Quantum Information Theory

Charles H. Bennett and Peter W. Shor

(Invited Paper)

Abstract—**We survey the field of quantum information theory. In particular, we discuss the fundamentals of the field, source coding, quantum error-correcting codes, capacities of quantum channels, measures of entanglement, and quantum cryptography.**

Index Terms— **Entanglement, quantum cryptography, quantum error-correcting codes, quantum information, quantum source coding.**

I. Introduction

RECENTLY, the historic connection between information and physics has been revitalized, as the methods of information and computation theory have been extended to treat the transmission and processing of intact quantum states, and the interaction of such "quantum information" with traditional "classical" information. Although many of the quantum results are similar to their classical analogs, there are notable differences. This new research has the potential to shed light both on quantum physics and on classical information theory.

In retrospect, this development seems somewhat belated, since quantum mechanics has long been thought to underlie all classical processes. But until recently, information itself had largely been thought of in classical terms, with quantum mechanics playing a supporting role of helping design the equipment used to process it, setting limits on the rate at which it could be sent through certain quantum channels. Now we know that a fully quantum theory of information and information processing offers, among other benefits, a brand of cryptography whose security rests on fundamental physics, and a reasonable hope of constructing quantum computers that could dramatically speed up the solution of certain mathematical problems. These feats depend on distinctively quantum properties such as uncertainty, interference, and entanglement.

At a more fundamental level, it has become clear that an information theory based on quantum principles extends and completes classical information theory, somewhat as complex numbers extend and complete the reals. The new theory includes quantum generalizations of classical notions such as sources, channels, and codes, as well as two complementary, quantifiable kinds of information—classical information and quantum entanglement. Classical information can be copied freely, but can only be transmitted forward in time, to a receiver in the sender's forward light cone. Entanglement, by contrast, cannot be copied, but can connect any two points in space–time. Conventional data-processing operations destroy entanglement, but quantum operations can create it, preserve it, and use it for various purposes, notably speeding up certain computations and assisting in the transmission of classical data ("quantum superdense coding") or intact quantum states ("quantum teleportation") from a sender to a receiver.

Any means, such as an optical fiber, for delivering quantum systems more or less intact from one place to another, may be viewed as a quantum channel. Unlike classical channels, such channels have three distinct capacities: a capacity C for transmitting classical data, a typically lower capacity Q for transmitting intact quantum states, and a third capacity Q_2, often between C and Q, for transmitting intact quantum states with the assistance of a two-way classical side-channel between sender and receiver.

How, then, does quantum information, and the operations that can be performed on it, differ from conventional digital data and data-processing operations? A classical bit (e.g., a memory element or wire carrying a binary signal) is generally a system containing many atoms, and is described by one or more continuous parameters such as voltages. Within this parameter space two well-separated regions are chosen by the designer to represent 0 and 1, and signals are periodically restored toward these standard regions to prevent them from drifting away due to environmental perturbations, manufacturing defects, etc. An n-bit memory can exist in any of 2^n logical states, labeled 000..0 to 111..1. Besides storing binary data, classical computers manipulate it, a sequence of Boolean operations (for example, NOT and AND) acting on the bits one or two at a time being sufficient to realize any deterministic transformation.

A quantum bit, or "qubit," by contrast is typically a microscopic system, such as an atom or nuclear spin or polarized photon. The Boolean states 0 and 1 are represented by a fixed pair of reliably distinguishable states of the qubit (for example, horizontal and vertical polarizations: $|0\rangle = \leftrightarrow$, $|1\rangle = \updownarrow$). A qubit can also exist in a continuum of intermediate states, or "superpositions," represented mathematically as unit vectors in a two-dimensional complex vector space (the "Hilbert space" $\mathcal{H}_2$) spanned by the basis vectors $|0\rangle$ and $|1\rangle$. For photons, these intermediate states correspond to other polarizations, for example,

$$⤢ = \sqrt{\tfrac{1}{2}}\,(|0\rangle + |1\rangle)$$
$$⤡ = \sqrt{\tfrac{1}{2}}\,(|0\rangle - |1\rangle)$$

Manuscript received January 4, 1998; revised June 2, 1998.

C. H. Bennett is with the IBM Research Division, T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: bennetc@watson.ibm.com).

P. W. Shor is with the AT&T Labs–Research, Florham Park, NJ 07932 USA (e-mail: shor@research.att.com).

Publisher Item Identifier S 0018-9448(98)06316-0.

and

$$\curvearrowright = \sqrt{\frac{1}{2}}\,(|0\rangle + i|1\rangle)$$

(right circular polarization). Unlike the intermediate states of a classical bit (e.g., voltages between the standard 0 and 1 values), these intermediate states cannot be reliably distinguished, even in principle, from the basis states. With regard to any measurement which distinguishes the states $|0\rangle$ and $|1\rangle$, the superposition $|\psi\rangle = \psi_0|0\rangle + \psi_1|1\rangle$ behaves like $|0\rangle$ with probability $|\psi_0|^2$ and like $|1\rangle$ with probability $|\psi_1|^2$. More generally, two quantum states are reliably distinguishable if and only if their vector representations are orthogonal; thus ↔ and ↕ are reliably distinguishable by one type of measurement, and ⤢ and ⤡ by another, but no measurement can reliably distinguish ↔ from ⤢. Multiplying a state vector by an arbitrary phase factor $e^{i\theta}$ does not change its physical significance: thus although they are usually represented by unit vectors, quantum states are more properly identified with *rays*, a ray being the equivalence class of a vector under multiplication by a complex constant.

It is convenient to use the so-called bracket or bra-ket notation, in which the inner product between two d-dimensional vectors $|\psi\rangle$ and $|\phi\rangle$ is denoted

$$\langle\psi|\phi\rangle = \sum_{x=1}^{d} \psi_x^* \phi_x$$

where the asterisk denotes complex conjugation. This may be thought of as matrix product of the row vector $\langle\psi| = (\psi_1^*, \cdots, \psi_d^*)$, by a column vector

$$|\phi\rangle = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_d \end{pmatrix}$$

where for any standard column (or "ket") vector $|\psi\rangle$, its row (or "bra") representation $\langle\psi|$ is obtained by transposing and taking the complex conjugate.

A pair of qubits (for example, two photons in different locations) is capable of existing in four basis states, $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$, as well as all possible superpositions of them. States of a pair of qubits thus lie in a four-dimensional Hilbert space. This space contains states like

$$\sqrt{\frac{1}{2}}\,(|00\rangle + |01\rangle) = \sqrt{\frac{1}{2}}\,|0\rangle(|0\rangle + |1\rangle) = \text{↔ ⤢}$$

which can be interpreted in terms of individual polarizations for the two photons, as well as "entangled" states, i.e., states like

$$\sqrt{\frac{1}{2}}\,(|00\rangle + |11\rangle)$$

in which neither photon by itself has a definite state, even though the pair together does.

More generally, where a string of n classical bits could exist in any of 2^n Boolean states $x = 000\cdots0$ through $111\cdots1$, a string of n qubits can exist in any state of the form

$$\Psi = \sum_{x=00\cdots0}^{11\cdots1} \psi_x |x\rangle \tag{1}$$

where the ψ_x are complex numbers such that $\sum_x |\psi_x|^2 = 1$. In other words, a quantum state of n qubits is represented by a complex unit vector Ψ (more properly a ray, since multiplying Ψ by a phase factor does not change its physical meaning) in the 2^n-dimensional Hilbert space $\mathcal{H}^{2^n} = (\mathcal{H}_2)^n$, defined as the tensor product of n copies of the two-dimensional Hilbert space representing quantum states of a single qubit. The exponentially large dimensionality of this space distinguishes quantum computers from classical analog computers, whose state is described by a number of parameters that grows only linearly with the size of the system. This is because classical systems, whether digital or analog, can be completely described by separately describing the state of each part. The vast majority of quantum states, by contrast, are entangled, and admit no such description. The ability to preserve and manipulate entangled states is the distinguishing feature of quantum computers, responsible both for their power and for the difficulty of building them.

An isolated quantum system evolves in such a way as to preserve superpositions and distinguishability; mathematically, such evolution is a unitary (i.e., linear and inner-product-conserving) transformation, the Hilbert-space analog of rigid rotation in Euclidean space. Unitary evolution and superposition are the central principles of quantum mechanics.

Just as any classical computation can be expressed as a sequence of one- and two-bit operations (e.g., NOT and AND gates), any quantum computation can be expressed as a sequence of one- and two-qubit quantum gates, i.e., unitary operations acting on one or two qubits at a time [1] (cf. Fig. 1). The most general one-qubit gate is described by a 2×2 unitary matrix[1] $\left(\begin{smallmatrix}\alpha & \gamma \\ \beta & \delta\end{smallmatrix}\right)$ mapping $|0\rangle$ to $\alpha|0\rangle + \beta|1\rangle$ and $|1\rangle$ to $\gamma|0\rangle + \delta|1\rangle$. One-qubit gates are easily implementable physically, e.g., by quarter- and half-wave plates acting on polarized photons, or by radio-frequency tipping pulses acting on nuclear spins in a magnetic field.

The standard two-qubit gate is the controlled-NOT or XOR gate, which flips its second (or "target") input if its first ("control") input is $|1\rangle$ and does nothing if the first input is $|0\rangle$. In other words, it interchanges $|10\rangle$ and $|11\rangle$ while leaving $|00\rangle$ and $|01\rangle$ unchanged. The XOR gate is represented by the 4×4 unitary matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Unlike one-qubit gates, two-qubit gates are difficult to realize in the laboratory, because they require two separate quantum information carriers to be brought into strong and controlled interaction. The XOR gate, together with the set of one-bit gates, form a universal set of primitives for quantum computation; that is, any quantum computation can be performed using just this set of gates without an undue increase in the number of gates used [1].

[1] A complex matrix is called unitary and represents a unitary transformation, iff its rows are orthogonal unit vectors. The inverse of any unitary matrix U is given by its *adjoint*, or conjugate transpose $U^\dagger$.

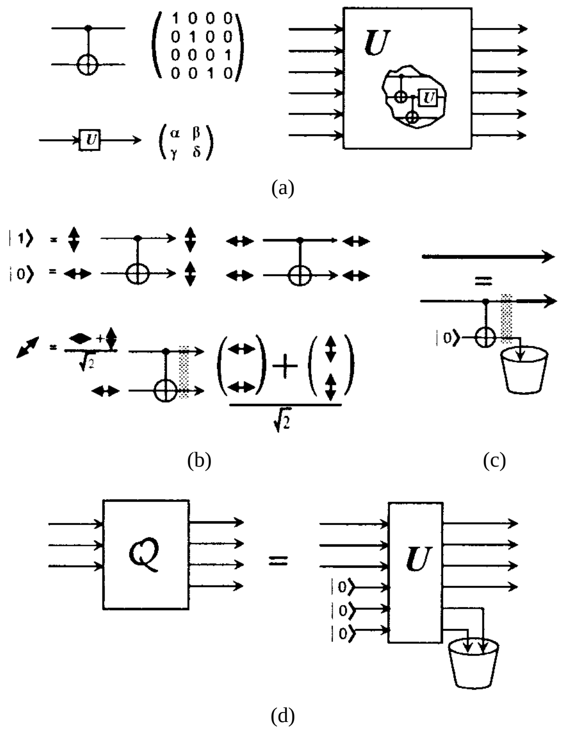


Fig. 1. (a) Any unitary operation U on quantum data can be synthesized from the two-qubit XOR gate and one-qubit unitary operations (U). (b) The XOR can clone Boolean-valued inputs, but if one attempts to clone intermediate superposition, an entangled state (shading) results instead. (c) A classical wire (thick line) conducts 0 and 1 faithfully but not superpositions or entangled states. It may be defined as a quantum wire that interacts (via an XOR) with an ancillary 0 qubit which is then discarded. (d) The most general treatment, or superoperator, that can be applied to quantum data is a unitary interaction with one or more 0 qubits, followed by discarding some of the qubits. Superoperators are typically irreversible.

The XOR gate is a prototype interaction between two quantum systems, and illustrates several key features of quantum information, in particular the impossibility of "cloning" an unknown quantum state, and the way interaction produces entanglement. If the XOR is applied to Boolean data in which the second qubit is 0 and the first is 0 or 1 (cf. Fig. 1(b)) the effect is to leave the first qubit unchanged while the second becomes a copy of it: $U_{\text{XOR}}|x, 0\rangle = |x, x\rangle$ for $x = 0$ or 1. One might suppose that the XOR operation could also be used to copy superpositions, such as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, so that $U_{\text{XOR}}|\psi, 0\rangle$ would yield $|\psi, \psi\rangle$, but this is not so. The unitarity of quantum evolution requires that a superposition of input states evolve to a corresponding superposition of outputs. Thus the result of applying U_{XOR} to $|\psi, 0\rangle$ must be $\alpha|0, 0\rangle + \beta|1, 1\rangle$, an entangled state in which neither output qubit alone has definite state. If one of the entangled output qubits is lost (e.g., discarded, or allowed to escape into the environment), the other thenceforth behaves as if it had acquired a random classical value 0 (with probability $|\alpha|^2$) or 1 (with probability $|\beta|^2$). Unless the lost output is brought back into play, all record of the original superposition $|\psi\rangle$ will have been lost. This behavior is characteristic not only of the XOR gate but of unitary interactions generally: their typical effect is to map most unentangled initial states of the interacting systems into entangled final states, which from the viewpoint of either system alone causes an unpredictable disturbance.

Since quantum physics underlies classical, there should be a way to represent classical data and operations within the quantum formalism. If a classical bit is a qubit having the value $|0\rangle$ or $|1\rangle$, a classical wire should be a wire that conducts $|0\rangle$ and $|1\rangle$ reliably, but not superpositions. This can be implemented using the XOR gate as described above, with an initial $|0\rangle$ in the target position which is later discarded (Fig. 1(c)). In other words, from the viewpoint of quantum information, classical communication is an irreversible process in which the signal interacts enroute with an environment or eavesdropper in such a way that Boolean signals pass through undisturbed, but other states become entangled with the environment. If the environment is lost or discarded, the surviving signal behaves as if it had irreversibly collapsed onto one of the Boolean states. Having defined a classical wire, we can then go on to define a classical gate as a quantum gate with classical wires on its inputs and outputs. The classical wire of Fig. 1(c) is an example of quantum information processing in an open system. Any processing that can be applied to quantum data, including unitary processing as a special case, can be described (Fig. 1(d)) as a unitary interaction of the quantum data with some *ancillary qubits*, initially in a standard $|0\rangle$ state, followed by discarding some of the qubits. Such a general quantum data processing operation (also called a trace-preserving completely positive map or superoperator [47], [64]) can therefore have an output Hilbert space larger or smaller than its input space (for unitary operations, the input and output Hilbert spaces are, of course, equidimensional).

Paradoxically, entangling interactions with the environment are thought to be the main reason why the macroscopic world seems to behave classically and not quantum-mechanically [74]. Macroscopically different states, e.g., the different charge states representing 0 and 1 in a VLSI memory cell, interact so strongly with their environment that information rapidly leaks out as to which state the cell is in. Therefore, even if it were possible to prepare the cell in superposition of 0 and 1, the superposition would rapidly evolve into a complex entangled state involving the environment, which from the viewpoint of the memory cell would appear as a statistical mixture, rather than a superposition, of the two classical values. This spontaneous decay of superpositions into mixtures is called decoherence.

The quantum states we have been talking about so far, identified with rays in Hilbert space, are called *pure states*. They represent situations of minimal ignorance, in which, in principle, there is nothing more to be known about the system. Pure states are fundamental in that the quantum mechanics of a closed system can be completely described as a unitary evolution of pure states in an appropriately dimensioned Hilbert space, without need of further notions. However, a very useful notion, the *mixed state* has been introduced to deal with situations of greater ignorance, in particular

- an ensemble $\mathcal{E}$ in which the system in question may be in any of several pure states $|\psi_1\rangle$, $|\psi_2\rangle \cdots$, with specified probabilities $p_1, p_2 \cdots$;
- a situation in which the system in question (call it A) is part of larger system (call it AB), which itself is in an entangled pure state Ψ.

In open systems, a pure state may naturally evolve into a mixed state (which can also be described as a pure state of a larger system comprising the original system and its environment). Mathematically, a mixed state is represented by a positive-semidefinite, self-adjoint *density matrix* ρ, having unit trace, and being defined in the first situation by

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \tag{2}$$

and in the second situation by

$$\rho = \mathrm{Tr}_B |\Psi\rangle\langle\Psi|. \tag{3}$$

Here Tr_B denotes a partial trace over the indices of the B subsystem. A pure state ψ is represented in the density-matrix formalism by the rank-one projection matrix $|\psi\rangle\langle\psi|$.

It is evident, in the first situation, that infinitely many different ensembles can give rise to the same density matrix. For example, the density matrix

$$\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

may be viewed as an equal mixture of the pure states $|0\rangle$ and $|1\rangle$, or as an equal mixture of $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle - |1\rangle)/\sqrt{2}$, or indeed as any other equal mixture of two orthogonal single-qubit pure states. Similarly, in the second situation, it is evident that infinitely many different pure states Ψ of the AB system can give rise to the same density matrix ρ for the A subsystem. One may therefore wonder in what sense a density matrix is an adequate description of a statistical ensemble of pure states, or of part of a larger system in a pure entangled state. The answer is that the density matrix ρ captures all and only that information that can be obtained by an observer allowed to examine infinitely many states sampled from the ensemble $\mathcal{E}$, or given infinitely many opportunities to examine part A of an AB system prepared in entangled pure state Ψ. This follows from the elementary fact that for any test vector ϕ, if a specimen drawn from ensemble $\mathcal{E} = \{\psi_i, p_i\}$ is tested for whether it is in state ϕ, the probability of a positive outcome is

$$\sum_i p_i |\langle\psi|\phi\rangle|^2 = \mathrm{Tr}\,(\rho|\phi\rangle\langle\phi|).$$

Similarly, for any test state ϕ of the A subsystem, the probability that an entangled state of the AB system having ρ as its partial trace will give a positive outcome is simply $\mathrm{Tr}\,(\rho|\phi\rangle\langle\phi|)$.

Perhaps more remarkable than the indistinguishability of the different ensembles compatible with ρ is the fact that any of them can be produced at will starting from any entangled state Ψ of the AB system having ρ as its partial trace. More specifically, if two parties (call them Alice and Bob) are in possession of the A and B parts, respectively, of a system in state Ψ, then for each compatible ensemble $\mathcal{E} = \{\psi_i, p_i\}$ that Bob might wish to create in Alice's hands, there is a measurement M he can perform on the B subsystem alone, without Alice's knowledge or cooperation, that will realize that ensemble in the sense that the measurement yields outcome i with probability p_i, and conditionally on that outcome having occurred, Bob will know that Alice holds pure state ψ_i. Bob's ability to decide Alice's ensemble in this unilateral, *post facto* fashion, has an important bearing on quantum cryptography as will be discussed later.

Since a mixed state represents incomplete information, it is natural to associate with any mixed state an entropy, given by the von Neumann formula

$$S(\rho) = -\mathrm{Tr}\,\rho \log \rho. \tag{4}$$

If the pure states Ψ_i comprising an ensemble are orthogonal, then they are mutually distinguishable, and can thus be treated as classical states. In this situation, the von Neumann entropy is equal to the Shannon entropy of the probabilities

$$H = -\sum_i p_i \log p_i.$$

When the pure states ψ_i are nonorthogonal, and thus not wholly distinguishable as physical states, the ensemble's von Neumann entropy is less than the Shannon entropy.

It is not hard to show that for any bipartite pure state Ψ, the density matrices ρ_A and ρ_B of its parts have equal rank and equal spectra of nonzero eigenvalues. Moreover, the original state has an especially simple expression in terms of these eigenvalues and eigenvectors

$$\Psi = \sum_k \sqrt{\lambda_k} |a_k\rangle \otimes |b_k\rangle \tag{5}$$

where $|a_k\rangle$ and $|b_k\rangle$ are eigenvectors of ρ_A and ρ_B, respectively, corresponding to the positive eigenvalues λ_k. This expression, known as the Schmidt decomposition, unfortunately has no simple counterpart for tripartite and higher systems.

The recent rapid progress in the theory of quantum information processing can be divided into two related parts: quantum computation and quantum information theory. Although major practical questions remain concerning the physical realization of quantum computers, many of the most important theoretical questions in quantum computation have already been answered: quantum algorithms are known to provide an exponential speedup, compared to known classical algorithms, for integer factoring and a few other problems, a quadratic speedup for a broad range of search and optimization problems, and no significant speedup for such problems as iterated function evaluation. The discovery of quantum error-correcting codes and fault-tolerant gate arrays (reviewed in [59]) means that, in principle, finitely reliable components are sufficient to perform arbitrarily large reliable quantum computations, just as in the theory of classical computation. But there is a quantitative difference. Today's classical devices (e.g., CMOS transistors) are so intrinsically reliable that fault-tolerant circuits are rarely needed. By contrast, today's primitive quantum hardware is several orders too unreliable to be corrected by known fault-tolerant circuit designs. Fortunately, there appears to be no fundamental reason why this gap cannot be closed by future improvements in hardware and software.

Here we concentrate on the second area: quantum information theory, where the classical notions of source, channel, code, and capacity have been generalized to encompass the

TABLE I
CLASSICAL-QUANTUM COMPARISON

Property	Classical	Quantum
State representation	String of bits $x \in \{0, 1\}^n$	String of qubits $\psi = \sum_x c_x \lvert x\rangle$
Computation primitives	Deterministic or stochastic one- and two-bit operations	One- and two-qubit unitary transformations
Reliable computations from unreliable gates	Yes, by classical fault-tolerant gate arrays	Yes, by quantum fault-tolerant gate arrays
Quantum computational speedups		Factoring: exponential speedup Search: quadratic speedup Black-box iteration: no speedup
Communication primitives	Transmitting a classical bit	Transmitting a classical bit Transmitting a qubit Sharing an EPR pair
Source entropy	$H = -\sum p(x) \log p(x)$	$S = -\mathrm{Tr}\, \rho \log \rho$
Error-correction techniques	Error-correcting codes	Quantum error-correcting codes Entanglement distillation
Noisy channel capacities	Classical capacity C_1 equals maximum mutual information through a single channel use	Classical capacity $C \geq C_1$ Unassisted quantum capacity $Q \leq C$ Assisted quantum capacity $Q_2 \geq Q$
Entanglement-assisted communication		Superdense coding Quantum teleportation
Communication complexity	Bit communication cost of distributed computation	Qubit cost, or entanglement-assisted bit cost, can be less
Agreement on a secret cryptographic key	Insecure against unlimited computing power, or if $P = NP$	Secure against general quantum attack and unlimited computing
Two-party bit commitment	Insecure against unlimited computing power, or if $P = NP$	Insecure against attack by a quantum computer
Digital signatures	Insecure against unlimited computing power, or if $P = NP$	No known quantum realization

optimal use of various channels, noiseless and noisy, for communicating not only classical information but also intact quantum states, and for sharing entanglement between separated observers. Although the fundamental physics and mathematics on which it is based is over fifty years old, the new theory has taken shape mostly over the last five years. Quantum data compression [3], [45], superdense coding [16], quantum teleportation [10], and entanglement concentration [8], [53] exemplify nontrivial ways in which quantum channels can be used, alone or in combination with classical channels, to transmit quantum and classical information. More recently, quantum error-correcting codes [14], [19]–[23], [25], [28], [33], [46], [48], [58], [66], [68], [69] and entanglement distillation protocols [12], [14], [24], [40], [41] have been discovered which allow noisy quantum channels, if not too noisy, to be substituted for noiseless ones in these applications. Important problems still open include finding exact expressions, rather than upper and lower bounds, for the classical and quantum capacities of noisy quantum channels. Some of the main similarities and differences between classical and quantum information processing are summarized in Table I.

As noted earlier, entanglement plays a central role in this enlarged information theory, complementary in several respects to the role of classical information. One of the important tasks of quantum information theory is therefore to devise quantitative measures of entanglement for bipartite and multipartite systems, pure and mixed. More generally, the theory should characterize the efficiency with which multipartite states can be transformed into one another by local operations and classical communication alone, without the exchange of quantum information among the parties. A complementary question is the extent to which prior entanglement among separated parties can reduce "communication complexity," i.e., the amount of classical communication needed to evaluate functions of several inputs, one held by each party.

II. SOURCE AND CHANNEL CODING

A natural quantum analog of a (discrete, memoryless) information *source* is an ensemble $\mathcal{E}$ of pure or mixed states [2] $\rho_1, \rho_2, \cdots, \rho_k$, emitted with known probabilities $p_1 \cdots p_k$. The quantum analog of a (discrete noiseless) channel is any quantum system capable of existing in an arbitrary state in some finite-dimensional Hilbert space, of being entangled with other similar quantum systems, and of remaining stably in this entangled or superposed state while enroute from sender to receiver. Just as a sequence of n bits, sent through a classical channel, can be used to transmit any of up to 2^n distinct classical messages, so a sequence of n elementary 2-state quantum systems or *qubits* can be used to transmit an arbitrary quantum state in a Hilbert space of up to 2^n dimensions. The quantum analog of a noisy channel is a quantum system that interacts unitarily with an outside environment while enroute from sender to receiver. Noisy channels (including noiseless ones as a special case) may thus be described as superoperators.

If the states ρ_i of a quantum source are all orthogonal, the source can be considered purely classical, because complete information about the source state can be extracted by a measurement at the sending end, transmitted classically to the receiving end, and used there to make arbitrarily many faithful replicas of the source state. On the other hand, if the source states are pure and nonorthogonal, then no classical measurement can extract complete information about the source state, and, whenever a source state is sent through a quantum channel, at most one faithful copy of the source state can be produced at the receiving end, and then only if no faithful copy remains behind at the sending end. An interesting intermediate situation occurs when the source states are nonorthogonal but commuting mixed states (i.e., the density matrices of the states commute). Such a source can be "broadcast," i.e., given an unknown one of the source states ρ_i, two systems A and B can be prepared in a joint state $\boldsymbol{\rho}_i(AB)$, which is not a clone of the source state (i.e., $\boldsymbol{\rho}_i(AB) \neq \rho_i(A) \otimes \rho_i(B)$), but whose partial trace over either subsystem agrees with the source state

$$\rho_i = \mathrm{Tr}_A(\boldsymbol{\rho}_i(AB)) = \mathrm{Tr}_B(\boldsymbol{\rho}_i(AB)) = \rho_i.$$

This "broadcasting" is essentially the same as classically copying a random variable—the resulting copies each have the right distribution, but the joint distribution of the copies is not equal to the product distribution of several copies of the original source. If the density matrices of the source states do not commute (this includes the case of pure nonorthogonal states), then the source can neither be cloned nor broadcast [2].

Because quantum information cannot be read or copied without disturbing it, whatever encoding apparatus is used at the sending end of a quantum channel must function rather blindly. If the channel is to transmit nonorthogonal pure states faithfully, it must operate on the states that pass through without knowing or learning anything about them. For the same reasons, assessment of the quality of quantum data transmission is a somewhat delicate matter. If the source states are pure, and a quantum channel produces output W_i (in general a mixed state) on input ψ_i, then the transmission's *fidelity* [45] is defined as

$$F = \sum_i p_i \langle \psi_i | W_i | \psi_i \rangle. \tag{6}$$

This is the expectation, averaged over channel inputs, that the output would pass a test for being the same as the input, conducted by someone who knew what the input was. When even the source states ρ_i are mixed states, fidelity must be defined [44] by a more complicated formula

$$F = \sum_i p_i \left(\mathrm{Tr} \sqrt{\sqrt{\rho_i}\, W_i \sqrt{\rho_i}} \right)^2 \tag{7}$$

which represents the maximum of (6) over "purifications" [44] of ρ_i, i.e., pure states ψ_i in a larger Hilbert space having ρ_i as their partial trace.

The two most important techniques of classical information theory are data compression and error correction. When the encoder and decoder are allowed to perform quantum operations, these techniques can be extended to quantum sources and channels, enabling data from a quantum source to be recovered with arbitrarily high fidelity after transmission through a noiseless or noisy quantum channel, provided the entropy of the source is less than the *quantum capacity* of the channel.

Classical data compression allows information from a redundant source, e.g., a binary source emitting 0 and 1 with unequal probability, to be compressed without distortion into a bulk asymptotically approaching the source's Shannon entropy. Similarly, quantum data compression (cf. Fig. 2(a)) allows signals from a redundant quantum source, e.g., one emitting horizontal ($\leftrightarrow$) and diagonal (⤢) photons with equal probability, to be compressed into a bulk approaching the source's von Neumann entropy, $S(\rho) = -\mathrm{Tr}\, \rho \log_2 \rho$, where $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, with fidelity approaching 1 in the limit of large n. Quantum data compression [45] is performed essentially by projecting the state of a sequence of n source signals onto the subspace spanned by the $2^{n(H(\rho)+\epsilon)}$ most important eigenvectors of their joint density matrix (the n'th tensor power of ρ). This *typical subspace* is the subspace spanned by typical sequences of eigenvectors, where the probability associated with each eigenvector is its eigenvalue. For every positive ϵ and δ there exists an n such that for block size n or greater, the resulting projection has probability less that $\delta/2$ of failing (i.e., having the state not fall into the designated subspace when measured), and fidelity greater than $1 - \delta/2$ if it does succeed. Thus the overall fidelity exceeds $1 - \delta$.

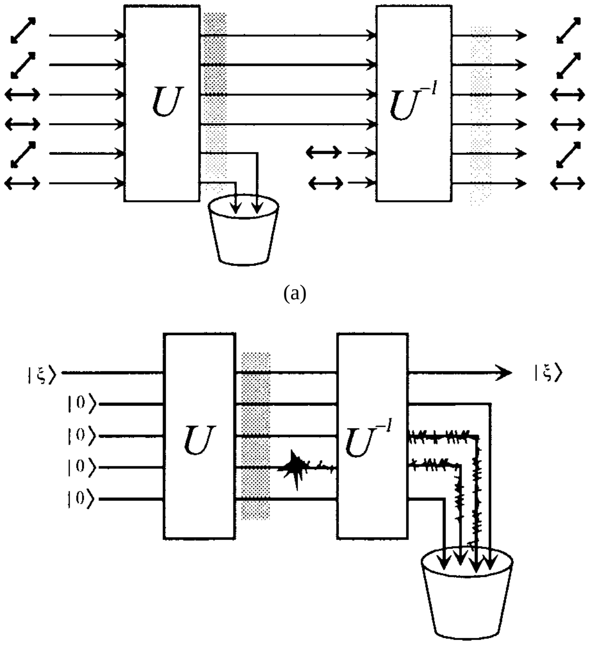
Slightly generalizing this example, suppose the source emits two equiprobable states $|\leftrightarrow\rangle$ and $\cos\theta |\leftrightarrow\rangle + \sin\theta |\updownarrow\rangle$, i.e., two polarizations differing by an angle θ. The corresponding density matrix is

$$\frac{1}{2} \begin{pmatrix} 1 + \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{pmatrix}. \tag{8}$$

This matrix has eigenvalues $\frac{1}{2}(1 \pm \cos\theta)$, so its von Neumann entropy, and the resulting compression factor, is $H_2(\frac{1}{2}(1 - \cos\theta))$, where H_2 is the dyadic entropy function

$$H_2(x) = -x \log_2 x - (1-x) \log_2 (1-x).$$

Though formally a close parallel to the classical noiseless coding theorem, quantum data compression is remarkable in that it can compress and re-expand each of 2^n distinct sequences of $\leftrightarrow$ and ⤢ photons, with fidelity approaching 1 *for the entire sequence*, even though the sequences cannot be reliably distinguished from one another by any measurement.



(a)

(b)

Fig. 2. (a) Quantum Data Compression: If classical data is redundant, due to unequal digit frequencies or correlations between the digits, it can be compressed to a smaller volume by techniques such as Huffman coding. Quantum data can be redundant in these ways but also in a third way: if the states in the data stream are nonorthogonal (e.g., a random stream of horizontal and 45° diagonal photons) and thus not wholly distinct as physical states. Such a data stream cannot be compressed by classical means, because the sending station would disturb the data by trying to read it. However, by performing unitary transformations on blocks of n incoming states, a quantum encoder can, without learning anything about the states, squeeze almost all their information into a smaller number of qubits, from which the original states can be reconstructed almost perfectly by an inverse transformation at the receiving station. The retained qubits leaving the encoder are heavily entangled (shading); the discarded qubits contain little information and are almost unentangled. The reconstruction approaches perfect fidelity, and the residual entanglement (pale shading) vanishes, in the limit of large n. (b) In a quantum error-correcting code, the encoder entangles the input state $|\xi\rangle$ with four standard qubits. The resulting entangled state can then withstand the corruption of any one of its qubits, and still allow recovery of the exact initial state by a decoder at the receiving end of the channel.

It is also quite interesting that the compression factor does not depend on the ensemble of states output by the source, but merely on the density matrix of these states. Although many different sources will give rise to the same density matrix, both the amount of compression achievable and the Schumacher–Jozsa algorithm for performing this compression depend only on the density matrix.

Quantum error-correcting codes [14], [19], [22], [23], [28], [33], [46], [48], [58], [66], [68], [69] have been the subject of intensive research since their discovery about three years ago. The idea that quantum error correction is possible is somewhat counterintuitive, because one familiar kind of classical error correction—in which the encoder makes several copies of the input and the decoder performs a majority vote over the channel outputs—cannot be used with quantum data because of the impossibility of accurately measuring or cloning an unknown quantum state. Nevertheless, quantum error-correcting codes (QECC) exist and are indeed a natural generalization of classical error-correcting codes ([cf. Fig. 2(b)). Rather than copying the input, the quantum encoder embeds it in a larger Hilbert space in such a way that the error processes the code is designed to correct—here interaction of any one of the qubits with the environment—do not allow any information about the encoded state to leak out. This enables the quantum decoder to restore the data qubit to its exact original state without duplicating any quantum information, while funneling the effects of the error into ancillary qubits, which are then discarded.

To see how quantum error-correcting codes are constructed, consider the following example. Suppose we take the simplest classical error-correcting code, the threefold repetition code, and try to turn it into a quantum error-correcting code in the most obvious way. We obtain the following transformation:

$$\begin{aligned}|0\rangle &\rightarrow |000\rangle\\ |1\rangle &\rightarrow |111\rangle.\end{aligned}\tag{9}$$

This code does not represent a cloning of the original qubit, but rather an embedding of its two-dimensional Hilbert space into a particular two-dimensional subspace of the eight-dimensional space of three qubits, namely that spanned by the vectors $|000\rangle$ and $|111\rangle$.

The problem with this encoding is that, while it protects against any single bit flip (i.e., $\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$ or $|0\rangle \rightarrow |1\rangle$, $|1\rangle \rightarrow |0\rangle$), a phase flip (i.e., $\left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)$, or $|0\rangle \rightarrow |0\rangle$, $|1\rangle \rightarrow -|1\rangle$) in any one of the qubits will produce a phase flip in the encoded state. Thus protection against bit flips has been obtained at the expense of increasing the vulnerability to phase flips. Equivalently, we have protected the states $|0\rangle$ and $|1\rangle$ (encoded as $|000\rangle$ and $|111\rangle$) but we have made the states $\frac{1}{\sqrt{2}}\,(|0\rangle \pm |1\rangle)$—encoded as $\frac{1}{\sqrt{2}}\,(|000\rangle \pm |111\rangle)$—more vulnerable, as a phase error in any one of the three encoding qubits will induce a phase error in the encoded qubit. The difficulty arises because a successful quantum error-correcting code must protect the entire subspace generated by superpositions of encoded states $|0\rangle$ and $|1\rangle$. In general, quantum error-correcting codes cannot merely protect specific quantum states, but must protect an entire subspace.

There is a duality between the role of bits and phases, as can be seen by considering the so-called Hadamard transformation $H = \frac{1}{\sqrt{2}}\left(\begin{smallmatrix}1&1\\1&-1\end{smallmatrix}\right)$

$$\begin{aligned}|0\rangle &\rightarrow \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\\ |1\rangle &\rightarrow \tfrac{1}{\sqrt{2}}\,(|0\rangle - |1\rangle).\end{aligned}\tag{10}$$

Under this transformation, bit flips become phase flips and *vice versa*. Applying the Hadamard transformation to the codewords of the triple-repetition code yields a code that protects against phase flips but not bit flips. This encoding is the following:

$$\begin{aligned}|0\rangle &\rightarrow \tfrac{1}{2}(|000\rangle + |011\rangle + |101\rangle + |110\rangle)\\ |1\rangle &\rightarrow \tfrac{1}{2}(|111\rangle + |100\rangle + |010\rangle + |001\rangle).\end{aligned}\tag{11}$$

Here phase flips can easily be seen to take an encoded $|0\rangle$ (or $|1\rangle$) to orthogonal states; for instance, a phase flip in the third qubit produces $\frac{1}{2}(|000\rangle - |011\rangle - |101\rangle + |110\rangle)$, orthogonal to both states in (11). Since the states resulting from a phase flip in any one of the three positions are all orthogonal to each other and to the original uncorrupted states, there is a von Neumann measurement to determine which phase flip has occurred, allowing it to be corrected. On the other hand, this encoding results in increased vulnerability to bit flips, as such an error in any qubit interchanges an encoded $|0\rangle$ and an encoded $|1\rangle$.

One objection that might be raised to the above analysis of the phase error-correcting code is phase flips are actually a very specific discrete form of error, while quantum-mechanical systems can undergo a continuous spectrum of errors. A general phase error is of the form $\left(\begin{smallmatrix}1&0\\0&e^{i\theta}\end{smallmatrix}\right)$, whereas a phase flip corresponds to $\theta = \pi$. The answer to this objection is to show that the above code will correct any phase error. In fact, any quantum code-correcting phase flips will also correct general phase errors.

To see that it is enough to consider phase flips, we first use the fact that a quantum state can be multiplied by an arbitrary phase factor, so we can rewrite the phase error above as $\left(\begin{smallmatrix}e^{i\phi}&0\\0&e^{-i\phi}\end{smallmatrix}\right)$ with $\phi = -\theta/2$. Consider what happens when this error is applied the first bit of the encoded state above. We obtain the mapping

$$\begin{aligned}|0\rangle \to\ & e^{i\phi}(|000\rangle + |011\rangle) + e^{-i\phi}(|101\rangle + |110\rangle)\\ &= \cos\phi(|000\rangle + |011\rangle + |101\rangle + |110\rangle)\\ &\quad + \sin\phi(|000\rangle + |011\rangle - |101\rangle - |110\rangle). \qquad (12)\end{aligned}$$

Observe that this is a superposition of an encoded $|0\rangle$ with no error, with amplitude $\cos\phi$, and an encoded $|0\rangle$ with a phase flip in the first bit, with amplitude $\sin\phi$. When we measure which bit is wrong, a phase flip in bit 1 will be observed with probability $\sin^2\phi$, and no error observed with probability $\cos^2\phi$. In either case, the act of measurement has reduced the state vector, so the measured error corresponds to the actual error in the state. This error can subsequently be corrected.

The trick to obtaining a code that protects against both bit flips and phase flips in any one qubit is apply the two preceding codes in a nested fashion to obtain a nine-qubit concatenated code:

$$\begin{aligned}|0\rangle &\to \frac{1}{2}\Big(|000000000\rangle + |111111000\rangle + |111000111\rangle\\ &\qquad + |000111111\rangle\Big)\\ |1\rangle &\to \frac{1}{2}\Big(|111111111\rangle + |000000111\rangle + |000111000\rangle\\ &\qquad + |111000000\rangle\Big). \qquad (13)\end{aligned}$$

Bit flips are corrected by the inner code and phase flips are corrected by the outer code. It is easy to check that these correction processes do not interfere with each other—even if one of the qubits suffers both a bit flip and a phase flip, the error will be corrected properly.

We now have a code which protects against bit and/or phase flips in any single qubit. As noted before, quantum mechanics has a continuous space of errors and bit and phase flips are only two specific possibilities. However, just as the ability to correct phase flips sufficed to correct any phase error, so the ability to correct both phase and bit flips suffices to correct any single-qubit error. This follows from the fact that the identity matrix I and the three Pauli matrices

$$\sigma_x = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \quad \sigma_y = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix} \quad \sigma_z = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \qquad (14)$$

form a basis for the space of all 4×4 matrices. The Pauli matrix σ_x corresponds to a bit flip, σ_z corresponds to a phase flip, and σ_y corresponds to a qubit where both the value and the phase have been flipped. In general, if any tensor product of up to t of the Pauli matrices in different qubits can be corrected, a general error in up to t qubits can be corrected.

One way of understanding quantum codes is viewing the encoding process as a way of separating the space of encoded states and the space of possible few-bit errors into orthogonal dimensions of Hilbert space. This makes it possible to perform a measurement of the error without disturbing the encoded state, allowing a subsequent unitary transformation to correct the error. In actuality, it is not necessary to measure the error; one can accomplish error correction by unitary transformations which separate the Hilbert space into a tensor product of a space containing the encoded state and a space containing the error.

Shortly after the nine-bit code was discovered, a code protecting one bit by mapping it into seven bits was discovered independently by Andrew Steane [69] (who had not seen the nine-bit construction) and by Calderbank and Shor [22]. This was just the first representative of an infinite class of codes also presented in these papers. Subsequently, a five-bit was discovered [14], [49] by searching the space of possible codes. Several other codes followed [59], [70], as well as several papers on the general theory of quantum error-correcting codes [28], [47]. By analyzing these examples, a more general class of quantum codes was discovered [20], [33]. This led to the discovery of a method of turning certain additive GF(4) codes into quantum codes [21]. Using this connection between classical and quantum codes made it possible to find MacWilliams identities [68] and linear programming bounds [61], [62] that apply to quantum codes. Although we will not discuss this subject further, the area is still quite active.

Despite the rapid progress in quantum error-correcting codes, the notion of quantum channel capacity is more complicated and less well understood than its classical counterpart. As noted above, a quantum channel generally has three distinct capacities: an unassisted capacity Q for transmitting intact quantum states, a typically larger capacity C for transmitting classical information, and a classically assisted quantum capacity Q_2, for transmitting intact quantum states with the help of a two-way classical side-channel. Paralleling the definition of capacity for classical channels, the unassisted quantum capacity $Q(\mathcal{N})$ of a noisy channel $\mathcal{N}$ may defined as the greatest rate (transmitted qubits per channel use) at which, for arbitrarily large n and arbitrarily small ϵ, every state ψ of n qubits can be recovered with fidelity greater than $1-\epsilon$ after block-encoding, transmission through the channel, and block-decoding. More precisely, $Q(\mathcal{N})$ is defined as

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\left\{\frac{n}{m}: \exists_{m,\mathcal{E},\mathcal{D}}\forall_{\psi\in H_{2^n}}\langle\psi|\mathcal{D}\mathcal{N}^{\otimes m}\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle > 1-\epsilon\right\}. \qquad (15)$$

Here $\mathcal{E}$ is an encoding superoperator from n qubits to m channel inputs and $\mathcal{D}$ is a decoding superoperator from m channel outputs to n qubits.

The classical capacity $C(\mathcal{N})$ of a noisy quantum channel is defined similarly as

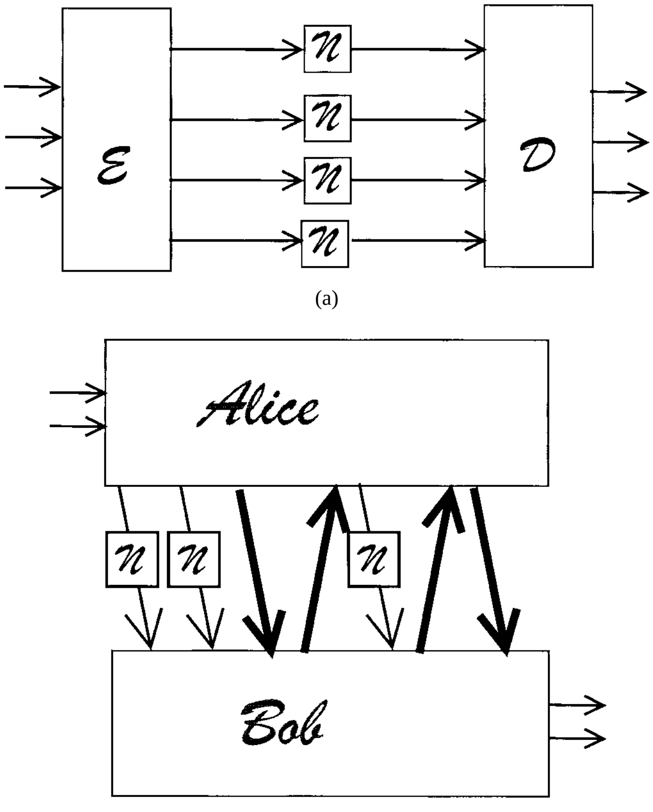
$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\left\{\frac{n}{m}: \exists_{m,\mathcal{E},\mathcal{D}}\forall_{\psi\in\{|0\rangle,\,|1\rangle\}^n} \langle\psi|\mathcal{D}\mathcal{N}^{\otimes m}\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle > 1-\epsilon\right\}. \quad (16)$$

This is the same as the definition of Q except that here the universal quantification is over all Boolean states $\psi \in \{|0\rangle, |1\rangle\}^n$, rather than all possible states $\psi \in H_{2^n}$ of the n qubits, because classical communication does not require superpositions of the Boolean states to be transmitted faithfully. Clearly, $Q(\mathcal{N}) \leq C(\mathcal{N})$ for all $\mathcal{N}$.

Equation (15) is the so-called *protected subspace* definition of quantum capacity. Other definitions of quantum capacity, based on the channel's ability to faithfully convey entanglement, or coherent information (a quantum analog of mutual information to be discussed later) in the limit of large block sizes, have been proposed [4], but there is reason [5], [50] to believe they are equivalent to the protected-subspace capacity.

Both Q and C can be understood in terms of the block diagrams like Fig. 3(a), in which n qubits are encoded, sent through m instances of the channel, and then decoded to yield a more-or-less faithful approximation of the input state, either for all input states in the case of Q, or for all Boolean input states in the case of C. The classically assisted capacity $Q_2(\mathcal{N})$ is defined in terms of a more complicated protocol (Fig. 3(b)) in which the sender Alice initially receives n qubits, after which she and the receiver Bob can perform local quantum operations and exchange classical messages freely in both directions, interspersed with m forward uses of the noisy quantum channel $\mathcal{N}$, with the goal of ultimately enabling Bob to output a faithful approximation of the n-qubit input state. The capacity Q_2 is then defined by a limiting expression like (15), but with the encoder/decoder combination $\mathcal{E}$, $\mathcal{D}$ replaced by an interactive protocol of the form of Fig. 3(b). Clearly, $Q_2(\mathcal{N}) \geq Q(\mathcal{N})$ for any quantum channel $\mathcal{N}$, and channels are known for which this inequality is strict; an open question is whether there are channels for which $Q_2 > C$.

The assisted quantum capacity Q_2 will be discussed later in detail; here it suffices to indicate why it can exceed the unassisted capacity Q. In a typical Q_2 protocol, Alice does not use the noisy channel $\mathcal{N}$ to transmit the input state ψ to Bob directly; instead, she prepares a number m of EPR pairs in her laboratory and sends one member of each pair to Bob through the noisy channel. The result is a set of impure EPR pairs (i.e., entangled mixed states) shared between Alice and Bob. By performing local operations and measurements on these impure EPR pairs, and engaging in classical discussion of the measurement results, Alice and Bob can distill from the m impure EPR pairs a smaller number n of nearly pure pairs, even when the channel $\mathcal{N}$ through which the pairs were shared is so noisy as to have zero unassisted quantum capacity. The purified EPR pairs are then used, in conjunction with further classical communication, to "teleport" the input state ψ to Bob with high fidelity. Teleportation and entanglement distillation are discussed at length in subsequent sections.



(a)

(b)

Fig. 3. (a) Unassisted quantum capacity Q of a noisy quantum channel is defined by the asymptotic performance of quantum error-correcting codes. (b) Classically assisted quantum capacity Q_2 is defined by the asymptotic performance of protocols involving forward uses of the noisy channel and unlimited classical communication between sender and receiver.

For most noisy quantum channels $\mathcal{N}$, none of the three capacities is known exactly, though upper and lower bounds have been obtained. This is true, for example, of the *depolarizing channel*, a rough analog of the classical binary-symmetric channel. The depolarizing channel transmits the input qubit intact with probability $1-p$ and replaces it by a random qubit with probability p. (Alternatively, the channel applies the identity operator to the input qubit with probability $1-3p/4$ and each of the three Pauli operators with probability $p/4$. The difference between $p/3$ and $p/4$ comes from the fact that applying at random one of the three Pauli operators or the identity operator effectively gives a random qubit.) A still simpler channel, whose capacities *are* known, is the *quantum erasure channel* [34], which transmits the input qubit intact with probability $1-p$ and with probability p replaces it by a recognizable erasure symbol orthogonal to both $|0\rangle$ and $|1\rangle$. For this channel, the classical capacity C and the assisted quantum capacity Q_2 are both known to be $1-p$ (the same as for a classical erasure channel) while the unassisted quantum capacity Q is known [13], [5] to be $\max\{1-2p, 0\}$.

The 50% erasure channel, for which $C = Q_2 = 1/2$ but $Q = 0$ nicely illustrates the difference between assisted and unassisted quantum capacities. The unassisted capacity Q of this channel must vanish by the following "no-cloning" argument, due to Smolin. Consider a (physically implementable) splitting device with one input port and two output ports which half the time sends the input qubit to the first output port and an erasure symbol to the second, the rest of the time sending the input qubit to the second and an erasure to the first. A recipient at either output port sees the source through a 50% erasure channel. Suppose there were a quantum error-correcting code by which an arbitrary unknown qubit could be encoded in such a way as to be reliably recoverable after transmission through a 50% erasure channel. Then each of the two recipients (call them Bob and Charlie) could recover a faithful copy of the unknown qubit, in effect cloning it. Since cloning is known to be impossible, a 50% erasure channel must have zero unassisted capacity. By the same token, a 50% depolarizing channel has zero unassisted quantum capacity. Nevertheless, either of these channels can still be used to transmit quantum information faithfully, when assisted by two-way classical communication in the manner of Fig. 3(b). In the case of the erasure channel, the sender shares EPR pairs through the channel, finds out by classical communication from the receiver which ones got through safely, and then uses the surviving EPR pairs to teleport the unknown qubits. Similar, but more complicated, arguments (cf. section on entanglement) show that the assisted quantum capacity of the depolarizing channel remains positive over the entire range $0 \leq p < 2/3$, while a stronger version of the no-cloning argument shows that its unassisted capacity vanishes for all $p \geq 1/3$. The no-cloning argument does not apply to Q_2 capacity because the classical communication breaks the symmetry between the two receivers, for example telling the sender, Alice, to perform different actions depending on whether she wants to send the input qubit to Bob or Charlie.

Mutual information between input and output plays an important role in classical theory, providing a nonasymptotic expression for the capacity, as the maximum (over input distributions) of the input:output mutual information for a single use of the channel, but this simplicity breaks down in the context of quantum channels. For none of the three capacities is there known a simple nonasymptotic quantity whose maximum for a single channel use equals the desired asymptotic capacity. Nevertheless, for both Q and C, mutual-information-like quantities have been devised that provide useful insight, and in some cases upper or lower bounds, on the asymptotic capacities. Here we discuss *coherent information* [4], a nonasymptotic quantity related to the unassisted quantum capacity Q. Later we will discuss the *Holevo bound*, a nonasymptotic quantity related to the classical capacity C of quantum channels. The Q_2 scenario, with its two-way communication, is obviously more complicated, and will not be discussed further in this context.

As noted earlier, a quantum channel, without loss of generality, may be viewed as a unitary interaction of the quantum information carrier q, enroute from sender to receiver, with an environment e initially in a standard state e_0. In order for the channel to carry intact quantum information, for example nonorthogonal states, or halves of EPR pairs, it is important that the environment not become too entangled with the quantum system passing through. At the other extreme, when the environment interacts strongly, as in Fig. 1(c), the channel loses its capacity to carry quantum information, though its classical capacity may be unimpaired. Considerations of this sort motivated the definition of the coherent information as the difference $S(\rho(q')) - S(\rho(e'))$ between the von Neumann entropy $S(\rho(q'))$ of the information carrier's mixed state after it has passed through the channel (the prime denotes an after-interaction state), and the von Neumann entropy of the channel environment's after-interaction mixed state $S(\rho(e'))$. The coherent information is thus a function both of the channel interaction and the density matrix $\rho(q)$ characterizing the channel input before interaction. One would like to say that a channel's asymptotic unassisted quantum capacity Q is simply the maximum of its one-shot coherent information over input distributions $\rho(q)$, and in simple cases this is indeed correct. For example, for a noiseless qubit channel or quantum wire, the environment does not entangle at all; hence $S(\rho(e')) = 0$ and $\rho(q') = \rho(q)$, so that the coherent information attains its maximum value of one qubit when the channel is connected to a maximal-entropy input (which can be thought of as a random, unknown qubit state, or as half an EPR pair). On the other hand, whatever input state $\rho(q)$ is supplied to the classical wire of Fig. 3(b), the environment gains so much entropy that the coherent information $S(\rho(q')) - S(\rho(e'))$ is zero. Similarly, for the erasure channel, the maximal one-shot coherent information accurately predicts the asymptotic capacity Q. However, in more complicated cases, for example the depolarizing channel, the coherent information underestimates Q due to a failure of additivity [25]: because of the possibility of entangling inputs across multiple channel uses, more coherent information can be sent through n channel uses than n times the maximum that can be sent through through one use.

The oldest branch of quantum information theory [35], [37], [38] concerns the use of quantum channels to transmit *classical* information. The classical capacity C is defined as the maximum asymptotic rate at which classical bits can be sent through the quantum channel with arbitrarily high reliability.[2] Even this seemingly pedestrian capacity is not easy to calculate for quantum channels, because it may depend on using a quantum encoder to prepare inputs entangled over multiple uses of the channel, and/or a quantum decoder to perform coherent measurements on multiple channel outputs (cf. Fig. 4). Recent results [39], [65] have made it possible to calculate the capacity C_{CQ} with classical encoding and quantum decoding for many channels, and to show that in some cases this exceeds the maximum mutual information C_{CC} that can be sent through a single use of the channel [30], [63]. The effect of a quantum *encoding* is less well understood. In particular it is not known whether there are channels for which entangled inputs are needed to achieve full capacity [15].

[2] The term classical capacity is often used for the maximum rate at which classical information can be sent through a quantum channel using a *fixed alphabet* of input states. In this case, the capacity is a function both of the input alphabet and the channel. Here we define the capacity without such a constraint, optimizing over all input alphabets including states entangled among multiple uses of the channel. The capacity thus defined is a function only of the channel.

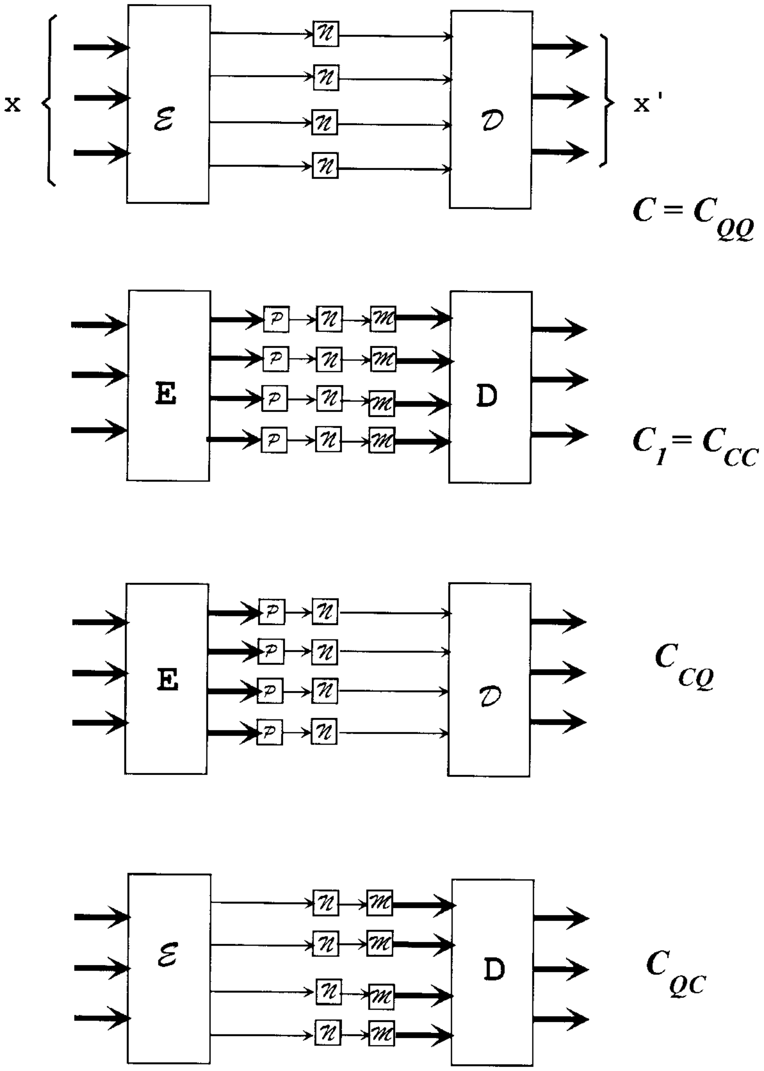


Fig. 4. The unrestricted classical capacity, C or C_{QQ}, of a quantum channel is defined using a quantum encoder and a quantum decoder. The one-shot capacity $C_1 = C_{CC}$ uses a classical encoder to direct the preparation $\mathcal{P}$ of inputs to independent instances of the quantum channel, whose outputs are independently measured $\mathcal{M}$ and then decoded by a classical decoder. It is equal to the classical maximal mutual information transmissible through a single instance of the channel. C_{CQ} and C_{QC} are intermediate cases.

Even without considering channels and codes, a nontrivial problem arises because of the incomplete distinguishability among the states comprising a quantum source (classical source states are, of course, always distinguishable by definition). Suppose that a quantum source is known to emit states ρ_i with probabilities p_i. How much classical information about i can be obtained by an optimal measurement? One of the first results in quantum information theory, stated (with no published proof) by Levitin [49] and independently (with proof) by Holevo [37], gives an upper bound on this so-called "accessible information" I

$$I \leq S\left(\sum_i p_i \rho_i\right) - \sum_i p_i S(\rho_i) \tag{17}$$

which holds with equality only when the states ρ_i commute. In general, no nice formulas are known for computing the accessible information, although there are a variety of formulas giving upper and lower bounds [29].

Before we can discuss accessible information in detail, we must treat quantum measurement more fully, as accessible information requires us to find the measurement of a source producing the most classical information. So far, we have only dealt with von Neumann measurements, which are measurements distinguishing among d orthogonal vectors in $\mathcal{H}_d$ (or possibly among subsets of them). More general quantum-mechanical measurements are possible; these can be made by introducing an *ancilla* quantum system, making a unitary transformation on the combined systems, and then performing a von Neumann measurement on the resulting state of the ancilla. This is the most general type of measurement possible, and there is a more convenient way of describing these measurements. Any such measurement in a d-dimensional Hilbert space corresponds to a collection of positive-semidefinite Hermitian operators $E_1, E_2, \cdots, E_k$ with $\sum_{i=1}^k E_i = I$, and k may be greater than the dimension d of the original Hilbert space. If a system in state ρ is measured, the probability of obtaining the ith outcome is $\mathrm{Tr}\,(\rho E_i)$. Since the E_i are positive operators, these are called positive operator valued measurements (abbreviated POVM or POM).

There are several subtleties associated with accessible information. Some of these are illustrated by a beautiful example investigated by Peres and Wootters [58], which will also prove relevant to the classical capacity C. They consider a source emitting with equal probabilities the three signal states which correspond to polarizations of a photon at 0°, 60°, 120°. These are

$$\begin{aligned} |a\rangle &= |0\rangle \\ |b\rangle &= \frac{1}{2}\,|0\rangle + \frac{\sqrt{3}}{2}\,|1\rangle \\ |c\rangle &= \frac{1}{2}\,|0\rangle - \frac{\sqrt{3}}{2}\,|1\rangle. \end{aligned} \tag{18}$$

The optimum measurement for these states is the POVM corresponding to $\frac{2}{3}\,|\overline{a}\rangle\langle\overline{a},\ \frac{2}{3}\,|\overline{b}\rangle\langle\overline{b}|,\ \frac{2}{3}\,|\overline{c}\rangle\langle\overline{c}|$, where the states $|\overline{a}\rangle$, $|\overline{b}\rangle$, $|\overline{c}\rangle$, are orthogonal to the original states $|a\rangle$, $|b\rangle$, $|c\rangle$, respectively; namely,

$$\begin{aligned} |\overline{a}\rangle &= |1\rangle \\ |\overline{b}\rangle &= \frac{1}{2}\,|1\rangle - \frac{\sqrt{3}}{2}\,|0\rangle \\ |\overline{c}\rangle &= \frac{1}{2}\,|1\rangle + \frac{\sqrt{3}}{2}\,|0\rangle. \end{aligned} \tag{19}$$

If the state $|a\rangle$, for example, is measured with respect to this POVM, the outcome $\overline{a}$ will never be observed, and there is probability $\frac{1}{2}$ of observing each of the outcomes $\overline{b}$ and $\overline{c}$. This gives accessible information of $\log_2 3 - 1 \approx 0.585$ bits.

If you consider two independent states output from the source above, and consider a joint measurement on them, this is equivalent to considering a source emitting the nine states $|aa\rangle$, $|ab\rangle, \cdots, |cc\rangle$, with equal probability. As might be expected, there is no way of gaining more than twice the information above, or 1.170 bits. However, the situation changes if you have a source emitting the three states $|aa\rangle$, $|bb\rangle$, $|cc\rangle$, with equal probability. We now describe a measurement that gives an accessible information of approximately 1.369 bits. The three states $|aa\rangle$, $|bb\rangle$, and $|cc\rangle$ span a three-dimensional vector space and are nearly orthogonal. Consider a von Neumann measurement which has axes close to these three states, namely, $|A\rangle$, $|B\rangle$, $|C\rangle$, where

$$|A\rangle = x|aa\rangle - y|bb\rangle - y|cc\rangle \tag{20}$$

and $|B\rangle$, $|C\rangle$ are defined similarly. Here x and y are chosen so that

$$\langle AB\rangle = \langle AC\rangle = \langle BC\rangle = 0$$

so

$$x = \left(4 + \sqrt{2}\right)/\sqrt{27} \approx 1.042$$

and

$$y = \left(2 - \sqrt{2}\right)/\sqrt{27} \approx 0.1127.$$

In this measurement, $|aa\rangle$ (for example) is measured as $|A\rangle$ with probability 0.9714 and as $|B\rangle$ and $|C\rangle$ with probability 0.0143 each. This gives an accessible information of 1.369 bits, producing more accessible information per qubit than the source emitting $|a\rangle$, $|b\rangle$, and $|c\rangle$ with equal probabilities.

After seeing this example, one might speculate that sources emitting more complicated codewords of length greater than two, and that joint measurements on these codewords could increase the accessible information still higher. This is indeed correct, and recently it was shown [39], [65] that Holevo's bound (17) is the correct formula for the limit of the maximum accessible information obtainable from tensor product sources having a marginal distribution equal to a given ensemble. Unlike the accessible information, the Holevo bound depends only on the density matrix of an ensemble and not the component states that make up this ensemble.

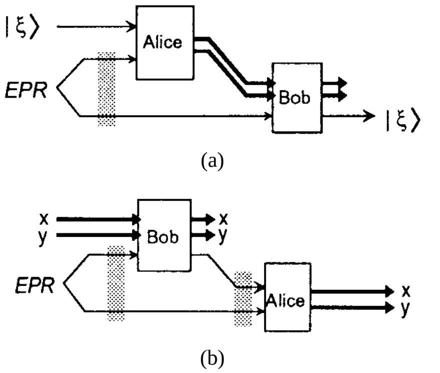
What does this example show? Suppose you are trying to communicate over a quantum channel, and are restricted to sending signals chosen from the three nonorthogonal states above. Achieving the largest capacity requires using codewords of length more than one and making joint (entangled) measurements on the signal. Now consider the problem of communicating classically over a general noisy quantum channel $\mathcal{N}$, where there is no restriction on the input alphabet, except that inputs are not allowed to be entangled across multiple channel uses. This is the situation of C_{CQ} in Fig. 4. If the marginal (one-shot) input ensemble consists of states ρ_i emitted with probabilities p_i, the corresponding output states (after passing through the channel) will be $\rho_i' = \mathcal{N}(\rho_i)$, and they will again occur with the probabilities p_i. Then the maximum C_{CQ} capacity achievable *with this input ensemble* is given by the Holevo bound

$$S\left(\sum_i p_i\rho_i'\right) - \sum_i p_i S(\rho_i') \tag{21}$$

and the unrestricted C_{CQ} is given by maximizing the above expression over the choice of input ensembles $\{\rho_i, p_i\}$. The ability to calculate the asymptotic capacity C_{CQ} by optimizing over the inputs to a single use of the channel have made it possible to show for a number of channels that C_{CQ} is greater than the one-shot capacity $C_1 = C_{CC}$ which is simply the maximum accessible information between input and output for a single channel use. Thus access to a quantum decoder spanning multiple channel outputs definitely increases the classical capacity of some quantum channels. It is not known whether access to a quantum encoder (allowing inputs to be entangled over multiple channel uses) increases the capacity. Thus of the four classical capacities depicted in Fig. 4, it is evident from the definitions that $C_{CC} \le \{C_{CQ}, C_{QC}\} \le C_{QQ}$. Specific examples are known where all four capacities are equal (e.g., the quantum erasure channel), and where C_{CQ} exceeds C_{CC}, but within these constraints the relation of C_{QC} and C_{QQ} to the other capacities is not known. Neither is the relation of these various classical capacities to the unassisted and assisted quantum capacities well understood. From the definitions it is clear that $C \ge Q$ for all channels, but is not known whether there are channels for which $C_1 < Q$.

III. Entanglement-Assisted Communication

While quantum source and channel codes optimize the use of one quantum resource—a quantum channel from sender to receiver—quantum teleportation [10] and quantum superdense



(a)

(b)

Fig. 5. (a) In quantum teleportation, prior sharing of an EPR pair, and transmission of a two-bit classical message from Alice to Bob, suffice to transmit an unknown quantum state even when nodirect quantum channel from Alice to Bob is available. (b) In quantum superdense coding, prior sharing of an EPR pair, and transmission of a qubit from Bob to Alice, suffices to transmit an arbitrary two-bit message (x, y).

coding [16] substitute a different quantum resource—namely, entanglement, in the form of entangled pairs of particles previously shared between sender and receiver—and use it to assist, respectively, in the performance of faithful quantum and classical communication.

In quantum teleportation (Fig. 5(a)) the sender (sometimes called "Alice") takes particle 1, whose unknown state $\xi_1 = \alpha \leftrightarrow_1 + \beta \updownarrow_1$ is to be teleported, and performs a joint measurement on it and particle 2, one member of the EPR pair. Particles 2 and 3 have been prepared beforehand in a maximally entangled EPR state, such as

$$\Phi^+_{23} = \sqrt{\tfrac{1}{2}}\,(\leftrightarrow_2\leftrightarrow_3 + \updownarrow_2\updownarrow_3).$$

The original state is thus

$$(\alpha \leftrightarrow_1 + \beta \updownarrow_1)\Phi^+_{23}. \tag{22}$$

It is easy to verify that this state can be rewritten as

$$\begin{aligned}\tfrac{1}{2}\,(\Phi^+_{12}(\alpha \leftrightarrow_3 + \beta \updownarrow_3) + \Phi^-_{12}(\alpha \leftrightarrow_3 - \beta \updownarrow_3)\\ + \Psi^+_{12}(\alpha \updownarrow_3 + \beta \leftrightarrow_3) + \Psi^-_{12}(\alpha \updownarrow_3 - \beta \leftrightarrow_3))\end{aligned} \tag{23}$$

where $\Phi^\pm$ and $\Psi^\pm$ are states in the so-called Bell basis, consisting of

$$\Phi^\pm_{12} = \sqrt{\tfrac{1}{2}}\,(\leftrightarrow_1\leftrightarrow_2 \pm \updownarrow_1\updownarrow_2)$$

and

$$\Psi^\pm_{12} = \sqrt{\tfrac{1}{2}}\,(\leftrightarrow_1\updownarrow_2 \pm \updownarrow_1\leftrightarrow_2) \tag{24}$$

four orthogonal maximally entangled states. The measurement on particles 1 and 2 projects them onto the Bell basis. The Bell measurement generates two bits of classical data, and leaves particle 3, now held by the receiver ("Bob"), in a residual state which can be unitarily transformed into a replica ξ_3 of the original quantum state ξ_1 which has been destroyed. This transformation is effected by subjecting particle 3 to one of four unitary operations $\mathbf{1}$, σ_z, σ_x, or σ_y according to which of the four outcomes, Φ^+, Φ^-, Ψ^+, or Ψ^- was obtained in the Bell measurement conducted by Alice.

Teleportation in effect splits the complete information in particle 1 into a classical part, carried by the two-bit message, and a purely quantum part, carried by the prior entanglement between particles 2 and 3. It avoids both cloning (the state ξ is destroyed in particle 1 before it is recreated in particle 3) and faster-than-light communication (the two-bit classical message must arrive at the receiver before the replica can be created). Note that we have somehow transmitted a continuously parameterized quantum state accurately using only two classical bits. Since the Holevo bound (21) shows that only one classical bit can be extracted from a qubit, teleportation does not give rise to any paradoxes, but the process is still quite counterintuitive. It is possible due to the entanglement that was originally shared between Alice and Bob, and is an example of what Einstein called the "spooky action at a distance" present in quantum mechanics.

A closely related effect is *superdense coding* (Fig. 5(b)), a scheme due to Wiesner [16]. Here also Alice and Bob begin by sharing an EPR pair. The sender (whom we now call Bob because he performs the same actions as Bob in teleportation) then encodes a two-bit classical message by performing one of the four unitary operations mentioned above on his member of the pair, thereby placing the pair as a whole into a corresponding one of the four orthogonal Bell states. The treated particle is then sent to Alice, who by measuring the particles together can reliably recover both bits of the classical message. Thus the full classical information capacity of two particles is made available, even though only one is directly handled by the sender.

Although the Bell states are quintessentially nonlocal, in the sense of being maximally entangled, they have a number of useful local properties. They can be converted into one another by local operations, in particular by the Pauli operators. For example, σ_x, applied to either member of an EPR pair (but not both), interconverts $\Phi^\pm$ and $\Psi^\pm$, while σ_y interconverts $\Phi^\pm$ with $\Psi^\mp$. If Alice and Bob are given an unknown one of the four Bell states, neither Alice nor Bob alone can learn anything about which one it is by a unilateral local measurement, but if they each perform a measurement and share the results, they can then test whether their unknown state belongs to an arbitrary set of two Bell states. For example, to test whether it is a Ψ or a Φ state, they measure each qubit in the standard basis and compare the results. If the results agree, they had a Φ state; otherwise, they had a Ψ state. Of course, this is a rather destructive kind of testing, because after the measurement the Bell state is gone and cannot be brought back by any local actions. A Bell state can be conveniently indexed by two classical bits, the first or amplitude bit saying whether the Bell state is Φ or Ψ, and the second or phase bit saying whether it is of the $+$ or $-$ type. Bilateral local measurements of the type we have described have the effect of extracting one bit of information about an unknown Bell state while destroying the other. By choice of local measurement, Alice and Bob can ascertain the phase bit, the amplitude bit, or the XOR of the phase and amplitude bits (this amounts to determining whether the unknown state is in the subset $\{\Phi^+, \Psi^-\}$) but in all cases, the other bit is destroyed. More generally [14], if Alice and Bob share n Bell states (Alice holding one qubit of each pair and Bob the other), they can, by local actions, determine the parity of an arbitrary subset of bits in the $2n$-bit string describing their Bell states, at the cost of measuring (and therefore sacrificing) one of the pairs. This ability to manipulate and extract information from Bell states by local actions has important consequences for entanglement distillation, classically assisted (Q_2) quantum communication, and quantum cryptography.

IV. Quantitative Theory of Entanglement

Since entanglement appears to be responsible for the remarkable behavior of information in quantum mechanics, a means of quantifying it would seem useful. We will discuss several such measures in this section. There is one "best" measure for entanglement of pure states, which is defined using entropy. However, there does not appear to be a unique best measure of entanglement for mixed states; which measure is best depends on what the measure is being used for. We will discuss several measures, all of which agree for pure quantum states.

Suppose Alice and Bob each hold one piece of a system in some quantum state. It is easy to define when these two pieces are entangled: this happens when the state of the entire system is not a mixture of tensor product states. More formally, the system is entangled iff its density matrix cannot be written in the form

$$\rho^{AB} = \sum_i p_i\, \rho_i^A \otimes \rho_i^B \tag{25}$$

where ρ_i^A and ρ_i^B are states of Alice's and Bob's subsystems and p_i are probabilities. Any unentangled state can be prepared by Alice and Bob using only local quantum-state preparation and classical communication: first they choose the index i with probability p_i and then ρ_i^A and ρ_i^B can be prepared locally without further communication. Preparation of an entangled state, on the other hand, requires that Alice and Bob either share some pre-existing entanglement or acquire such entanglement by transmission of quantum states between them. Determining whether a given density matrix has nonzero entanglement is not easy, especially if Alice and Bob each have higher than two-dimensional Hilbert spaces to work with. We shall later discuss a test that addresses this question. However, we will first discuss measures of entanglement.

If ρ^{AB} is a pure state (i.e., a rank-1 matrix $|\Psi^{AB}\rangle\langle\Psi^{AB}|$), there is a unique entanglement measure which has the properties one might hope for, namely, the entropy of entanglement

$$E(\Psi^{AB}) = S(\mathrm{Tr}_B\, \rho^{AB}) = S(\mathrm{Tr}_A\, \rho^{AB}). \tag{26}$$

This is the von Neumann entropy of the mixed state obtained when either Bob's or Alice's subsystem is disregarded. Why is this the right definition? To answer that, we need first decide what properties a good measure of entanglement must have. If entanglement is a resource, Alice and Bob ought not to be able to increase their entanglement by operations that involve only classical communication between them and local quantum operations. The ideal situation would then be if, whenever we are given two states ρ_1 and ρ_2 with entanglement E_1 and E_2, respectively, with $E_1 > E_2$, we could always reach the second state from the first state using only local quantum operations and classical communication. Although this turns out to be too much to ask, an asymptotic version of this ideal situation holds when ρ_1 and ρ_2 are pure states. For any two pure bipartite states Ψ and Ψ' (since we will always be considering bipartite states we omit the superscript AB), in the large n limit, n independent copies of Ψ (in other words, the state $\Psi^{\otimes n}$) can be transformed by local actions and classical communication into a state arbitrarily close to $\Psi'^{\otimes n'}$, with the fidelity of the approximation tending to 1 and the yield n'/n tending to $E(\Psi)/E(\Psi')$ in the limit of large n [8], [32], [53]. An important property of this definition of entanglement is that it is *additive*. That is, if Alice and Bob share two independent systems with entanglement E_1 and E_2, respectively, the combined system will have entanglement $E_1 + E_2$.

For mixed states, it appears that the amount of pure-state entanglement asymptotically required to prepare a mixed state may, in general, be larger than the amount of pure-state entanglement that asymptotically can be extracted from that mixed state. We call the first quantity *entanglement of formation* and the second *distillable entanglement.* The formal definitions of these involve, for entanglement of formation, the number of EPR pairs required to create many copies of the state with high fidelity; and for distillable entanglement, the number of nearly perfect EPR pairs distillable with high fidelity from many copies of the state. It has been conjectured that the entanglement of formation of a mixed state can be strictly larger than the distillable entanglement. That we cannot prove or disprove this conjecture shows how much is still unknown in quantum information theory.

The entanglement of formation and distillable entanglement as defined above are asymptotic quantities. For the entanglement of formation, there is also a related natural nonasymptotic quantity. As we saw previously, mixed quantum states can, in general, be represented as probabilistic mixtures of pure states, and can be so represented in many different ways. The *one-shot* entanglement of formation can thus be defined as the average entanglement of the pure states in a mixture giving the desired mixed state most efficiently. That is,

$$E_F^{(1)}(\rho) = \min \left\{ \sum_i p_i E(\Psi_i) \middle| \rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i| \right\}. \tag{27}$$

This quantity is clearly at least as large as the entanglement of formation, since the component pure states can asymptotically be created efficiently and then mixed together. It is not known whether the one-shot entanglement of formation is additive, that is, if

$$E_F^{(1)}(\rho_1 \otimes \rho_2) = E_F^{(1)}(\rho_1) + E_F^{(1)}(\rho_2). \tag{28}$$

If it is additive, then it must agree with the asymptotic definition of entanglement of formation. Hill and Wootters have obtained an exact expression for the one-shot entanglement of formation for arbitrary states of two qubits [36], [73].[3]

To illustrate how pure EPR pairs can be distilled from partly entangled mixed states, we first define a particularly

[3] In [12], [14], [36], and [73] the term "entanglement of formation" is used for what we call here the one-shot entanglement of formation $E_F^{(1)}$.

simple kind of bipartite mixed state of two qubits, the so-called Werner state, which may be thought of as a partly depolarized EPR pair. A Werner state of *fidelity* F, denoted W_F, is a mixture of F parts of a canonical Bell state (without loss of generality the state $\Phi^+ = \sqrt{\frac{1}{2}}\,(|00\rangle + |11\rangle)$ with $(1-F)/3$ parts of the other three Bell states, that is,

$$\begin{aligned} W_F =& F|\Psi^+\rangle\langle\Psi^+| \\ &+ (1-F)/3(|\Psi^-\rangle\langle\Psi^-| + |\Psi^+\rangle\langle\Psi^+| + |\Psi^-\rangle\langle\Psi^-|). \end{aligned} \tag{29}$$

There is a simple equivalence between p-depolarizing channels and Werner states W_F, where $F = 1 - 3p/4 \geq 1/4$: the Werner state W_F can be produced by transmitting a perfect Φ^+ pair through a p-depolarizing channel; conversely a p-depolarizing channel can be produced by performing teleportation with a Werner state W_F substituting for the perfect EPR pair called for in the teleportation protocol. This simple relation breaks down for other, less symmetric, channels and mixed states.

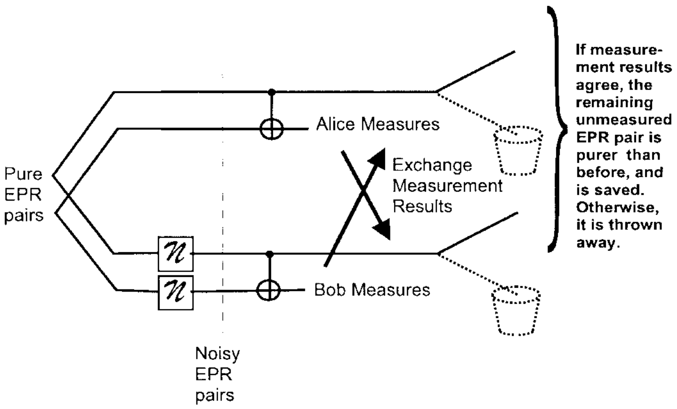
The final entanglement measure we wish to discuss is *relative entropy of entanglement,* introduced by Plenio and Vedral [72], [71] as a way of bounding the amount of distillable entanglement of a bipartite mixed state ρ. It is defined as

$$E_{RE}(\rho) = \min\{\mathrm{Tr}\,(\rho\log\rho - \rho\log\rho')|\rho' \text{ unentangled}\} \tag{30}$$

where the minimum is taken over all unentangled states ρ'. To explain why this gives a bound on distillable entanglement, we need to introduce the quantum analog of Sanov's theorem. Recall that Sanov's theorem expresses the distinguishability of two probability distributions in terms of relative entropy. The quantum Sanov's theorem [26] says that the probability of mistaking m copies of a quantum states ρ' for m copies of ρ after a measurement designed to test for ρ is at least $2^{-mS(\rho||\rho')}$, where

$$S(\rho||\rho') = \mathrm{Tr}\,(\rho(\log\rho - \log\rho'))$$

is the relative entropy. Further, there is a measurement that achieves this bound asymptotically. Now, if m copies of a state ρ can be purified to n EPR pairs, then the probability of distinguishing these n EPR pairs from a nonseparable state cannot be greater than the probability of distinguishing the m copies of the original state ρ from a nonseparable state, as the purification could be used as the first part of such a test. Using the projection onto the EPR state for distinguishing EPR pairs from unentangled states, we find that the probability of an unentangled state passing this test is at most 2^{-n}. This shows that the relative entropy of entanglement is an upper bound on the amount of distillable entanglement. It turns out that for Werner states $E_{RE} = 1 - H_2(F)$; more generally, for Bell-diagonal states $E_{RE} = 1 - H_2(\lambda)$ where λ is the largest eigenvalue in the density matrix. Rains [62] proved the same bound on distillable entanglement for Bell-diagonal states using a quite different method involving weight enumerators. E_{RE} is the best known upper bound on the distillable entanglement of Bell-diagonal states, and indeed for mixtures of *two* Bell states it gives the distillable entanglement exactly.



If measurement results agree, the remaining unmeasured EPR pair is purer than before, and is saved. Otherwise, it is thrown away.

Fig. 6. One step of entanglement distillation by the iterative comparison method.

As noted earlier, one reason for wanting to distill good EPR pairs from Werner states, or other entangled mixed states, is in order to use a noisy quantum channel for reliable quantum communication. Given a depolarizing channel, Alice and Bob can use it to build up a supply of Werner states, distill some good EPR pairs from these, then use the EPR pairs, in conjunction with classical communication, to teleport the desired quantum information from Alice to Bob, or vice versa, even though no direct noiseless quantum channel is available, and even though the noisy quantum channel that is available may be too noisy to be salvaged by quantum error-correcting codes.

We now describe one way of distilling good EPR pairs from Werner states. Suppose that Alice and Bob share two EPR pairs. It is easily checked that if both Alice and Bob XOR their halves of the two perfect Φ^+ states together (this so-called bilateral XOR operation is shown in Fig. 6), they obtain two new Φ^+ states. More generally, if Alice and Bob bilaterally XOR any two Bell states together, the result will again be two Bell states, which depend on the initial Bell states in a simple reversible manner (the source Bell state's amplitude bit gets XORed into the target's amplitude bit, while the target's phase bit gets XORed into the source's phase bit, leaving the amplitude of the source and the phase of the target unchanged). Suppose Alice and Bob share two Werner states. When the halves of these are XOR'ed together, then if either EPR pair is imperfect (i.e., not a Φ^+ state), the impurity is likely to spread to both output pairs. One of the resulting pairs (say the target) can be measured in the standard basis, and the results compared to see whether they agree, as they should if both incoming pairs were good Φ^+ states. This destroys the measured pair, but gives an indication as to whether the surviving unmeasured pair is good or not. If the measurement results agree, then Alice and Bob keep the other pair, and if not, they throw it away. We can easily analyze the first-order efficiency of this process. When we XOR the two Werner pairs together, if we start with fidelity $1 - \epsilon$, then for small ϵ, the outgoing fidelity of each is roughly $1 - 2\epsilon$. Now, when one of the two impure EPR pairs is checked, the measurement process catches two of the three possible Bell state errors (Ψ^+, Ψ^-, Φ^-) thus leaving the unmeasured pair in a state with fidelity roughly $1 - \frac{2}{3}\epsilon$. Doing the calculations in detail, we find this process increases the fidelity of Werner states if the initial fidelity is better than 50%. On the other hand, any Werner state with less than 50% fidelity can be shown to be unentangled [12], so it is hopeless to try to extract entanglement from it.

This procedure is not the best distillation method known for Werner states. It can be improved in a number of ways. One improvement is to substitute a more intelligent iterative comparison procedure due to Macchiavello [24] (cf. also the explanation in [14]) which takes advantage of the fact that after the first iteration the output is no longer a Werner state, but a more structured mixture of Bell states. A second improvement is that when the Werner states' fidelity has been sufficiently improved (to above about 0.83), Alice and Bob should stop the iterative technique and finish the distillation by another technique, called hashing [14]. This involves taking a large number n of the partially purified pairs and locally operating on them to measure parities of random subsets of the $2n$ phase and amplitude bits. This hashing procedure (essentially the decoding of a random linear code), allows the remaining Bell-state errors to be efficiently found and corrected at a cost of discarding a number of pairs asymptotically approaching the entropy of the $2n$-bit sequence. The asymptotic yield of arbitrarily pure EPR pairs by hashing is thus $1-S(\rho)$, where ρ is the density matrix of the input states to the hashing process.

The amount of noise an entanglement distillation protocol can tolerate depends on whether the protocol requires two-way communication between Alice and Bob (like the iterative comparison method), or can work with only one-way communication (like the hashing method). Protocols requiring only one-way communication are equivalent to quantum error-correcting codes, attaining the capacity Q in the limit of large n. Protocols depending on two-way communication, such as iterative comparison followed by hashing, give rise to the classically assisted Q_2 protocols of the type shown in Fig. 3(b). The yield of pure EPR pairs from Werner states, from hashing (D_H) and from the best known two-way (D_M) distillation protocol is shown in Fig. 7, along with the one-shot entanglement of formation E_F^1 and the relative entropy of entanglement E_{RE} of the Werner states. D_H vanishes for $F < 0.81071$; the best known one-way protocol [25] is marginally better, having a cutoff at $F = 0.80944$. The other three quantities, D_M, E_{RE}, and E_F^1, all vanish at $F = \frac{1}{2}$. Because of the equivalence between Werner states and depolarizing channels, the two lower curves in this figure also give the best known lower bounds on the quantum capacities Q_2 and Q for the depolarizing channels with $p = 4(1-F)/3$, while E_{RE} is the best known upper bound on Q_2.

We now discuss an elegant necessary condition for separability, the so-called partial transpose criterion of Peres [56]. Consider a density matrix ρ. Its transpose is also a density matrix. If a mixed state shared between two parties is unentangled, then it is a mixture of unentangled pure states, each of which is a tensor product of pure states of the two subsystems. Thus if the transformation corresponding to a transpose by only one of the two parties is applied, the result, the so-called partial transpose $\tilde{\rho}$ will still be a density matrix. However, if the original mixed state ρ is entangled, then $\tilde{\rho}$ need not be a density matrix. It may, for example, have negative eigenvalues. To take a simple example, the density matrix corresponding to Bell state $\Psi^- = 1/\sqrt{2}\,(|01\rangle - |10\rangle)$, is

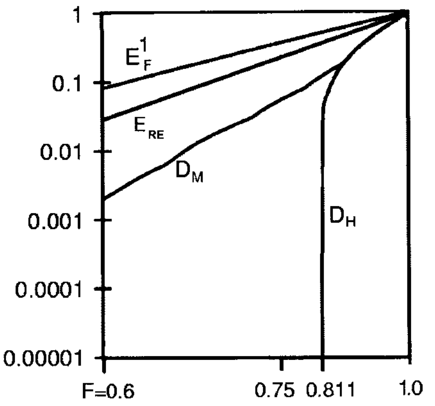


Fig. 7. Log plot of entanglement distillable from Werner states of fidelity F by hashing D_H and best known two-way protocol D_M, compared with their relative entropy of entanglement E_{RE} and their one-shot entanglement of formation E_F^1. (Abscissa is logarithmic in $F - \frac{1}{2}$.)

$$\frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The partial transpose of this matrix is

$$\frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

which has three eigenvalues of $\frac{1}{2}$ and one of $-\frac{1}{2}$, proving what we knew already, that Φ^- is entangled.

Subsequent work by the Horodecki family [40], [41] showed that positivity of the partial transpose is a necessary and sufficient condition for separability for mixed states of two qubits, and mixed states of a qubit with a three-state particle. Further research by the Hordeckis [42], [43] showed that no state with distillable entanglement, in any number of dimensions, can have a positive partial transpose, and led to the discovery of states with so-called "bound entanglement," i.e., entangled mixed states from which *no* pure entanglement can be distilled. These states (of which the simplest examples occur in $2\otimes 4$- and $3\otimes 3$-dimensional systems) are entangled in the sense that they are not mixtures of products states (in other words, their one-shot entanglement of formation is positive), yet their partial transposes are positive; hence they have no distillable entanglement.

V. Quantum Cryptography

Quantum cryptography is the art of applying the unique properties of quantum systems to cryptographic goals, that is, the protection of classical information from tampering or unauthorized disclosure in a multiparty setting where not all the parties trust one another.

This idea began with the proposal of quantum money by Wiesner [72]. Suppose that a piece of money has included with it some number n of quantum systems which are in a random sequence of nonorthogonal states, known only to the preparer, for example, a random sequence of the four polarizations ↔, ↕, ⤢, ⤡. Then anybody who tries to duplicate the bill must clone nonorthogonal states, an impossible task. There are drawbacks of this scheme—available quantum information carriers have a decoherence time shorter than inflationary half-life of most currencies, and only the issuing bank (which would have a classical record of the secret sequence) can check the validity of the money, so a counterfeiter could indeed pass fake copies off to anybody else. However, this idea formed the basis of quantum key distribution or QKD. Quantum key distribution [6], [7], [9], [11], [24], [27], [52], [54] is a task involving both quantum and classical communication among three parties, the legitimate users Alice and Bob and an eavesdropper Eve. Alice and Bob's goal is to use quantum uncertainty to do something that would be impossible by purely classical public communication—agree on a secret random bit string K, called a cryptographic key, that is informationally secure in the sense that Eve has little or no information on it.[4] In the quantum protocol (cf. Fig. 8), Eve is allowed to interact with the quantum information carriers (e.g., photons) enroute from Alice to Bob—at the risk of disturbing them—and can also passively listen to all classical communication between Alice and Bob, but she cannot alter or suppress the classical messages. Sometimes (e.g., if Eve jams or interacts strongly with the quantum signals) Alice and Bob will conclude that the quantum signals have been excessively disturbed, and therefore that no key can safely be agreed upon (designated by a frown in the figure); but, conditionally on Alice and Bob's concluding that it is safe to agree on a key, Eve's expected information on that key should be negligible.

The practical implementation of quantum key distribution is much further advanced than other kinds of quantum information processing, owing to the fact that the standard QKD protocols require no two-qubit interactions, only preparation and measurement of simple quantum states, along with classical communication and computations. Optical prototypes working over tens of kilometers of fiber, or even through a kilometer of open air (at night), have been built and tested. In principle, however, a quantum key distribution protocol could involve quantum computations by Alice and Bob; and to be sure of its security, one ought to allow Eve the full power of a quantum computer, even though Alice and Bob do not need one for the standard protocols. Fig. 8 shows the most general situation, where Alice, Eve, and Bob each have a separate quantum computer.

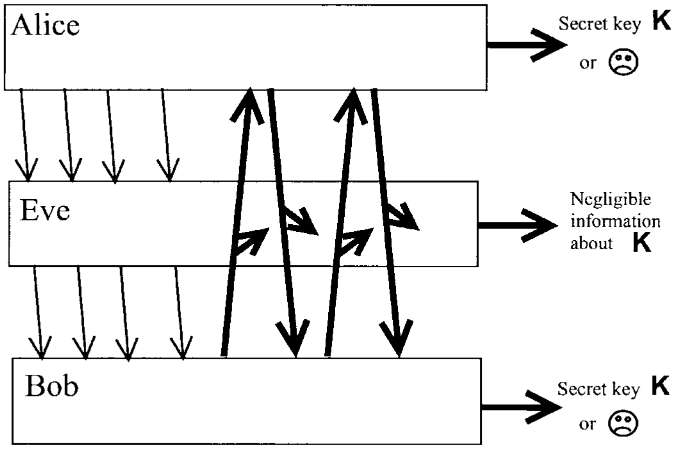


Fig. 8. In quantum key distribution, Alice and Bob, who trust each other but distrust Eve, wish to arrive at a secret key K of which Eve is ignorant. Alice and Bob have at their disposal a quantum channel on which Eve may eavesdrop actively, and a classical two-way channel on which she eavesdrops passively.

Various proofs of security of quantum key distribution protocols, especially the four-state "BB84" protocol of [9], have been offered. A complete security proof should encompass all attacks allowed by the laws of quantum mechanics, should be able to cope with noise, and should be applicable in realistic settings where noise arises not only from eavesdropping but also from noisy channels and detectors. Finally it should provide a way of calculating a safe rate of key generation as a function of the noise level observed by Alice and Bob. Early discussions of security treated only limited kinds of attacks, such as the interception and resending of individual photons. Subsequently [31] considered optimal eavesdropping strategies on individual qubits, while [17] considered a general collective attack on ancillas that have interacted separately with individual qubits. Most recently, Mayers and Yao [55] have given a security proof for BB84 against general quantum attacks, in the presence of finite channel and detector noise. Although they did not calculate an explicit noise threshold, their techniques can probably be used to show that BB84 still yields secure key in the presence of practical noise levels in the several percent range. The Mayers and Yao proof shows that QKD requires only current technology, yet is secure against attacks using any future technology consistent with the laws of quantum mechanics.

If we now allow Alice and Bob the full power of quantum computation, what further advantages do they gain? In contrast to BB84, such quantum-computational protocols for QKD, first proposed in [24] and rigorously developed in [52] are based on entanglement distillation, and require Alice and Bob to perform nontrivial gate operations on their quantum data. Aside from simplifying the proof of security, the main advantages of quantum-computational protocols for QKD are that they can tolerate more noise and, unlike BB84, can operate over arbitrarily great distances, using distributed quantum error correction along a chain of intermediate stations. In this approach to QKD, Alice and Bob use their insecure quantum channel to share a large number n of EPR pairs end-to-end, then distill a smaller number m of presumably good EPR pairs from the n distributed pairs, using the same sorts of distillation protocols they would use if they were distilling Werner states.

[4] Purely classical protocols for key agreement exist and are in widespread use, but these result in a key that is only computationally secure—an adversary with sufficient computing power could infer it from the messages exchanged between Alice and Bob. In particular, the most widely used classical key agreement protocols could be easily broken by a quantum computer, if one were available.

Finally, they use hashing to verify that the state of the m surviving pairs differs by only an exponentially small amount from a perfect product of EPR pairs such as $(\Phi^+)^{\otimes m}$. Now, if Alice and Bob actually possessed the perfect state $(\Phi^+)^{\otimes m}$, they could distill m bits of perfectly secret key from it by locally measuring each pair in the standard basis. Instead, they merely have a state approximating $(\Phi^+)^{\otimes m}$; but because the approximation is so good, they can go ahead as if it were perfect, and Eve's expected information on the resulting m bits of key will be bounded at an exponentially small fraction of bits.

Given the success of quantum key distribution, there was high hope that quantum techniques could help with another task, called two-party oblivious function evaluation. This is the problem, especially important in commerce and post-cold-war diplomacy, of enabling two mutually distrustful parties to cooperate in evaluating some publicly agreed function of private data held separately by each party, without compromising the private data any more than it would be compromised had they assigned the job of evaluating the function to a trusted intermediary. Initially Alice knows data x and Bob knows data y; when the protocol is finished, Alice and Bob should each also know $f(x, y)$, but neither party should know any more about the other party's private input than can logically be inferred from a knowledge of their own data and the common function value $f(x, y)$. Classical protocols for oblivious function evaluation exist, but like classical key agreement protocols they are only computationally secure, based on the assumption that certain problems are hard. Hopes for finding a secure quantum foundation for oblivious function evaluation were dampened last year by the proof that a different specific cryptographic problem, called bit commitment, is insecure in principle against quantum attacks [56], [52]. In bit commitment Alice has a bit which she does not wish to reveal to Bob. However, she wants to send some information to Bob which ensures that she can later reveal the value of her bit, and prove to him that it originally had the value she claims. A proposed protocol for quantum bit commitment [18] was found to have a flaw, and this flaw led to a proof that quantum bit commitment was impossible [52], [56]. This makes the design of quantum cryptographic protocols for tasks other than key distribution much more problematic, since bit commitment is a crucial ingredient in the classical implementations of many of them, and since many cryptographic primitives can be used to implement bit commitment, showing that they too are impossible.

ACKNOWLEDGMENT

The authors wish to thank numerous collaborators and colleagues for discussions of their work which is reviewed here, and A. R. Calderbank for helpful comments on a draft of this paper.

REFERENCES

[1] A. Barenco, C. H. Bennett, R. Cleve, D. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. Smolin, and H. Weinfurter, "Elementary gates for quantum computation," *Phys. Rev. A*, vol. 52, pp. 3457–3467, 1995.

[2] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, "Noncommuting mixed states cannot be broadcast," *Phys. Rev. Lett.*, vol. 76, pp. 2818–2821, 1996 (eprint quant-ph/9511010).

[3] H. Barnum, C. A. Fuchs, R. Jozsa, and B. Schumacher, "A general fidelity limit for quantum channels," 1996 (eprint quant-ph/9603014).

[4] H. Barnum, M. A. Nielsen, and B. Schumacher, "Information transmission through noisy quantum channels," 1997 (eprint quant-ph/9702049).

[5] H. Barnum, J. Smolin, and B. Terhal, "Results on quantum channel capacity," submitted to *Phys. Rev A* (eprint quant-ph/9711032).

[6] C. H. Bennett, "Quantum cryptography using any two nonorthogonal states," *Phys. Rev. Lett.*, vol. 68, pp. 3121–3124, 1992.

[7] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, "Experimental quantum cryptography," *J. Cryptol.*, vol. 5, pp. 3–28, 1992.

[8] 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.

[9] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *Proc. IEEE Int. Conf. Computers, Systems and Signal Processing* (Bangalore, India, 1984), pp. 175–179.

[10] 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.

[11] C. H. Bennett, G. Brassard, and D. Mermin, "Quantum cryptography without Bell's theorem," *Phys. Rev. Lett.*, vol. 68, pp. 557–559, 1992.

[12] C. H. Bennett, G. Brassard, B. Schumacher, S. Popescu, J. Smolin, and W. K. Wootters, "Purification of noisy entanglement and faithful teleportation via noisy channels," *Phys. Rev. Lett.*, vol. 76, pp. 722–725, 1996.

[13] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, "Capacities of quantum erasure channels," *Phys. Rev. Lett.*, vol. 78, pp. 3217–3220, 1997.

[14] C. H. Bennett, D. P. DiVincenzo, J. Smolin, and W. K. Wootters, "Mixed state entanglement and quantum error correction," *Phys. Rev. A*, vol. 54, pp. 3824–3851, 1996 (eprint quant-ph/9604024).

[15] C. H. Bennett, C. A. Fuchs, and J. A. Smolin, "Entanglement-enhanced classical communication on a noisy quantum channel," 1996 (eprint quant-ph/9611006).

[16] 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.

[17] E. Biham, M. Boyer, G. Brassard, J. van de Graaf, and T. Mor, "Security of quantum key distribution against all collective attacks" (eprint quant-ph/9801022).

[18] G. Brassard, C. Créepeau, R. Jozsa and D. Langlois, "A quantum bit commitment scheme provably unbreakable by both parties," in *Proc. 34th Annu. IEEE Symp. Foundations of Computer Science*, 1993, pp. 362–371.

[19] S. L. Braunstein, "Quantum error correction of dephasing in 3 qubits," 1996 (eprint quant-ph/9603024).

[20] A. R. Calderbank, E. Rains, P. W. Shor, and N. J. A. Sloane, "Quantum error correction and orthogonal geometry," *Phys. Rev. Lett.*, vol. 78, pp. 405–408, 1997 (eprint quant-ph/9605005).

[21] ———, "Quantum error correction via codes over GF(4), I" submitted to *IEEE Trans. Inform. Theory* (eprint quant-ph/9608006).

[22] A. R. Calderbank and P. W. Shor, "Good quantum error-correcting codes exist," *Phys. Rev. A*, vol. 54, pp. 1098–1106, 1995 (eprint quant-ph/9512032).

[23] I. L. Chuang and R. Laflamme, "Quantum error correction by coding," 1995 (eprint quant-ph/9511003).

[24] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, "Quantum privacy amplification and the security of quantum cryptography over noisy channels," *Phys. Rev. Lett.*, vol. 77, pp. 2818–2821, 1996. Erratum: vol. 80, p. 2022, 1998.

[25] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin, "Quantum-channel capacity of very noisy channels," *Phys. Rev. A.*, vol. 57, pp. 830–839, 1998 (eprint quant-ph/9706061).

[26] M. J. Donald, "*A priori* probability and localized observers," *Found. Phys.*, vol. 22, pp. 1111–1172, 1992.

[27] A. Ekert "Quantum cryptography based on Bell's theorem," *Phys. Rev. Lett.*, vol. 67, pp. 661–663, 1991.

[28] A. Ekert and C. Machiavello, "Quantum error correction for communication," *Phys. Rev. Lett.*, vol. 77, pp. 2585–2588, 1996 (eprint quant-ph/9602022).

[29] C. A. Fuchs, "Distinguishability and accessible information in quantum theory," Ph.D. dissertation, Univ. New Mexico, Albuquerque, 1996 (eprint quant-ph/9601020).

[30] ———, “Nonorthogonal quantum states maximize classical information capacity,” *Phys. Rev. Lett.*, vol. 79, pp. 1162–1165, 1997.

[31] C. A. Fuchs, N. Gisin, R. B. Griffiths, C.-S. Niu, and A. Peres “Optimal eavesdropping in quantum cryptography I,” quant-ph/9701039.

[32] N. Gisin, “Nonlocality criteria for quantum teleportation,” *Phys. Lett. A*, vol. 210, pp. 151–156, 1996.

[33] D. Gottesman, “A class of quantum error-correcting codes saturating the Hamming bound,” *Phys. Rev. A*, vol. 54, pp. 1862–1868, 1996 (eprint quant-ph/9604038).

[34] M. Grassl, T. Beth, and T. Pellizzari, “Codes for the quantum erasure channel” (eprint quant-ph/9610042).

[35] C. W. Helstrom *Quantum Detection and Estimation Theory*. New York: Academic, 1976.

[36] S. Hill and W. K. Wootters “Entanglement of a pair of quantum bits,” 1997 (eprint quant-ph/9703041).

[37] A. S. Holevo, “Information theoretical aspects of quantum measurement,” *Probl. Pered. Inform.*, vol. 9, no. 2, pp. 31–42, 1973 (in Russian), translation in *Probl. Inform. Transm.*, vol. 9, pp. 177–183, 1973.

[38] ———, “Problems in the mathematical theory of quantum communication channels,” *Rep. Math. Phys.*, vol. 12, pp. 273–278, 1977.

[39] ———, “The capacity of quantum channel with general signal states,” submitted to *IEEE Trans. Inform. Theory* (eprint quant-ph/9611023).

[40] M. Horodecki, P. Horodecki, and R. Horodecki, “Separability of mixed states: Necessary and sufficient conditions,” *Phys. Lett. A*, vol. 223, pp. 1–8 (eprint quant-ph/9605038).

[41] ———, “Distillability of inseparable quantum systems,” 1996 (eprint quant-ph/9607009).

[42] M. Horodecki, P. Horodecki, and R. Horodecki “Mixed-state entanglement and distillation: Is there a ‘bound’ entanglement in nature?” 1998 (eprint quant-ph/9801069).

[43] P. Horodecki, “Separability criterion and inseparable mixed states with positive partial transposition,” *Phys. Lett. A*, vol. 232, pp. 333–339, 1997 (eprint quant-ph/9703004).

[44] R. Jozsa, “Fidelity for mixed quantum states,” *J. Modern Opt.*, vol. 41, 2315–2323, 1994.

[45] R. Jozsa and B. Schumacher, “A new proof of the quantum noiseless coding theorem,” *J. Modern Opt.*, vol. 41, pp. 2343–2349, 1994.

[46] E. Knill and R. Laflamme, “A general theory of quantum error correction codes,” *Phys. Rev. A*, vol. 55, pp. 900–911, 1997 (eprint quant-ph/9604034).

[47] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory*. Berlin, Germany: Springer, 1983.

[48] R. Laflamme, C. Miquel, J.-P. Paz, and W. H. Zurek, “Perfect quantum error correction code,” *Phys. Rev. Lett.*, vol. 77, pp. 198–201, 1996.

[49] L. B. Levitin, “On the quantum measure of information,” in *Proc. 4th All-Union Conf. Information and Coding Theory* (Tashkent, 1969), in Russian.

[50] S. Lloyd, “The capacity of the noisy quantum channel,” 1996 (eprint quant-ph/9604015).

[51] H.-K. Lo and H. F. Chau, “Why quantum bit commitment and ideal quantum coin tossing are impossible,” 1997 (eprint quant-ph/9711065).

[52] ———, “Quantum computers render quantum key distribution unconditionally secure over arbitrarily long distance,” 1998 (eprint quant-ph/9711065).

[53] H.-K. Lo and S. Popescu, “Concentrating entanglement by local actions—beyond mean values,” 1997 (eprint quant-ph/9707038).

[54] D. Mayers and A. C. C. Yao, “Unconditional security in quantum cryptography,” 1998 (eprint quant-ph/9802025).

[55] D. Mayers, “The trouble with quantum bit commitment,” 1996 (eprint quant-ph/9603015).

[56] A. Peres, “Separability criterion for density matrices,” *Phys. Rev. Lett.*, vol. 77, pp. 1413–1415, 1996.

[57] A. Peres and W. K. Wootters, “Optimal detection of quantum information,” *Phys. Rev. Lett.*, vol. 66, pp. 1119–1122, 1991.

[58] M. B. Plenio, V. Vedral, and P. L. Knight, “Quantum error correction in the presence of spontaneous emission,” *Phys. Rev. A*, vol. 55, pp. 67–72, 1997 (eprint 9603022).

[59] J. Preskill, “Reliable quantum computers,” 1997 (eprint quant-ph/9705031).

[60] E. Rains,“Quantum weight enumerators” (eprint quant-ph 9611001).

[61] ———, “Quantum shadow enumerators,” eprint quant-ph 9612015.

[62] ———, “Entanglement purification via separable superoperators,” 1997 (eprint quant-ph/9707002).

[63] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota “A simple quantum channel having superadditivity of channel capacity,” 1997 (eprint quant-ph/9705043).

[64] B. Schumacher, “Sending entanglement through noisy channels,” 1996 (eprint quant-ph/9604023).

[65] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, *Phys Rev. A*, vol. 54, p. 1869, 1996.

[66] P. W. Shor, “Scheme for reducing decoherence in quantum computer memory,” *Phys. Rev. A*, vol. 52, pp. 2493–2496, 1995.

[67] P. Shor and R. Laflamme, “Quantum analog of the MacWilliams identities in classical coding theory,” *Phys. Rev. Lett.*, vol. 78, pp. 1600–1602, 1997.

[68] A. Steane, “Multiple particle interference and quantum error correction,” in *Proc. Roy. Soc. London A*,vol. 452, pp. 2551–2577, 1996 (eprint quant-ph/9601029).

[69] L. Vaidman, L. Goldenberg, and S. Wiesner, “Error prevention scheme with four particles” (eprint quant-ph/9603031).

[70] V. Vedral and M. Plenio, “Entanglement measures and purification procedures” (quant-ph/9707035).

[71] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, “Quantifying entanglement,” *Phys. Rev. Lett.*, vol. 78, pp. 2275–2279, 1997 (eprint quant-ph/9702027).

[72] S. Wiesner, “Conjugate coding,” *Sigact News*, vol. 15, no. 1, pp. 78–88, 1983.

[73] W. K. Wootters, “Entanglement of formation of an arbitrary state of two qubits,” 1997 (eprint quant-ph/9709029).

[74] W. H. Zurek, “Decoherence and the transition from quantum to classical,” *Phys. Today*, vol. 44, pp. 36–44, 1991.