unified approach to quantum capacities: Towards a quantum noisy coding theorem," LANL e-print quant-ph/0003040, available at `http://xxx.lanl.gov`.

[20] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *J. Modern Optics*, vol. 41, pp. 2343-2349, 1994.

[21] L. B. Levitin, "On the quantum measure of information," in *Proceedings of the Fourth All-Union Conference on Information and Coding Theory, Sec. II*, Tashkent, 1969.

[22] L. B. Levitin, "Optimal quantum measurements for pure and mixed states," in *Quantum Communications and Measurement*, (V. P. Belavkin, O. Hirota and R. L. Hudson, eds., Plenum Press, New York and London, 1995) pp. 439–448.

[23] H.-K. Lo and H. F. Chau, "Unconditional security of quantum key distribution over arbitrarily long distances," *Sciehnce*, vol. 283, pp. 2050–2056, 1999.

[24] D. Mayers, "Unconditional security in quantum cryptography," *J. ACM*, to appear, also LANL e-print quant-ph/9802025, available at `http://xxx.lanl.gov`.

[25] J. R. Pierce, "The early days of information theory," *IEEE Trans. Info. Theory*, vol. 19, pp. 3–8, 1973.

[26] B. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, pp. 2738–2747, 1995.

[27] B. Schumacher and Westmoreland, "Sending classical information via a noisy quantum channel," *Phys. Rev. A*, vol. 56, pp. 131–138, 1997.

[28] C. E. Shannon, "A mathematical theory of communication," *The Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[29] P. W. Shor and J. A. Preskill, "Simple proof of security of the BB84 quantum key distribution protocol," *Phys. Rev. Lett*, vol. 85, pp. 441-444, 2000.

[30] W. K. Wootters and W. H. Zurek, "A single quantum cannot be cloned," *Nature*, vol. 299, pp. 802–803, 1982.