Mixed-state entanglement and quantum error correction

Charles H. Bennett,[1] David P. DiVincenzo,[1] John A. Smolin,[2] and William K. Wootters[3]
[1]*IBM Research Division, Yorktown Heights, New York 10598*
[2]*Physics Department, University of California at Los Angeles, Los Angeles, California 90024*
[3]*Physics Department, Williams College, Williamstown, Massachusetts 01267*
(Received 23 April 1996; revised manuscript received 8 August 1996)

Entanglement purification protocols (EPPs) and quantum error-correcting codes (QECCs) provide two ways of protecting quantum states from interaction with the environment. In an EPP, perfectly entangled pure states are extracted, with some yield D, from a mixed state M shared by two parties; with a QECC, an arbitrary quantum state $|\xi\rangle$ can be transmitted at some rate Q through a noisy channel χ without degradation. We prove that an EPP involving one-way classical communication and acting on mixed state $\hat{M}(\chi)$ (obtained by sharing halves of Einstein-Podolsky-Rosen pairs through a channel χ) yields a QECC on χ with rate $Q=D$, and vice versa. We compare the amount of entanglement $E(M)$ required to prepare a mixed state M by local actions with the amounts $D_1(M)$ and $D_2(M)$ that can be locally distilled from it by EPPs using one- and two-way classical communication, respectively, and give an exact expression for $E(M)$ when M is Bell diagonal. While EPPs require classical communication, QECCs do not, and we prove Q is not increased by adding one-way classical communication. However, both D and Q can be increased by adding two-way communication. We show that certain noisy quantum channels, for example a 50% depolarizing channel, can be used for reliable transmission of quantum states if two-way communication is available, but cannot be used if only one-way communication is available. We exhibit a family of codes based on universal hashing able to achieve an asymptotic Q (or D) of $1-S$ for simple noise models, where S is the error entropy. We also obtain a specific, simple 5-bit single-error-correcting quantum block code. We prove that iff a QECC results in high fidelity for the case of no error then the QECC can be recast into a form where the encoder is the matrix inverse of the decoder. [S1050-2947(96)07711-6]

PACS number(s): 03.65.Bz, 42.50.Dv, 89.70.+c

I. INTRODUCTION

A. Entanglement and nonlocality in quantum physics

Among the most celebrated features of quantum mechanics is the Einstein-Podolsky-Rosen [1] (EPR) effect, in which anomalously strong correlations are observed between presently noninteracting particles that have interacted in the past. These nonlocal correlations occur only when the quantum state of the entire system is *entangled*, i.e., not representable as a tensor product of states of the parts. In Bohm's version of the EPR paradox, a pair of spin-1/2 particles, prepared in the singlet state

$$\Psi^- = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle), \tag{1}$$

and then separated, exhibit perfectly anticorrelated spin components when locally measured along any axis. Bell [2] and Clauser *et al.* [3] showed that these statistics violate inequalities that must be satisfied by any classical local hidden variable model of the particles' behavior. Repeated experimental confirmation [4] of the nonlocal correlations predicted by quantum mechanics is regarded as strong evidence in its favor.

Besides helping to confirm the validity of quantum mechanics, entanglement has assumed an important role in quantum information theory, a role in many ways complementary to the role of classical information. Much recent work in quantum information theory has been aimed at characterizing the channel resources necessary and sufficient to transmit unknown quantum states, rather than classical data, from a sender to a receiver. To avoid violations of physical law, the intact transmission of a general quantum state requires both a quantum resource, which cannot be cloned, and a directed resource, which cannot propagate superluminally. The sharing of entanglement requires only the former, while purely classical communication requires only the latter. In quantum teleportation [5] the two requirements are met by two separate systems, while in the direct, unimpeded transmission of a quantum particle, they are met by the same system. Quantum data compression [6] optimizes the use of quantum channels, allowing redundant quantum data, such as a random sequence of two nonorthogonal states, to be compressed to a bulk approximating its von Neumann entropy, then recovered at the receiving end with negligible distortion. On the other hand, quantum superdense coding [7] uses previously shared entanglement to double a quantum channel's capacity for carrying classical information.

Probably the most important achievement of classical information theory is the ability, using error-correcting codes, to transmit data reliably through a noisy channel. Quantum error-correcting codes (QECCs) [8–16] use coherent generalizations of classical error-correction techniques to protect quantum states from noise and decoherence during transmission through a noisy channel or storage in a noisy environment. Entanglement purification protocols (EPPs) [17] achieve a similar result indirectly, by distilling pure entangled states (e.g., singlets) from a larger number of impure entangled states (e.g., singlets shared through a noisy channel). The purified entangled states can then be used for reli-

able teleportation, thereby achieving the same effect as if a noiseless storage or transmission channel had been available. The present paper develops the quantitative theory of mixed state entanglement and its relation to reliable transmission of quantum information.

Entanglement is a property of bipartite systems—systems consisting of two parts A and B that are too far apart to interact, and whose state, pure or mixed, lies in a Hilbert space $\mathcal{H}=\mathcal{H}_A\otimes\mathcal{H}_B$ that is the tensor product of Hilbert spaces of these parts. Our goal is to develop a general theory of state transformations that can be performed on a bipartite system without bringing the parts together. We consider these transformations to be performed by two observers, ''Alice'' and ''Bob,'' each having access to one of the subsystems. We allow Alice and Bob to perform local actions, e.g., unitary transformations and measurements, on their respective subsystems along with whatever ancillary systems they might create in their own laboratories. Sometimes we will also allow them to coordinate their actions through one-way or two-way classical communication; however, we do not allow them to perform nonlocal quantum operations on the entire system nor to transmit fresh quantum states from one observer to the other. Of course two-way or even one-way classical communication is itself an element of nonlocality that would not be permitted, say, in a local hidden variable model, but we find that giving Alice and Bob the extra power of classical communication considerably enhances their power to manipulate bipartite states, without giving them so much power as to make all state transformations trivially possible, as would be the case if nonlocal quantum operations were allowed. We will usually assume that $\mathcal{H}_A$ and $\mathcal{H}_B$ have equal dimension N (no generality is lost, since either subsystem's Hilbert space can be embedded in a larger one by local actions).

B. Pure-state entanglement

For pure states, a sharp distinction can be drawn between entangled and unentangled states: a pure state is entangled or nonlocal if and only if its state vector Υ cannot be expressed as a product $\Upsilon_A\otimes\Upsilon_B$ of pure states of its parts. It has been shown that every entangled pure state violates some Bell-type inequality [19], while no product state does. Entangled states cannot be prepared from unentangled states by any sequence of local actions of Alice and Bob, even with the help of classical communication.

Quantitatively, a pure state's entanglement is conveniently measured by its entropy of entanglement,

$$E(\Upsilon)=S(\rho_A)=S(\rho_B), \tag{2}$$

the apparent entropy of either subsystem considered alone. Here $S(\rho)=-\mathrm{Tr}\rho\log_2\rho$ is the von Neumann entropy and $\rho_A=\mathrm{Tr}_B|\Upsilon\rangle\langle\Upsilon|$ is the reduced density matrix obtained by tracing the whole system's pure-state density matrix $|\Upsilon\rangle\langle\Upsilon|$ over Bob's degrees of freedom. Similarly $\rho_B=\mathrm{Tr}_A|\Upsilon\rangle\langle\Upsilon|$ is the partial trace over Alice's degrees of freedom.

The quantity E, which we shall henceforth often call simply *entanglement*, ranges from zero for a product state to $\log_2 N$ for a maximally entangled state of two N-state particles. $E=1$ for the singlet state Ψ^- of Eq. (1), either of whose spins, considered alone, appears to be in a maximally mixed state with one bit of entropy. Paralleling the term *qubit* for any two-state quantum system (e.g., a spin-$\frac{1}{2}$ particle), we define an *ebit* as the amount of entanglement in a maximally entangled state of two qubits, or any other pure bipartite state for which $E=1$.

Properties of E that make it a natural entanglement measure for pure states include the following.

(1) The entanglement of independent systems is additive, n shared singlets, for example, having n ebits of entanglement.

(2) E is conserved under local unitary operations, i.e., under any unitary transformation U that can be expressed as a product $U=U_A\otimes U_B$ of unitary operators on the separate subsystems.

(3) The expectation of E cannot be increased by local nonunitary operations: if a bipartite pure state Υ is subjected to a local nonunitary operation (e.g., measurement by Alice) resulting in residual pure states Υ_j with respective probabilities p_j, then the expected entanglement of the final states $\Sigma_j p_j E(\Upsilon_j)$ is no greater, but may be less, than the original entanglement $E(\Upsilon)$ [20]. In the present paper we generalize this result to mixed states: see Sec. II A.

(4) Entanglement can be concentrated and diluted with unit asymptotic efficiency [20], in the sense that for any two bipartite pure states Υ and Υ', if Alice and Bob are given a supply of n identical systems in a state $\mathbf{\Upsilon}=(\Upsilon)^n$, they can use local actions and one-way classical communication to prepare m identical systems in state $\mathbf{\Upsilon}'\approx(\Upsilon')^m$, with the yield m/n approaching $E(\Upsilon)/E(\Upsilon')$, the fidelity $|\langle\mathbf{\Upsilon}'|(\Upsilon')^m\rangle|^2$ approaching 1, and probability of failure approaching zero in the limit of large n.

With regard to entanglement, a pure bipartite state Υ is thus completely parametrized by $E(\Upsilon)$, with $E(\Upsilon)$ being both the asymptotic number of standard singlets required to locally prepare a system in state Υ—its ''entanglement of formation''—and the asymptotic number of standard singlets that can be prepared from a system in state Υ by local operations—its ''distillable entanglement.''

C. Mixed-state entanglement

One aim of the present paper is to extend the quantitative theory of entanglement to the more general situation in which Alice and Bob share a *mixed* state M, rather than a pure state Υ as discussed above. Entangled mixed states may arise (cf. Fig. 1) when one or both parts of an initially pure entangled state interact, intentionally or inadvertently, with other quantum degrees of freedom (shown in the diagram as noise processes N_A and N_B and shown explicitly in quantum channel χ in Fig. 13) resulting in a nonunitary evolution of the pure state Υ into a mixed state M. Another principal aim is to elucidate the extent to which mixed entangled states, or the noisy channels used to produce them, can nevertheless be used to transmit quantum information reliably. In this connection we develop a family of one-way entanglement purification protocols [17] and corresponding quantum error-correcting codes, as well as two-way entanglement purification protocols which can be used to transmit quantum states reliably through channels too noisy to be used reliably with any quantum error-correcting code.

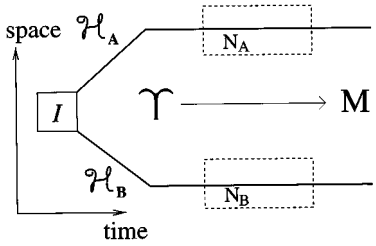

FIG. 1. Typical scenario for creation of entangled quantum states. At some early time and at location I, two quantum systems A and B interact [18], then become spatially separated, one going to Alice and the other to Bob. The joint system's state lies in a Hilbert space $\mathcal{H}=\mathcal{H}_A\otimes\mathcal{H}_B$ that is the tensor product of the spaces of the subsystems, but the state itself is not expressible as a product of states of the subsystems: $\Upsilon\neq\Upsilon_A\otimes\Upsilon_B$. State Υ, its pieces acted upon separately by noise processes N_A and N_B, evolves into mixed state M.

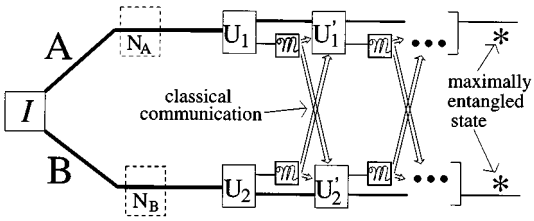

FIG. 2. Entanglement purification protocol involving two-way classical communication (2-EPP). In the basic step of 2-EPP, Alice and Bob subject the bipartite mixed state to two local unitary transformations U_1 and U_2. They then measure some of their particles $\mathcal{M}$, and interchange the results of these measurements (classical data transmission indicated by double lines). After a number of stages, such a protocol can produce a pure, near-maximally-entangled state (indicated by *'s).

The theory of mixed-state entanglement is more complicated and less well understood than that of pure-state entanglement. Even the qualitative distinction between local and nonlocal states is less clear. For example, Werner [21] has described mixed states which violate no Bell inequality with regard to simple spin measurements, yet appear to be nonlocal in other subtler ways. These include improving the fidelity of quantum teleportation above what could be achieved by purely classical communication [22], and giving nonclassical statistics when subjected to a sequence of measurements [23].

Quantitatively, no single parameter completely characterizes mixed state entanglement the way E does for pure states. For a generic mixed state, we do not know how to distill out of the mixed state as much pure entanglement (e.g., standard singlets) as was required to prepare the state in the first place; moreover, for some mixed states, entanglement can be distilled with the help of two-way communication between Alice and Bob, but not with one-way communication. In order to deal with these complications, we introduce three entanglement measures $D_1(M)\leqslant D_2(M)\leqslant E(M)$, each of which reduces to E for pure states, but at least two of which (D_1 and D_2) are known to be inequivalent for a generic mixed state.

Our fundamental measure of entanglement, for which we continue to use the symbol E, will be a mixed state's *entanglement of formation* $E(M)$, defined as the least expected entanglement of any ensemble of pure states realizing M. We show that local actions and classical communication cannot increase the expectation of $E(M)$ and we give exact expressions for the entanglement of formation of a simple class of mixed states: states of two spin-$\frac{1}{2}$ particles that are diagonal in the so-called *Bell basis*. This basis consists of four maximally entangled states — the singlet state of Eq. (1), and the three triplet states

$$\Psi^+=\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle+|\downarrow\uparrow\rangle), \tag{3}$$

$$\Phi^\pm=\frac{1}{\sqrt{2}}(|\uparrow\uparrow\rangle\pm|\downarrow\downarrow\rangle). \tag{4}$$

We also give lower bounds on the entanglement of formation of other, more general mixed states. Nonzero $E(M)$ will again serve as our qualitative criterion of nonlocality; thus, a mixed state will be considered local if can be expressed as a mixture of product states, and nonlocal if it cannot.

By *distillable entanglement* we will mean the asymptotic yield of arbitrarily pure singlets that can be prepared locally from mixed state M by entanglement purification protocols (EPPs) involving one-way or two-way communication between Alice and Bob. Distillable entanglement for one- and two-way communication will be denoted $D_1(M)$ and $D_2(M)$, respectively. Except in cases where we have been able to prove that D_1 or D_2 is identically zero, we have no explicit values for distillable entanglement, but we will exhibit various upper bounds, as well as lower bounds given by the yield of particular purification protocols.

D. Entanglement purification and quantum error correction

Entanglement purification protocols will be the subject of a large portion of this paper; we describe them briefly here. The most powerful protocols, depicted in Fig. 2, involve two-way communication. Alice and Bob begin by sharing a bipartite mixed state $\mathbf{M}=(M)^n$ consisting of n entangled pairs of particles each described by the density matrix M, then proceed by repeated application of three steps: (1) Alice and Bob perform unitary transformations on their states; (2) they perform measurements on some of the particles; and (3) they share the results of these measurements, using this information to choose which unitary transformations to perform in the next stage. The object is to sacrifice some of the particles, while maneuvering the others into a close approximation of a maximally entangled state such as $\mathbf{\Upsilon}=(\Psi^-)^m$, the tensor product of m singlets, where $0<m<n$. No generality is lost by using only unitary transformations and von Neumann measurements in steps (1) and (2), because Alice and Bob are free at the outset to enlarge the Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ to include whatever ancillas they might need to perform nonunitary operations and generalized measurements on the original systems.

A restricted version of the purification protocol involving only one-way communication is illustrated in Fig. 3. Here, without loss of generality, we permit only one stage of unitary operation and measurement, followed by a one-way

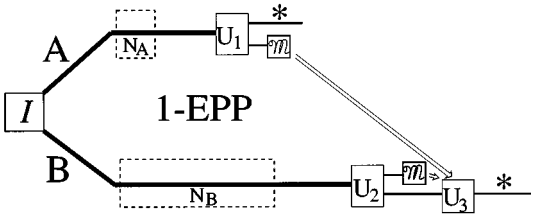

FIG. 3. One-way entanglement purification protocol (1-EPP). In 1-EPP there is only one stage; after unitary transformation U_1 and measurement $\mathcal{M}$, Alice sends her classical result to Bob, who uses it in combination with his measurement result to control a final transformation U_3. The unidirectionality of communication allows the final, maximally entangled state (*) to be separated both in space and in time.

classical communication. The principal advantage of such a protocol is that the components of the resulting purified maximally entangled state indicated by * can be separated both in space *and in time*. In Secs. V and VI we show that the time-separated EPR pairs resulting from such a one-way protocol (1-EPP) always permit the creation of a quantum error-correction code (QECC) whose rate and fidelity are, respectively, equal to the yield m/n and fidelity of the purified states produced by the 1-EPP.

The link between 1-EPP and QECC is provided by quantum teleportation [5]. As Fig. 4 illustrates, the availability of the time-separated EPR state (*) means that an arbitrary quantum state $|\xi\rangle$ (in a Hilbert space no larger than 2^m) can be teleported forward in time: the teleportation is initiated with Alice's Bell measurement $\mathcal{M}_4$, and is completed by Bob's unitary transformation U_4. The net effect is that an exact replica of $|\xi\rangle$ reappears at the end, despite the presence of noise ($N_{A,B}$) in the intervening quantum environment. Moreover, we will show in detail in Sec. VI that the protocol of Fig. 4 can be converted into a much simpler protocol with the same quantum communication capacity but involving neither entanglement nor classical communication, and having the topology of a quantum error correcting code (Fig. 16) [8–16].

Many features of mixed-state entanglement, along with their consequences for noisy-channel coding, are illustrated by a particular mixed state, the Werner state [21],

$$W_{5/8}=\frac{5}{8}|\Psi^-\rangle\langle\Psi^-|+\frac{1}{8}(|\Psi^+\rangle\langle\Psi^+|+|\Phi^+\rangle\langle\Phi^+| +|\Phi^-\rangle\langle\Phi^-|). \tag{5}$$

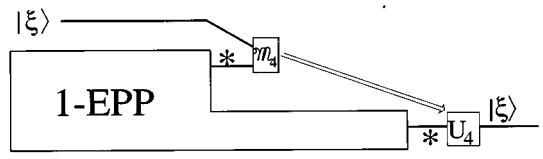

FIG. 4. If the 1-EPP of Fig. 3 is used as a module for creating time-separated EPR pairs (*), then by using quantum teleportation [5], an arbitrary quantum state $|\xi\rangle$ may be recovered exactly after U_4, despite the presence of intervening noise. This is the desired effect of a quantum error-correcting code (QECC).

This state, a 5/8 vs 3/8 singlet-triplet mixture, can be produced by mixing equal amounts of singlets and random uncorrelated spins, or equivalently by sending one spin of an initially pure singlet through a 50% depolarizing channel. (An x-depolarizing channel is one in which a state is transmitted unaltered with probability $1-x$ and is replaced with a completely random qubit with probability x.) These recipes suggest that $E(W_{5/8})$, the amount of pure entanglement required to prepare a Werner state, might be 0.5, but we show (Sec. II) that in fact that $E(W_{5/8})\approx 0.117$. The Werner state $W_{5/8}$ is also remarkable in that pure entanglement can be distilled from it by two-way protocols but not by any one-way protocol. In terms of noisy-channel coding, this means that a 50% depolarizing channel, which has a positive capacity for transmitting classical information, has zero capacity for transmitting intact quantum states if used in a one-way fashion, even with the help of quantum error-correcting codes. This will be proved in Sec. IV. If the same channel is used in a two-way fashion, or with the help of two-way classical communication, it has a positive capacity due to the nonzero distillable entanglement $D_2(W_{5/8})$, which is known to lie between 0.004 57 and 0.117 pure singlets out per impure pair in. The lower bound is from an explicit 2-EPP, while the upper bound comes from the known entanglement of formation, which is always an upper bound on distillable entanglement.

The remainder of this paper is organized as follows. Section II contains our results on the entanglement of formation of mixed states. Section III explains purification of pure, maximally entangled states from mixed states. Section IV exhibits a class of mixed states for which $D_1=0$ but $D_2>0$. Section V shows the relationship between mixed states and quantum channels. Section VI shows how a class of quantum error-correction codes may be derived from one-way purification protocols and contains our efficient 5-qubit code. Finally, Sec. VII reviews several important remaining open questions.

II. ENTANGLEMENT OF FORMATION

A. Justification of the definition

As noted above, we define the entanglement of formation $E(M)$ of a mixed state M as the least expected entanglement of any ensemble of pure states realizing M. The point of this subsection is to show that the designation ''entanglement of formation'' is justified: in order for Alice and Bob to create the state M without transferring quantum states between them, they must already share the equivalent of $E(M)$ pure singlets; moreover, if they do share this much entanglement already, then they will be able to create M. (Both of these statements are to be taken in the asymptotic sense explained in the Introduction.) In this sense $E(M)$ is the amount of entanglement needed to create M.

Consider any specific ensemble of pure states that realizes the mixed state M. By means of the asymptotically entanglement-conserving mapping between arbitrary pure states and singlets [20], such an ensemble provides an asymptotic recipe for locally preparing M from a number of singlets equal to the mean entanglement of the pure states in the ensemble. Clearly some ensembles are more economical than others. For example, the totally mixed state of two qubits can be prepared at zero cost, as an equal mixture of four

product states, or at unit cost, as an equal mixture of the four Bell states. The quantity $E(M)$ is the minimum cost in this sense. However, this fact does not yet justify calling $E(M)$ the entanglement of formation, because one can imagine more complicated recipes for preparing M: Alice and Bob could conceivably start with an initial mixture whose expected entanglement is less than $E(M)$ and somehow, by local actions and classical communication, transform it into another mixture with greater expected entanglement. We thus need to show that such entanglement-enhancing transformations are not possible. We start by summarizing the definitions that lead to $E(M)$.

Definition: The entanglement of formation of a bipartite pure state Υ is the von Neumann entropy $E(\Upsilon) = S(\mathrm{Tr}_A|\Upsilon\rangle\langle\Upsilon|)$ of the reduced density matrix as seen by Alice or Bob [see Eq. (2)].

Definition: The entanglement of formation $E(\mathcal{E})$ of an ensemble of bipartite pure states $\mathcal{E}=\{p_i,\Upsilon_i\}$ is the ensemble average $\Sigma_i p_i E(\Upsilon_i)$ of the entanglements of formation of the pure states in the ensemble.

Definition: The entanglement of formation $E(M)$ of a bipartite mixed state M is the minimum of $E(\mathcal{E})$ over ensembles $\mathcal{E}=\{p_i,\Upsilon_i\}$ realizing the mixed state: $M=\Sigma_i p_i|\Upsilon_i\rangle\langle\Upsilon_i|$.

We now prove that $E(M)$ is nonincreasing under local operations and classical communication. First we prove two lemmas about the entanglement of bipartite *pure* states under local operations by one party, say Alice. Any such local action can be decomposed into four basic kinds of operations: (i) appending an ancillary system not entangled with Bob's part, (ii) performing a unitary transformation, (iii) performing an orthogonal measurement, and (iv) throwing away, i.e., tracing out, part of the system. (There is no need to add generalized measurements as a separate category, since such measurements can be constructed from operations of the above kinds.) It is clear that neither of the first two kinds of operation can change the entanglement of a pure state shared by Alice and Bob: the entanglement in these cases remains equal to the von Neumann entropy of Bob's part of the system. However, for the last two kinds of operations, the entanglement can change. In the following two lemmas we show that the expected entanglement in these cases cannot increase.

Lemma: If a bipartite pure state Υ is subjected to a measurement by Alice, giving outcomes k with probabilities p_k, and leaving residual bipartite pure states Υ_k, then the expected entanglement of formation $\Sigma_k p_k E(\Upsilon_k)$ of the residual states is no greater than the entanglement of formation $E(\Upsilon)$ of the original state.

$$\sum_k p_k E(\Upsilon_k) \leq E(\Upsilon). \tag{6}$$

Proof. Because the measurement is performed locally by Alice, it cannot affect the reduced density matrix seen by Bob. Therefore the reduced density matrix seen by Bob before measurement, $\rho=\mathrm{Tr}_A|\Upsilon\rangle\langle\Upsilon|$, must equal the ensemble average of the reduced density matrices of the residual states after measurement: $\rho_k=\mathrm{Tr}_A|\Upsilon_k\rangle\langle\Upsilon_k|$ after measurement. It is well known that von Neumann entropy, like classical Shannon entropy, is convex, in the sense that the entropy of a weighted mean of several density matrices is no less than the corresponding mean of their separate entropies [24]. Therefore

$$S(\rho) \geq \sum_k p_k S(\rho_k). \tag{7}$$

But the left-hand side of this expression is the original pure state's entanglement before measurement, while the right-hand side is the expected entanglement of the residual pure states after measurement. QED.

Lemma: Consider a tripartite pure state Υ, in which the parts are labeled A, B, and C. (We imagine Alice holding parts A and C and Bob holding part B.) Let $M=\mathrm{Tr}_C|\Upsilon\rangle\langle\Upsilon|$. Then $E(M)\leq E(\Upsilon)$, where the latter is understood to be the entanglement between Bob's part B and Alice's part AC. That is, Alice cannot increase the minimum expected entanglement by throwing away system C.

Proof. Again, whatever pure-state ensemble one takes as the realization of the mixed state M, the entropy at Bob's end of the *average* of these states must equal $E(\Upsilon)$, because the density matrix held by Bob has not changed. By the above argument, then, the average of the entropies of the reduced density matrices associated with these pure states cannot exceed the entropy of Bob's overall density matrix; that is, $E(M)\leq E(\Upsilon)$. QED.

We now prove a theorem that extends both of the above results to mixed states.

Theorem: If a bipartite mixed state M is subjected to an operation by Alice, giving outcomes k with probabilities p_k, and leaving residual bipartite mixed states M_k, then the expected entanglement of formation $\Sigma_k p_k E(M_k)$ of the residual states is no greater than the entanglement of formation $E(M)$ of the original state.

$$\sum_k p_k E(M_k) \leq E(M). \tag{8}$$

(If the operation is simply throwing away part of Alice's system, then there will be only one value of k, with unit probability.)

Proof. Given mixed state M there will exist some minimal-entanglement ensemble

$$\mathcal{E}=\{p_j,\Upsilon_j\} \tag{9}$$

of pure states realizing M.

For any ensemble $\mathcal{E}'$ realizing M,

$$E(M)\leq E(\mathcal{E}'). \tag{10}$$

Applying the above lemmas to each pure state in the minimal-entanglement ensemble $\mathcal{E}$, we get, for each j,

$$\sum_k p_{k|j} E(M_{jk}) \leq E(\Upsilon_j), \tag{11}$$

where M_{jk} is the residual state if pure state Υ_j is subjected to Alice's operation and yields result k, and $p_{k|j}$ is the conditional probability of obtaining this outcome when the initial state is Υ_j.

Note that when the outcome k has occurred the residual mixed state is described by the density matrix

$$M_k=\sum_j p_{j|k}M_{jk}. \tag{12}$$

Multiplying Eq. (11) by p_j and summing over j gives

$$\sum_{j,k} p_j p_{k|j}E(M_{jk})\leqslant\sum_j p_j E(\Upsilon_j)=E(M). \tag{13}$$

By Bayes theorem,

$$p_{j,k}=p_j p_{k|j}=p_k p_{j|k}, \tag{14}$$

Eq. (13) becomes

$$\sum_{j,k} p_k p_{j|k}E(M_{jk})\leqslant E(M). \tag{15}$$

Using the bound Eq. (10), we get

$$\sum_k p_k E(M_k)\leqslant\sum_k p_k\sum_j p_{j|k}E(M_{jk})\leqslant E(M). \tag{16}$$

QED.

Although the above theorem concerns a single operation by Alice, it evidently applies to any finite preparation procedure, involving local actions and one- or two-way classical communication, because any such procedure can be expressed as sequence of operations of the above type, performed alternately by Alice and Bob. Each measurement-type operation, for example, generates a new classical result, and partitions the before-measurement mixed state into residual after-measurement mixed states whose mean entanglement of formation does not exceed the entanglement of formation of the mixed state before measurement. Hence we may summarize the result of this section by saying that expected entanglement of formation of a bipartite system's state does not increase under local operations and classical communication. As noted in [20], entanglement itself *can* increase under local operations, even though its expectation cannot. Thus it is possible for Alice and Bob to gamble with entanglement, risking some of their initial supply with a chance of winning more than they originally had.

B. Entanglement of formation for mixtures of Bell states

In the preceding subsection it was shown that an ensemble of *pure states* with minimum average pure-state entanglement realizing a given density matrix defines a maximally economical way of creating that density matrix. In general it is not known how to find such an ensemble of minimally entangled states for a given density matrix M. We have, however, found such minimal ensembles for a particular class of states of two spin-$\frac{1}{2}$ particles, namely, mixtures that are diagonal when written in the Bell basis Eqs. (1), (3), and (4). We have also found a lower bound on $E(M)$ applicable to any mixed state of two spin-$\frac{1}{2}$ particles. We present these results in this subsection.

As a motivating example consider the Werner states of [21]. A Werner state is a state drawn from an ensemble of F parts pure singlet, and $(1-F)/3$ parts of each of the other Bell states — that is, a generalization of Eq. (5):

$$W_F=F|\Psi^-\rangle\langle\Psi^-|+\frac{1-F}{3}(|\Psi^+\rangle\langle\Psi^+|+|\Phi^+\rangle\langle\Phi^+|+|\Phi^-\rangle\langle\Phi^-|). \tag{17}$$

This is equivalent to saying it is drawn from an ensemble of $x=(4F-1)/3$ parts pure singlet, and $1-x$ parts the totally mixed ''garbage'' density matrix (equal to the identity operator)

$$G=I=\frac{1}{4}(|\Psi^+\rangle\langle\Psi^+|+|\Psi^-\rangle\langle\Psi^-|+|\Phi^+\rangle\langle\Phi^+|+|\Phi^-\rangle\langle\Phi^-|), \tag{18}$$

which was Werner's original formulation. We label these generalized Werner states W_F, with their F value, which is their fidelity or purity $\langle\Psi^-|W_F|\Psi^-\rangle$ relative to a perfect singlet (even though this fidelity is defined nonlocally, it can be computed from the results of local measurements, as $1-3P_\parallel/3$, where $P_\parallel$ is the probability of obtaining parallel outcomes if the two spins are measured along the same random axis).

It would take $x=(4F-1)/3$ pure singlets to create a mixed state W_F by directly implementing Werner's ensemble. One might assume that this prescription is the one requiring the least entanglement, so that the $W_{5/8}$ state would cost 0.5 ebits to prepare. However, through a numerical minimization technique we found four pure states, each having only 0.117 ebits of entanglement, that when mixed with equal probabilities create the $W_{5/8}$ mixed state much more economically. Below we derive an explicit minimally entangled ensemble for any Bell-diagonal mixed state W, including the Werner states W_F as a special case, as well as a giving a general lower bound for general mixed states M of a pair of spin-$\frac{1}{2}$ particles. For pure states and Bell-diagonal mixtures $E(M)$ is simply equal to this bound.

The lower bound is expressed in terms of a quantity $f(M)$ which we call the ''fully entangled fraction''of M and define as

$$f(M)=\max\langle e|M|e\rangle, \tag{19}$$

where the maximum is over all completely entangled states $|e\rangle$. Specifically, we will see that for all states of a pair of spin-$\frac{1}{2}$ particles, $E(M)\geqslant h[f(M)]$, where the function h is defined by

$$h(f)=\begin{cases} H[\frac{1}{2}+\sqrt{f(1-f)}] & \text{for } f\geqslant\frac{1}{2} \\ 0 & \text{for } f<\frac{1}{2}. \end{cases} \tag{20}$$

Here $H(x)=-x\log_2 x-(1-x)\log_2(1-x)$ is the binary entropy function. For mixtures of Bell states, the fully entangled fraction $f(M)$ is simply the largest eigenvalue of M.

We begin by considering the entanglement of a single pure state $|\phi\rangle$. It is convenient to write $|\phi\rangle$ in the following orthogonal basis of completely entangled states:

$$|e_1\rangle=|\Phi^+\rangle,$$

$$|e_2\rangle=i|\Phi^-\rangle,$$

$$|e_3\rangle=i|\Psi^+\rangle,$$

$$|e_4\rangle=|\Psi^-\rangle. \tag{21}$$

Thus we write

$$|\phi\rangle=\sum_{j=1}^{4}\alpha_j|e_j\rangle. \tag{22}$$

The entanglement of $|\phi\rangle$ can be computed directly as the von Neumann entropy of the reduced density matrix of either of the two particles. On doing this calculation, one finds that the entanglement of $|\phi\rangle$ is given by the simple formula

$$E=H[\tfrac{1}{2}(1+\sqrt{1-C^2})], \tag{23}$$

where $C=|\Sigma_j\alpha_j^2|$. (Note that one is squaring the complex numbers α_j, not their moduli.) E and C both range from 0 to 1, and E is a monotonically increasing function of C, so that C itself is a kind of measure of entanglement. According to Eq. (23), any *real* linear combination of the states $|e_j\rangle$ is another completely entangled state (i.e., $E=1$). In fact, *every* completely entangled state can be written, up to an overall phase factor, as a real linear combination of the $|e_j\rangle$'s. (To see this, choose α_1 to be real without loss of generality. Then if the other α_j's are not all real, C will be less than unity, and thus so will E.)

Note that if one of the α_j's, say α_1, is sufficiently large in magnitude, then the other α_j's will not have enough combined weight to make C equal to zero, and thus the state will have to have some entanglement. This makes sense: if one particular completely entangled state is sufficiently strongly represented in $|\phi\rangle$, then $|\phi\rangle$ itself must have some entanglement. Specifically, if $|\alpha_1|^2>\frac{1}{2}$, then because the sum of the squares of the three remaining α_j's cannot exceed $1-|\alpha_1|^2$ in magnitude, C must be at least $|\alpha_1|^2-(1-|\alpha_1|^2)$, i.e., $2|\alpha_1|^2-1$. It follows from Eq. (23) that E must be at least $H[\frac{1}{2}+\sqrt{|\alpha_1|^2(1-|\alpha_1|^2)}]$. That is, we have shown that

$$E(|\phi\rangle)\geq h(|\alpha_1|^2), \tag{24}$$

where h is defined in Eq. (20). This inequality will be very important in what follows.

As one might expect, the properties just described are not unique to the basis $\{|e_j\rangle\}$. Let $|e_j'\rangle=\Sigma_k R_{jk}|e_k\rangle$, where R is any real, orthogonal matrix (i.e., $R^TR=I$). We can expand $|\phi\rangle$ as $|\phi\rangle=\Sigma_j\alpha_j'|e_j'\rangle$, and the sum $\Sigma_j\alpha_j'^2$ is guaranteed to be equal to $\Sigma_j\alpha_j^2$ because of the properties of orthogonal transformations. Thus one can use the components α_j' in Eq. (23) just as well as the components α_j. In particular, the inequality (24) can be generalized by substituting for α_1 the component of $|\phi\rangle$ along *any* completely entangled state $|e\rangle$. That is, if we define $w=|\langle e|\phi\rangle|^2$ for some completely entangled $|e\rangle$, then

$$E(|\phi\rangle)\geq h(w). \tag{25}$$

We now move from pure states to mixed states. Consider an arbitrary mixed state M, and consider any ensemble $\mathcal{E}=p_k,\phi_k$ which is a decomposition of M into pure states

$$M=\sum_k p_k|\phi_k\rangle\langle\phi_k|. \tag{26}$$

For an arbitrary completely entangled state $|e\rangle$, let $w_k=|\langle e|\phi_k\rangle|^2$, and let $w=\langle e|M|e\rangle=\Sigma_k p_k w_k$. We can bound the entanglement of the ensemble (26) as follows:

$$E(\mathcal{E})=\sum_k p_k E(|\phi_k\rangle)\geq\sum_k p_k h(w_k)\geq h\left[\sum_k p_k w_k\right]=h(w). \tag{27}$$

This equation is true in particular for the minimal entanglement ensemble realizing M for which $E(M)=E(\mathcal{E})$. The second inequality follows from the convexity of the function h. Clearly we obtain the best bound of this form by maximizing $w=\langle e|M|e\rangle$ over all completely entangled states $|e\rangle$. This maximum value of w is what we have called the fully entangled fraction $f(M)$. We have thus proved that

$$E(M)\geq h[f(M)], \tag{28}$$

as promised.

To make the bound (28) more useful, we give the following simple algorithm for finding the fully entangled fraction f of an arbitrary state M of a pair of qubits. First, write M in the basis $\{|e_j\rangle\}$ defined in Eq. (21). In this basis, the completely entangled states are represented by the real vectors, so we are looking for the maximum value of $\langle e|M|e\rangle$ over all real vectors $|e\rangle$. But this maximum value is simply the largest eigenvalue of the real part of M. We have then: f is the maximum eigenvalue of Re M, when M is written in the basis of Eq. (21).

We now show that the bound (28) is actually achieved for two cases of interest: (i) pure states and (ii) mixtures of Bell states. That is, in these cases, $E(M)=h[f(M)]$.

(i) *Pure states*. Any pure state can be changed by local rotations into a state [25] of the form $|\phi\rangle=\alpha|\uparrow\uparrow\rangle+\beta|\downarrow\downarrow\rangle$, where $\alpha,\beta\geq 0$ and $\alpha^2+\beta^2=1$. Entanglement is not changed by such rotations, so it is sufficient to show that the bound is achieved for states of this form. For $M=|\phi\rangle\langle\phi|$, the completely entangled state maximizing $\langle e|M|e\rangle$ is $|\Phi^+\rangle$, and the value of f is $|\langle\Phi^+|\phi\rangle|^2=(\alpha+\beta)^2/2=\frac{1}{2}+\alpha\beta$. By straightforward substitution one finds that $h(\frac{1}{2}+\alpha\beta)=H(\alpha^2)$, which we know to be the entanglement of $|\phi\rangle$. Thus $E(M)=h[f(M)]$, which is what we wanted to show.

(ii) *Mixtures of Bell states.* Consider a mixed state of the form

$$W=\sum_{j=1}^{4}p_j|e_j\rangle\langle e_j|. \tag{29}$$

Suppose first that one of the eigenvalues p_j is greater than or equal to $\frac{1}{2}$, and without loss of generality take this eigenvalue to be p_1. The following eight pure states, mixed with equal probabilities, yield the state W:

$$\sqrt{p_1}|e_1\rangle+i(\pm\sqrt{p_2}|e_2\rangle\pm\sqrt{p_3}|e_3\rangle\pm\sqrt{p_4}|e_4\rangle). \tag{30}$$

Moreover, all of these pure states have the same entanglement, namely,

$$E = h(p_1). \tag{31}$$

[See Eq. (23).] Therefore the average entanglement of the mixture is also $\langle E\rangle = h(p_1)$. But $p_1 = f(W)$ for this density matrix, so for this particular mixture, we have $\langle E\rangle = h[f(W)]$. Since the right-hand side is our lower bound on E, this mixture must be a minimum-entanglement decomposition of W, and thus $E(W) = h[f(W)]$.

If none of the eigenvalues p_j is greater than $\frac{1}{2}$, then there exist phase factors θ_i such that $\Sigma_j p_j e^{i\theta_j} = 0$. In that case we can express W as an equal mixture of a different set of eight states:

$$\sqrt{p_1}e^{i\theta_1/2}|e_1\rangle \pm \sqrt{p_2}e^{i\theta_2/2}|e_2\rangle \pm \sqrt{p_3}e^{i\theta_3/2}|e_3\rangle \pm \sqrt{p_4}e^{i\theta_4/2}|e_4\rangle. \tag{32}$$

For each of these states, the quantity C [Eq. (23)] is equal to zero, and thus the entanglement is zero. It follows that $E(W) = 0$, so that again the bound is achieved. (The bound $h[f(W)]$ is zero in this case because f, the greatest of the p_j's, is less than $\frac{1}{2}$.)

It is interesting to ask whether the bound $h[f(M)]$ is in fact *always* equal to $E(M)$ for general mixed states M, not necessarily Bell-diagonal. It turns out that it is not. Consider, for example, the mixed state

$$M = \tfrac{1}{2}|\uparrow\uparrow\rangle\langle\uparrow\uparrow| + \tfrac{1}{2}|\Psi^+\rangle\langle\Psi^+|. \tag{33}$$

The value of f for this state is $\frac{1}{2}$, so that $h(f) = 0$. And yet, as we now show, it is impossible to build this state out of unentangled pure states; hence $E(M)$ is greater than zero and is not equal to $h(f)$.

To see this, let us try to construct the density matrix of Eq. (33) out of unentangled pure states. That is, we want

$$M = \sum_k p_k |\phi_k\rangle\langle\phi_k|, \tag{34}$$

where each $|\phi_k\rangle$ is unentangled. That is, each $|\phi_k\rangle$ is such that when we write it in the basis of Eq. (21), i.e., as $|\phi_k\rangle = \Sigma_{j=1}^4 \alpha_{k,j}|e_j\rangle$, the α's satisfy the condition

$$\sum_{j=1}^{4} \alpha_{k,j}^2 = 0. \tag{35}$$

Now the density matrix M, when written in the $|e_j\rangle$ basis, looks like this:

$$M = \begin{bmatrix} \frac{1}{4} & \frac{i}{4} & 0 & 0 \\ \frac{-i}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{36}$$

Thus, in order for Eq. (34) to be true, the α's must be consistent with the following conditions:

$$\begin{aligned} \sum_k p_k|\alpha_{k,1}|^2 &= \tfrac{1}{4}, \\ \sum_k p_k|\alpha_{k,2}|^2 &= \tfrac{1}{4}, \\ \sum_k p_k|\alpha_{k,3}|^2 &= \tfrac{1}{2}, \\ \sum_k p_k|\alpha_{k,4}|^2 &= 0, \\ \sum_k p_k\alpha_{k,1}\alpha_{k,2}^* &= \frac{i}{4}. \end{aligned} \tag{37}$$

Evidently all the $\alpha_{k,4}$'s are equal to zero. By Eq. (35) the remaining α's satisfy

$$|\alpha_{k,1}|^2 + |\alpha_{k,2}|^2 \geqslant |\alpha_{k,3}|^2 \quad \text{for every } k. \tag{38}$$

In fact, the ''$\geqslant$'' of this last relation must be an equality, or else the sum conditions of Eq. (37) would not work out. That is,

$$|\alpha_{k,1}|^2 + |\alpha_{k,2}|^2 = |\alpha_{k,3}|^2 \quad \text{for every } k. \tag{39}$$

Combining this last equation with Eq. (35), we arrive at the conclusion that for each k, the ratio of $\alpha_{k,1}$ to $\alpha_{k,2}$ is real. But in that case there is no way to generate the imaginary sum required by the last of the conditions (37). It is thus impossible to build M out of unentangled pure states; that is, $E(M) > 0$.

We conclude, then, that our bound is only a bound and not an exact formula for E. It turns out, in fact, that there are two other ways to prove that the state M has nonzero entanglement of formation. Peres [26] and Horodecki *et al.* [27] have recently developed a general test for nonzero entanglement for states of two qubits and has applied it explicitly to states like our M, showing that $E(M)$ is nonzero. Also, in Sec. III B 2 below, we show that one can distill some pure entanglement from M, which would not be possible if $E(M)$ were zero.

III. PURIFICATION

Suppose Alice and Bob have n pairs of particles, each pair's state described by a density matrix M. Such a mixed state results if one or both members of an initially pure Bell state is subjected to noise during transmission or storage (cf. Fig. 1). Given these n impure pairs, how many pure Bell singlets can they distill by local actions; indeed, can they distill any at all? In other words, how much entanglement can they ''purify'' out of their mixed state without further use of a quantum channel to share more entanglement?

The complete answer is not yet known, but upper and lower bounds are [17]. An upper bound is $E(M)$ per pair, because if Alice and Bob could get more good singlets than that they could use them to create more mixed states with

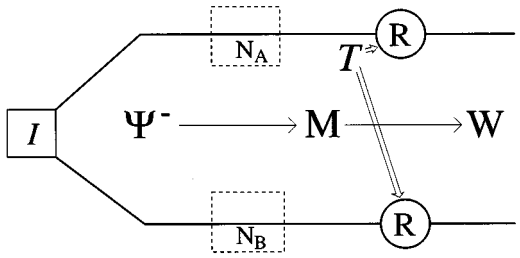

FIG. 5. The general mixed state M of Fig. 1 can be converted into one of the Werner form W_F of Eq. (17) if the particles on both Alice's and Bob's side are subjected to the same random rotation R [we refer to the act of choosing a random SU(2) rotation and applying it to both particles as a ''twirl'' T].

density matrix M than the number with which they started, thereby increasing their entanglement by local operations, which we have proven impossible (Sec. II A). Lower bounds are given by construction. We have found specific procedures which Alice and Bob can use to purify certain types of mixed states into a lesser number of pure singlets. We call these schemes *entanglement purification protocols* (EPPs), which should not be confused with the *purifications* of a mixed state of [28].

A. Purification basics

Our purification procedures all stem from a few simple ideas.

(1) A general two-particle mixed state M can be converted to a Werner state W_F [Eq. (17)] by an irreversible preprocessing operation which increases the entropy $[S(W_F)>S(M)]$, perhaps wasting some of its recoverable entanglement, but rendering the state easier to deal with because it can thereafter be regarded as a classical mixture of the four orthogonal Bell states [Eqs. (1), (3), and (4)] [29]. The simplest such preprocessing operation, a *random bilateral rotation* [17] or ''twirl,'' consists of choosing an independent, random SU(2) for each impure pair and applying it to both members of the pair (cf. Fig. 5). Because of the singlet state's invariance under bilateral rotation, twirling has the effect of removing off-diagonal terms in the two-particle density matrix in the Bell basis, as well as equalizing the triplet eigenvalues. Actually, removing the off-diagonal terms is sufficient as all of our EPPs operate successfully (with only minor modification) on a Bell-diagonal mixed state W with, in general, unequal triplet eigenvalues. Equalization of the triplet eigenvalues only adds unnecessary entropy to the mixture. In Appendix A it is shown that a continuum of rotations is unnecessary: an arbitrary mixed state of two qubits can be converted into a Werner W_F or Bell-diagonal W mixture by a ''discrete twirl,'' consisting of a random choice among an appropriate discrete set of bilateral rotations [30]. We use T to denote the nonunitary operation of performing either a discrete or a continuous twirl.

(2) Once the initial mixed state M has been rendered into Bell-diagonal form W, it can be purified as if it were a classical mixture of Bell states, without regard to the original mixed state M or the noisy channel(s) that may have generated it [31]. This is extremely convenient for the development of all our protocols. However, as we show in Appendix B, all the purification protocols we will develop will also work just as well on the original non-Bell-diagonal mixtures M.

(3) Bell states map onto one another under several kinds of local unitary operations (cf. Table I). These three sets of operations are of two types: *unilateral* operations, which are performed by Bob or Alice but not both, and *bilateral* operations, which can be written as a tensor product of an Alice part and a Bob part, each of which are the same. The three types of operations used are (1) unilateral rotations by π rad, corresponding to the three Pauli matrices σ_x, σ_y, and σ_z; (2) bilateral rotations by $\pi/2$ rad, henceforth denoted B_x, B_y, and B_z; and (3) the bilateral application of the two-bit quantum XOR (or controlled- NOT) [32,33], hereafter referred to as the BXOR operation (see Fig. 6). These operations and the Bell state mappings they implement, along with individual particle measurements, are the basic tools Alice and Bob use to purify singlets out of W.

(4) Alice and Bob can distinguish Φ states from Ψ states by locally measuring their particles along the z direction. If they get the same results they have a Φ; if they get opposite results they have a Ψ. Note that if only one of the observers (say, Bob) needs to know whether the state was a Φ or a Ψ, the process can be done without two-way communication. Alice simply makes her measurement and sends the result to Bob. After Bob makes his measurement, he can then determine whether the state had been a Φ or a Ψ by comparing his measurement result with Alice's, without any further communication.

(5) For convenience we take $|\Phi^+\rangle$ as the standard state for the rest of the paper. This is because it is the state which, when used as both source and target in a BXOR, remains unchanged. It is not necessary to use this convention but it is algebraically simpler. We note that $|\Phi^+\rangle$ states can be converted to singlet ($|\Psi^-\rangle$) states using the unilateral σ_y rotation, as shown in Table I. The only complication is that the nonunitary twirling operation T of item 1 works only when $|\Psi^-\rangle$ is taken as the standard state. But a modified twirl T' which leaves $|\Phi^+\rangle$ invariant and randomizes the other three Bell states may easily be constructed: the modified twirl would consist of a unilateral σ_y (which swaps the $|\Phi^+\rangle$'s and $|\Psi^-\rangle$'s) followed by a conventional twirl T, followed by another unilateral σ_y (which swaps them back).

(6) The preceding points all suggest a new notation for the Bell states. We use two classical bits to label each of the Bell states and write

$$\Phi^+ = 00,$$

$$\Psi^+ = 01,$$

$$\Phi^- = 10,$$

$$\Psi^- = 11. \tag{40}$$

The right, low-order or ''amplitude'' bit identifies the Φ/Ψ property of the Bell state, while the left, high-order or ''phase'' bit identifies the $+/-$ property. Both properties could be distinguished simultaneously by a nonlocal measurement, but local measurements can only distinguish one

TABLE I. The unilateral and bilateral operations used by Alice and Bob to map Bell states to Bell states. Each entry of the BXOR table has two lines, the first showing what happens to the source state, the second showing what happens to the target state.

		Source			
		Ψ^-	Φ^-	Φ^+	Ψ^+
Unilateral π rotations:	I	Ψ^-	Φ^-	Φ^+	Ψ^+
	σ_x	Φ^-	Ψ^-	Ψ^+	Φ^+
	σ_y	Φ^+	Ψ^+	Ψ^-	Φ^-
	σ_z	Ψ^+	Φ^+	Φ^-	Ψ^-

		Source			
		Ψ^-	Φ^-	Φ^+	Ψ^+
Bilateral $\pi/2$ rotations:	I	Ψ^-	Φ^-	Φ^+	Ψ^+
	B_x	Ψ^-	Φ^-	Ψ^+	Φ^+
	B_y	Ψ^-	Ψ^+	Φ^+	Φ^-
	B_z	Ψ^-	Φ^+	Φ^-	Ψ^+

		Source				
	Target	Ψ^-	Φ^-	Φ^+	Ψ^+	
Bilateral XOR:		Ψ^+	Φ^+	Φ^-	Ψ^-	(source)
	Ψ^-	Φ^-	Ψ^-	Ψ^-	Φ^-	(target)
		Ψ^+	Φ^+	Φ^-	Ψ^-	(source)
	Φ^-	Ψ^-	Φ^-	Φ^-	Ψ^-	(target)
		Ψ^-	Φ^-	Φ^+	Ψ^+	(source)
	Φ^+	Ψ^+	Φ^+	Φ^+	Ψ^+	(target)
		Ψ^-	Φ^-	Φ^+	Ψ^+	(source)
	Ψ^+	Φ^+	Ψ^+	Ψ^+	Φ^+	(target)

of the properties at a time, randomizing the other. For example, a bilateral z spin measurement distinguishes the amplitude while randomizing the phase.

B. Purification protocols

We now present several two- and one-way purification protocols. All begin with a large collection of n impure pairs each in mixed state M, use up $n-m$ of them (by measurement), while maneuvering the remaining m pairs into a collective state M' whose fidelity $\langle(\Phi^+)^m|M'|(\Phi^+)^m\rangle$ relative to a product of m standard Φ^+ states approaches 1 in the limit of large n. The yield a purification protocol P on input mixed states M is defined as

$$D_P(M) = \lim_{n\to\infty} m/n. \tag{41}$$

If the original impure pairs M arise from sharing pure EPR pairs through a noisy channel χ, then the yield $D_P(M)$ defines the asymptotic number of qubits that can be reliably transmitted (via teleportation) per use of the channel. For one-way protocols the yield is equal to the rate of a corresponding quantum error-correcting code (cf. Sec. V). For two-way protocols, there is no corresponding quantum error-correcting code. We will compare the yields from our protocols with the rates of quantum error-correcting codes introduced by other authors, and with known upper bounds on the one-way and two-way distillable entanglements $D_1(W)$ and $D_2(W)$. These are defined in the obvious way, e.g., $D_1(W) = \max\{D_P(W): P \text{ is a } 1-\text{EPP}\}$. No entanglement purification protocol has been proven optimal, but all give lower bounds on the amount of entanglement that can be distilled from various mixed states.

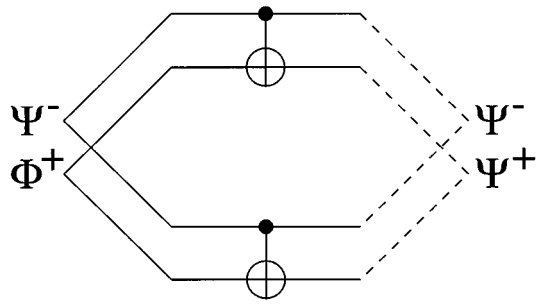

FIG. 6. The BXOR operation. A solid dot indicates the source bit of an XOR operation [32] and a crossed circle indicates the target. In this example a Ψ^- state is the source and a Φ^+ is the target. If the pairs are later brought back together and measured in the Bell basis the source will remain a Ψ^- and the target will have become a Ψ^+, as per Table I.

1. Recurrence method

A purification procedure presented originally in [17] is the recurrence method. This is an explicitly two-way protocol.

TABLE II. Probabilities for each initial configuration of source and target in a pair of Bell states drawn from the same ensemble, and the resulting state configuration after a BXOR operation is applied. The final column shows whether the target state passes (P) or fails (F) the test for being parallel along the z axis (this is given by the rightmost bit of the target state after the BXOR). This table, ignoring the probability column, is just the BXOR table of Table I written in the bitwise notation of item (6) of Sec. III A.

	Initial		After BXOR		Test
Probability	S	T	S	T	result
p_{00}^2	00	00	00	00	P
$p_{00}p_{01}$	00	01	00	01	F
$p_{00}p_{10}$	00	10	10	10	P
$p_{00}p_{11}$	00	11	10	11	F
$p_{01}p_{00}$	01	00	01	01	F
p_{01}^2	01	01	01	00	P
$p_{01}p_{10}$	01	10	11	11	F
$p_{01}p_{11}$	01	11	11	10	P
$p_{10}p_{00}$	10	00	10	00	P
$p_{10}p_{01}$	10	01	10	01	F
p_{10}^2	10	10	00	10	P
$p_{10}p_{11}$	10	11	00	11	F
$p_{11}p_{00}$	11	00	11	01	F
$p_{11}p_{01}$	11	01	11	00	P
$p_{11}p_{10}$	11	10	01	11	F
p_{11}^2	11	11	01	10	P

Two states are drawn from an ensemble which is a mixture of Bell states with probabilities p_i where i labels the Bell states in our two-bit notation. (As noted earlier, if the original impure state is not Bell-diagonal, it can be made so by twirling). The 00 state is again taken to be the standard state and we take $p_{00}=F$. The two states are used as the source and target for the BXOR operation. Their initial states and probabilities, and states after the BXOR operation, are shown in Table II. Alice and Bob test the target states, and then separate the source states into the ones whose target states passed and the ones whose target state failed. Each of these subsets is a Bell state mixture, with new probabilities. These *a posteriori* probabilities for the ''passed'' subset are

$$p'_{00}=(p_{00}^2+p_{10}^2)/p_{\text{pass}}, \quad p'_{01}=(p_{01}^2+p_{11}^2)/p_{\text{pass}}, \quad p'_{10}=2p_{00}p_{10}/p_{\text{pass}}, \quad p'_{11}=2p_{01}p_{11}/p_{\text{pass}}, \tag{42}$$

with

$$p_{\text{pass}}=p_{00}^2+p_{01}^2+p_{10}^2+p_{11}^2+2p_{00}p_{10}+2p_{01}p_{11}. \tag{43}$$

Consider the situation where Alice and Bob begin with a large supply of Werner states W_F. They apply the preceding procedure and are left with a subset of states which passed and a subset which failed. For the members of the ''passed'' subset $p'_{00}>p_{00}$ for all $p_{00}>0.5$. The members of the ''failed'' subset have $p_{00}=p_{01}=p_{10}=p_{11}=1/4$. Since the entanglement E of this mixture is 0, it will clearly not be possible to extract any entanglement from the ''failed'' subset, so all members of this subset are discarded. Note that this is where the protocol explicitly requires two-way communication. Both Alice and Bob need to know the results of the test in order to determine which pairs to discard.

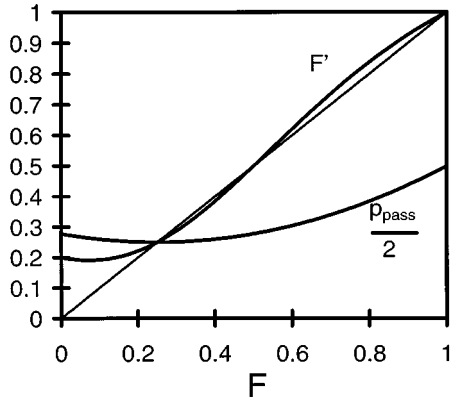

FIG. 7. Effect on the fidelity of Werner states of one step of purification, using the recurrence protocol. F is the initial fidelity of the Werner state [Eq. (17)], F' is the final fidelity of the ''passed'' pairs after one iteration. Also shown is the fraction $p_{pass}/2$ of pairs remaining after one iteration [cf. Eq. (43)].

The members of the ''passed'' subset have a greater p_{00} than those in the original set of impure pairs. The new density matrix is still Bell diagonal, but is no longer a Werner state W_F. Therefore, a twirl T' is applied (Sec. III A, items 1 and 5), leaving the p_{00} component alone and equalizing the others [34]. (It is appropriate in this situation to use the modified twirl T' which leaves Φ^+ invariant, as explained in item 5 of Sec. III A.) We are left with a new situation similar to the starting situation, but with a higher fidelity $F'=p'_{00}$. Figure 7 shows the resulting F' versus F. The process is then repeated; iterating the function of Fig. 7 will continue to improve the fidelity. This can be continued until the fidelity is arbitrarily close to 1. Macchiavello [34] has found that faster convergence can be achieved by substituting a deterministic bilateral B_x rotation for the twirl T'. With this modification, the density matrix remains Bell diagonal, but no longer has the Werner form W_F after the first iteration; nevertheless its p_{00} component increases more rapidly with successive iterations.

Even with this improvement the recurrence method is rather inefficient, approaching zero yield in the limit of high output fidelity, since in each iteration at least half the pairs are lost (one out of every two is measured, and the failures are discarded). Figure 7 shows the fraction of pairs lost on each iteration. A positive yield D_2, even in the limit of perfect output fidelity, can be obtained by switching over from the recurrence method to the hashing method, to be described in Sec. III B 3, as soon as so doing will produce more good singlets than doing another step of recurrence. The yield versus initial fidelity of these combined recurrence-hashing protocols is shown in Fig. 8. (See also Fig. 9.)

It is important to note that the recurrence-hashing method gives a positive yield of purified singlets from all Werner states with fidelity greater than 1/2. Werner states of fidelity 1/2 or less have $E=0$ and therefore can yield no singlets. The pure hashing and breeding protocols, described below, which are one-way protocols, work only down to

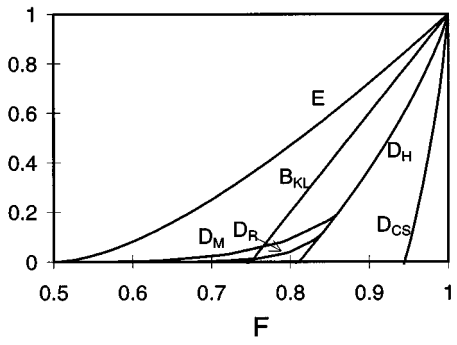

FIG. 8. Measures of entanglement versus fidelity F for Werner states W_F of Eq. (17). E is the entanglement of formation, Eq. (27). D_R is the yield of the recurrence method of Sec. III B 1 continued by the hashing method of (Sec. III B 4). D_M is the yield of the modified recurrence method of Macchiavello [34], continued by hashing. D_H is the yield of the one-way hashing and breeding protocols (Sec. III B 4) used alone. D_{CS} is the rate of the quantum error correcting codes proposed by Calderbank and Shor [10] and Steane [11]. B_{KL} is the upper bound for D_1 as shown in Sec. VI E (following Knill and Laflamme [40]).

$F\approx 0.8107$, and even the best known one-way protocol [35] works only down to $F\approx 0.8094$.

2. Direct purification of non-Bell-diagonal mixtures

Most of the purification strategies discussed in this paper assume that the state to be purified is first brought to the Werner form, or at least to Bell-diagonal form, by means of a twirling operation. As we have said, though, this strategy is somewhat wasteful and we use it only to make the analysis manageable. In this subsection we give a simple example showing how a state can be purified directly with no twirling. For this particular example, it happens that the purification is accomplished in a single step rather than in a series of steps that gradually raise the fidelity.

Consider again the state M of Eq. (33):

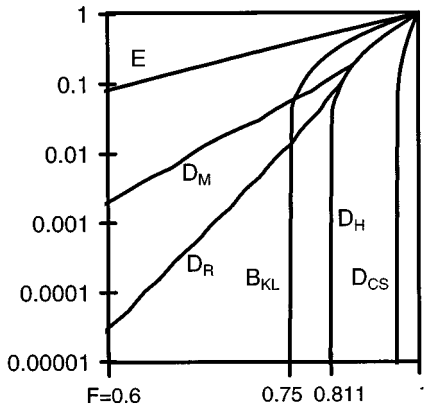

FIG. 9. The same as Fig. 8 exhibited on logarithmic scales. The value along the x axis is proportional to the logarithm of $(F-0.5)$. In this form it is clear that E, D_M, and D_R follow power laws $(F-0.5)^{\alpha}$. The ripples in D_M and D_R are real, and arise from the variable number of recurrence steps performed before switching over to the hashing protocol [17].

$$M=\tfrac{1}{2}|\uparrow\uparrow\rangle\langle\uparrow\uparrow|+\tfrac{1}{2}|\Psi^+\rangle\langle\Psi^+|. \tag{44}$$

Note that because the fully entangled fraction [Eq. (19)] $f=1/2$ for this state, it cannot be purified by the recurrence method. However, a collection of pairs in this state can be purified as using the following two-way protocol [36]: as in the recurrence method, Alice and Bob first perform the BXOR operation between pairs of pairs, and then bilaterally measure each target pair in the up-down basis. One can show that if the outcome of this measurement on a given target pair is ''down-down,'' then the corresponding source pair is left in the completely entangled state Ψ^+. Alice and Bob therefore keep the source pair only when they get this outcome, and discard it otherwise. The probability of getting the outcome ''down-down'' is 1/8, and since each target pair had to be sacrificed for the measurement, the yield from this procedure is $D_2=1/16$. The same strategy works for any state of the form

$$M=(1-p)|\uparrow\uparrow\rangle\langle\uparrow\uparrow|+p|\Psi^+\rangle\langle\Psi^+|, \tag{45}$$

with yield $D_2=p^2/4$.

A recent paper by Horodecki *et al.* [37] presents a more general approach to the purification of mixed states which, like the above scheme, does not start by bringing the states to Bell-diagonal form. Their strategy begins with a filtering operation aimed at increasing the fully entangled fraction f [Eq. (19)] of the surviving pairs; these pairs are then subjected to the recurrence procedure described above. These authors have shown that by this technique, one can distill some amount of pure entanglement from any state of two qubits having a nonzero entanglement of formation. In other words, they have obtained for such systems the very interesting result that if $E(M)$ is nonzero, then so is $D_2(M)$.

3. One-way hashing method

This protocol uses methods analogous to those of universal hashing in classical privacy amplification [38]. (We will give a self-contained treatment of this hashing scheme here.) Given a large number n of impure pairs drawn from a Bell-diagonal ensemble of known density matrix W, this protocol allows Alice and Bob to distill a smaller number $m\approx n[1-S(W)]$ of purified pairs (e.g., near-perfect Φ^+ states) whenever $S(W)<1$. In the limit of large n, the output pairs approach perfect purity, while the asymptotic yield m/n approaches $1-S(W)$. This hashing protocol supersedes our earlier breeding protocol [17], which we will review briefly in Sec. III B 4.

The hashing protocol works by having Alice and Bob each perform BXOR's and other local unitary operations (Table I) on corresponding members of their pairs, after which they locally measure some of the pairs to gain information about the Bell states of the remaining unmeasured pairs. By the correct choice of local operations, each measurement can be made to reveal almost one bit about the unmeasured pairs; therefore, by sacrificing slightly more than $nS(W)$ pairs, where $S(W)$ is the von Neumann entropy [see Eq. (2)] of the impure pairs, the Bell states of all the remaining unmeasured pairs can, with high probability, be ascertained. Then local unilateral Pauli rotations ($\sigma_{x,y,z}$) can be used to restore each unmeasured pair to the standard Φ^+ state.

The hashing protocol requires only one-way communication: after Alice finishes her part of the protocol, in the process having measured $n-m$ of her qubits, she is able to send Bob classical information which, when combined with his measurement results, enables him to transform his corresponding unmeasured qubits into near-perfect Φ^+ twins of Alice's unmeasured qubits, as shown in Fig. 3.

Let δ be a small positive parameter that will later be allowed to approach zero in the limit of large n. The initial sequence of n impure pairs can be conveniently represented by a $2n$-bit string x_0 formed by concatenating the two-bit representations [Eq. (40)] of the Bell states of the individual pairs, the sequence $\Psi^-\Phi^+\Phi^-$, for example, being represented 110010. The *parity* of a bit string is the modulo-2 sum of its bits; the parity of a subset s of the bits in a string x can be expressed as a Boolean inner product $s\cdot x$, i.e., the modulo-2 sum of the bitwise AND of strings s and x. For example $1101\cdot 0111=0$ in accord with the fact that there are an even number of ones in the subset consisting of the first, second, and fourth bits of the string 0111. Although the inner product $s\cdot x$ is a symmetric function of its two arguments, we use a slanted font for the first argument to emphasize its role as a subset selection index, while the second argument (in roman font) is the bit string representing an unknown sequence of Bell states to be purified.

The hashing protocol takes advantage of the following facts.

(1) The distribution P_{X_0} of initial sequences x_0, being a product of n identical independent distributions, receives almost all its weight from a set of $\approx 2^{nS(W)}$ ''likely'' strings. If the likely set $\mathcal{L}$ is defined as comprising the $2^{n(S(W)+\delta)}$ most probable strings in P_{X_0}, then the probability that the initial string x falls outside $\mathcal{L}$ is $O(\exp(-\delta^2 n))$ [6].

(2) As will be described in more detail later, the local Bell-preserving unitary operations of Table I (bilateral $\pi/2$ rotations, unilateral Pauli rotations, and BXOR's), followed by local measurement of one of the pairs, can be used to learn the parity of an arbitrary subset s of the bits in the unknown Bell-state sequence x, leaving the remaining unmeasured pairs in definite Bell states characterized by a two-bits-shorter string $f_s(x)$ determined by the initial sequence x and the chosen subset s.

(3) For any two distinct strings $x\neq y$, the probability that they agree on the parity of a random subset of their bit positions, i.e., that $s\cdot x=s\cdot y$ for random s, is exactly 1/2. This is an elementary consequence of the distributive law $(s\cdot x)\oplus(s\cdot y)=s\cdot(x\oplus y)$.

The hashing protocol consists of $n-m$ rounds of the following procedure. At the beginning of the $(k+1)$st round. $k=0,1,\ldots,n-m-1$, Alice and Bob have $n-k$ impure pairs whose unknown Bell state is described by a $2(n-k)$-bit string x_k. In particular, before the first round, the Bell sequence x_0 is distributed according to the simple *a priori* probability distribution P_{X_0} noted above. Then in the $(k+1)$st round, Alice first chooses and tells Bob a random $2(n-k)$-bit string s_k. Second, Alice and Bob perform local unitary operations and measure one pair to determine the subset parity $s_k\cdot x_k$, leaving behind $n-k-1$ unmeasured pairs in a Bell state described by the $[2(n-k)-2]$-bit string $x_{k+1}=f_{s_k}(x_k)$.

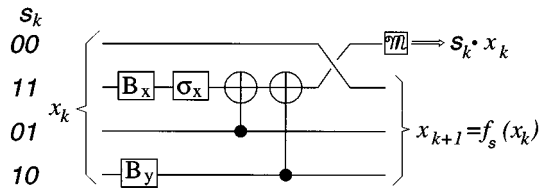

FIG. 10. Step k of the one-way hashing protocol, used to determine the parity $s_k\cdot x_k$, for an arbitrary unknown set of four Bell states represented by an unknown 8-bit string x relative to a known subset index string $s=00,11,01,10$. If bilateral measurement $\mathcal{M}$ yields a Ψ state (i.e., if the measurement result is 1), then half the candidates for x are excluded (e.g., $x=00,00,00,00$), but half are still allowed (e.g., $x=00,11,00,00$). For each allowed x, the after-measurement Bell states of the three remaining unmeasured pairs are a described by a 6-bit sequence $x_{k+1}=f_s(x_k)$ deterministically computable from x and s.

Consider the trajectories of two arbitrary but distinct strings $x_0\neq y_0$ under this procedure. Let x_k and y_k denote the images of x_0 and y_0, respectively, after k rounds, where the same sequence of operations $f_{s_0},f_{s_1},\ldots,f_{s_{n-m-1}}$, parameterized by the same random-subset index strings $s_0,s_1,\ldots,s_{n-m-1}$, is used for both trajectories. It can readily be verified that for any $r<n$ the probability

$$P((x_r\neq y_r)\&\forall_{k=0}^{r-1}(s_k\cdot x_k=s_k\cdot y_k)) \tag{46}$$

(i.e., the probability that x_r and y_r remain distinct while nevertheless having agreed on all r subset parities along the way, $s_k\cdot x_k=s_k\cdot y_k$ for $k=0,\ldots,r-1$) is at most 2^{-r}. This follows from the fact that at each iteration the probability that x and y remain distinct is $\leqslant 1$, while the probability that, if they were distinct at the beginning of the iteration they will give the same subset parity, is exactly 1/2. Recalling that the likely set $\mathcal{L}$ of initial candidates has only $2^{n[S(W)+\delta]}$ members, but with probability greater than $1-O(\exp(-\delta^2 n))$ includes the true initial sequence x_0, it is evident that after $r=n-m$ rounds, the probability of failure, i.e., of no candidate, or of more than one candidate, remaining at the end for x_m, is at most $2^{n[S(W)+\delta]-(n-m)}+O(\exp(-\delta^2 n))$. Here the first term upper-bounds the probability of more than one candidate surviving, while the second term upper-bounds the probability of the true x_0 having fallen outside the likely set. Letting $n-m=n[S(M)+2\delta]$ and taking $\delta\approx n^{-1/4}$, we get the desired result that the error probability approaches 0 and the yield m approaches $n[1-S(M)]$ in the limit of large n.

It remains to show how the local operations of Table I can be used to collect the parity of an arbitrary subset of bits of x into the amplitude bit of a single pair. We choose as the destination pair, into which we wish to collect the parity $s\cdot x$, that pair corresponding to the first nonzero bit of s. For example if $s=00,11,01,10$ (see Fig. 10), the destination will be the second pair of x_k. Our goal will be to make the amplitude bit of that pair after round k equal to the parity of both bits of the second pair, the right bit of the third pair, and the left bit of the fourth pair in the unknown input x_k. Pairs such as the first, having 00 in the index string s, have no effect on the desired subset parity, and accordingly are bypassed by all the operations described below.

The first step in collecting the parity is to operate separately on each of the pairs having a *01, 10,* or *11* in the index string, so as to collect the desired parity *for that pair* into the amplitude (right) bit of the pair. This can be achieved by doing nothing to pairs having *01* in the index string, performing a B_y on pairs having *10* (since B_y has the effect of interchanging the phase and amplitude bits of a Bell state), and performing the two rotations B_x and σ_x on pairs with *11* in the index string ($B_x\sigma_x=\sigma_xB_x$ has the effect of XORing a Bell state's phase bit into its amplitude bit).

The next step consists of BXORing all the pairs except those with *00* in the index string into the selected destination, in this case the second pair. The selected destination pair is used as the common target for all these BXOR's, causing its amplitude bit to accumulate the desired subset parity $s\cdot x$. This follows from the fact (cf. Table I) that the BXOR leaves the source's amplitude bit unaffected while causing the target's amplitude bit to become the XOR of the previous amplitude bits of source and target. Recall that phase bits behave oppositely under BXOR: the target's phase bit is unaffected while the source's phase bit becomes the XOR of the previous values of source and target phase bits; this ''back-action'' must be accounted for in determining the function f_s. Figure 10 illustrates this step of the hashing method on an unknown four-Bell-state sequence x using the subset index string $s=00,11,01,10$ mentioned before.

The hashing protocol distills a yield $D_H=1-S(W)$, which we have called D_0 in our previous work [17]. For the Werner channel, parametrized completely by F,

$$S(W_F)=-F\log_2(F)-(1-F)\log_2[(1-F)/3], \quad (47)$$

giving a positive yield for Werner states with $F\gtrsim 0.8107$. Figures 8 and 9 show $D_H(F)$, comparing it with E and with other purification protocols.

4. Breeding method

This protocol, introduced in Ref. [17], will not be described here in detail, as it has been superseded by the one-way hashing protocol described in the preceding section. The breeding protocol assumes that Alice and Bob have a shared pool of pure $|\Phi^+\rangle=00$ states, previously prepared by some other method (e.g., the recurrence method) and also a supply of Bell-diagonal impure states which they wish to purify. The protocol consumes the Φ^+ states from the pool, but, if the impure states are not too impure, produces more newly purified pairs than the number of pool states consumed (in the manner of a breeder reactor).

The basic step of breeding is very similar to that of hashing and is shown in Fig. 11. Again a random subset s of the amplitude and phase bits of the Bell states is selected. The parity of this selected set is again gathered up in exactly the same way, except that the target of the BXOR operations is one of the prepurified 00 states. The use of the pure target simplifies the action of the BXOR, in that the ''back action'' which changes the state of the source bits is avoided in this scheme. This means that the input string x can be restored to exactly its original value by a simple undoing of the one-qubit local operations, as shown. This offers the advantage that the (possibly very complicated) sequence of Boolean functions $f_{s_0}, f_{s_1}, \ldots, f_{s_{n-m-1}}$ do not have to be calculated in this case. Once again, the result of the parity measurement $\mathcal{M}$ is to reduce the number of candidates for x by almost exactly 1/2. Thus, by the same argument as before, after $n-m\approx nS(W)$ rounds of parity measurements, it is probable that x has been narrowed down to be just one member of the likely set $\mathcal{L}$. Thus, all n of these pairs can be turned into pure Φ^+ states; however, since $n-m$ pure Φ^+'s have been used up in the process, the net yield is $m/n=D_H(F)$, exactly the same as in the hashing protocol.

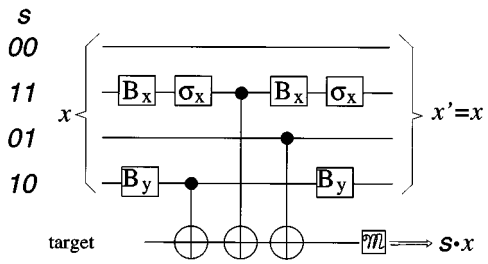

FIG. 11. Step k of the one-way breeding protocol. The scheme is very similar to the hashing protocol of Fig. 10, except that the target for the BXOR's is guaranteed to be a perfect Φ^+ state. This allows the one-bit operations to be undone so that there is no back-action on the string x.

IV. ONE-WAY *D* AND TWO-WAY *D* ARE PROVABLY DIFFERENT

It has already been noted that some of the entanglement purification schemes use two-way communication between the two parties Alice and Bob while others use only one-way communication. The difference is significant because one-way protocols can be used to protect quantum states during storage in a noisy environment, as well as during transmission through a noisy channel, while two-way protocols can only be used for the latter purpose (cf. Sec. VI). Thus it is important to know whether there are mixed states for which D_1 is properly less than D_2. Here we show that there are, and indeed that the original Werner state $W_{5/8}$ (i.e., the result of sharing singlets through a 50% depolarizing channel) cannot be purified at all by one-way protocols, even though it has a positive yield under two-way protocols.

To show this, consider an ensemble where a state preparer gives Alice n singlets, half shared with Bob and half shared with another person (Charlie). Alice is unaware of which pairs are shared with Bob and which with Charlie. Bob and Charlie are also given enough extra garbage particles (either randomly selected qubits or any state totally entangled with the environment but with no one else) so that they each have a total of n particles as well. This situation is diagrammed in Fig. 12. From Alice and Bob's point of view, each state has the density matrix $W_{5/8}$.

Alice, without hearing any information from Bob or Charlie, is supposed to do her half of a purification protocol and then send on classical data to the others. Therefore, each particle Alice has looks like a totally mixed state to her. By symmetry, anything she could do to assure herself that a particular particle is half of a good EPR pair shared with Bob will also assure her that the same particle is half of a good EPR pair shared with Charlie. No such three-sided EPR pair

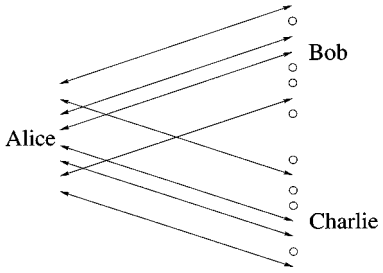

FIG. 12. A symmetric situation in which Bob and Charlie are each equally entangled with Alice. Two-headed arrows denote maximally entangled pairs, and open circles denote garbage states [Eq. (18)].

can exist. If she used it to teleport a qubit to Bob she would also have teleported it to Charlie, violating the no-cloning theorem [39]. Therefore, she cannot distill even one good EPR pair from an arbitrarily large supply of $W_{5/8}$ states. On the other hand the combined recurrence-hashing method (D_M in Fig. 9) gives a positive lower bound on the two-way yield $D_2(W_{5/8})>0.004\,57$, so we can write

$$D_1(W_{5/8})=0<0.004\,57\leqslant D_2(W_{5/8}). \tag{48}$$

It is also clear that any ensemble of Werner states can be reduced to one of lower fidelity by local action [combining with totally mixed states of Eq. (18)]. Therefore $D_1(W_F)=0$ for all $F<5/8$. Knill and Laflamme prove [40] that $D_1(W_F)=0$ for all $F<3/4$. In Sec. VI E we explain their proof and, using the argument of Sec. V B, obtain the bound

$$D_1<4F-3, \tag{49}$$

as shown in Figs. 8 and 9.

A similar argument can be used to show that for some ensembles D_1 is not symmetric, depending on whether it is Alice or Bob who starts the communication. Suppose in the symmetric situation of Fig. 12 that Bob and Charlie *know* which pairs are shared with Alice and which are garbage. For this ensemble the symmetry argument for Alice remains the same and $D_{A\to B}=0$. If the communication is from Bob to Alice, though, it is easy to see that he can use half of his particles, the ones he knows are good pairs shared with Alice. The other half are useless since they have $E=0$ and could have been manufactured locally. Thus we have $D_{B\to A}=1/2$ and $D_{A\to B}=0$.

Our no-cloning argument shows that Alice and Bob cannot generate good EPR pairs by applying a 1-EPP to the mixed state $W_{5/8}$ generated by sharing singlets through a 50% depolarizing channel. As a consequence, there is no quantum error-correcting code which can transmit unknown quantum states reliably through a 50% depolarizing channel, as will be shown in the next section.

V. NOISY CHANNELS AND BIPARTITE MIXED STATES

In preceding sections we have considered the preparation and purification of bipartite mixed states, and we have shown that two-way entanglement purification protocols can purify some mixed states that cannot be purified by any one-way protocol. When used in conjunction with teleportation, purification protocols, whether one-way or two-way, offer a means of transmitting quantum information faithfully via noisy channels; and one-way protocols, by producing time-separated entanglement, can additionally be used to protect quantum states during storage in a noisy environment. In this section we discuss the close relation between one-way entanglement purification protocols and the other well-known means of protecting quantum information from noise, namely quantum error-correcting codes (QECC) [8–16].

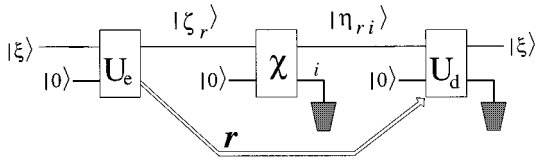

FIG. 13. A general one-way QECC. A classical side channel from Alice to Bob is allowed in addition to quantum channel χ.

A quantum channel χ, operating on states in an N-dimensional Hilbert space, may be defined as (cf. [9]) a unitary interaction of the input state with an environment, in which the environment is supplied in a standard pure initial state $|\mathbf{0}\rangle$ and is traced out (i.e., discarded) after the interaction to yield the channel output, generally a mixed state. The quantum capacity $Q(\chi)$ of such a channel is the maximum asymptotic rate of reliable transmission of unknown quantum states $|\xi\rangle$ in $\mathcal{H}_2$ through the channel that can be achieved by using a QECC to encode the states before transmission and decode them afterward.

As in quantum teleportation [5] we will also consider the possibility that the quantum channel is supplemented with classical communication. This leads us to define the augmented quantum capacities $Q_1(\chi)$ and $Q_2(\chi)$, of a channel supplemented by unlimited one- and two-way classical communication. For example, Fig. 13 shows a quantum error-correcting code, consisting of encoding transformation U_e and decoding transformation U_d, used to transmit unknown quantum states $|\xi\rangle$ reliably through the noisy quantum channel χ, with the help of a one-way classical side channel (operating in the same direction as the quantum channel). Perhaps surprisingly, this one-way classical channel provides no enhancement of quantum capacity:

$$Q_1=Q. \tag{50}$$

This will be shown in Sec. V A.

We consider also the case of a noisy quantum channel supplemented by a *noiseless* quantum channel. We will show in Sec. V B that the capacity of n uses of a noisy channel supplemented by m uses of a noiseless channel of unit capacity is no greater than the sum of their individual capacities, i.e., their quantum capacities are no more than additive. We have no similar result for the case of two different imperfect channels.

In contrast to Eq. (50) we will show that for many quantum channels two-way classical communication can be used to transmit quantum states through the channel at a rate $Q_2(\chi)$ considerably exceeding the one-way capacity $Q(\chi)$. This is typically done by using the channel to share EPR pairs between Alice and Bob, purifying the resulting bipartite

mixed states by a two-way entanglement purification protocol, then using the resulting purified pairs to teleport unknown quantum states $|\xi\rangle$ from Alice to Bob.

The analysis of Q and Q_2 is considerably simplified by the fact that an important class of noisy channels, including depolarizing channels, can be mapped in a one-to-one fashion onto a corresponding class of bipartite mixed states, with the consequence that the channel's quantum capacity $Q_1=Q$ is given by the one-way distillable entanglement D_1 of the mixed state, and vice versa. For example, a depolarizing channel of depolarization probability $p=1-x$ [cf. above Eq. (18)] corresponds to a Werner state W_F of fidelity $F=1-(3p/4)$ and has $Q=D_1(W_F)$ and $Q_2=D_2(W_F)$.

The correspondence between channels and mixed states is established by two functions, $\hat{M}(\chi)$ defining the bipartite mixed state obtained from channel χ and $\hat{\chi}(M)$ defining the channel obtained from bipartite mixed state M. The bipartite mixed state $\hat{M}(\chi)$ is obtained by preparing a standard maximally entangled state of two N-state subsystems,

$$\Upsilon=N^{-1/2}\sum_{i=1}^{N}|e_j\rangle\otimes|e_j\rangle, \tag{51}$$

and transmitting Bob's part through the channel χ. For example, a Werner state W_F, with $F=1-3p/4$, results when half a standard EPR pair is transmitted through a p-depolarizing channel.

The mapping in the other direction, from mixed states to channels, is obtained by teleportation. Given a bipartite mixed state M of two subsystems, each having Hilbert space of dimension N, the channel $\hat{\chi}(M)$ is defined by using mixed state M, instead of the standard maximally entangled state $|\Upsilon\rangle\langle\Upsilon|$, in a teleportation [5] channel (see Fig. 4). It can be readily shown that for Bell-diagonal mixed states the two mappings are mutually inverse $\hat{M}(\hat{\chi}(M))=M$; we shall call the channels corresponding to such mixed states ''generalized depolarizing channels.''

For more general channels and mixed states, the two mappings are not generally mutually inverse. For example, $\hat{\chi}(M)$, for the bipartite state $M=|\uparrow\uparrow\rangle\langle\uparrow\uparrow|$, is the $p=1$ depolarizing channel, and $\hat{M}(\hat{\chi}(M))=G$ of Eq. (18).

Nevertheless, two quite general inequalities will be demonstrated in Secs. V C and V D:

$$\forall_M \quad D_1(M)\geqslant Q(\hat{\chi}(M)) \tag{52}$$

and

$$\forall_\chi \quad D_1(\hat{M}(\chi))\leqslant Q(\chi). \tag{53}$$

If (as in the case of a Bell-diagonal state and its corresponding generalized depolarizing channel) the mapping is reversible, so that $M=\hat{M}(\chi)$ and $\chi=\hat{\chi}(M)$, the two inequalities are both satisfied, resulting in the equality mentioned earlier, viz.,

$$D_1(M)=Q(\chi). \tag{54}$$

Equation (52) follows from the ability, to be demonstrated in Sec. V C, to transform a QECC on $\hat{\chi}(M)$ into a 1-EPP on M; Eq. (53) follows, as shown in Sec. V D, from the fact that any 1-EPP on $\hat{M}(\chi)$, followed by quantum teleportation, results in a QECC on χ with a classical side channel.

A trivial extension of these arguments also shows that the corresponding results for two-way classical communication are true, namely,

$$\forall_M \quad D_2(M)\geqslant Q_2(\hat{\chi}(M)) \tag{55}$$

and

$$\forall_\chi \quad D_2(\hat{M}(\chi))\leqslant Q_2(\chi), \tag{56}$$

and if $\hat{M}(\hat{\chi}(M))=M$ then

$$D_2(M)=Q_2(\chi). \tag{57}$$

A. A forward classical side channel does not increase quantum capacity

To demonstrate Eq. (50), we note that any one-way protocol for transmitting $|\xi\rangle$ through channel χ can be described as in Fig. 13. The sender Alice codes $|\xi\rangle$ and an ancillary state $|\mathbf{0}\rangle$ using unitary transformation U_e. She then performs an incomplete measurement on the coded system giving classical results r which she sends on to Bob, the receiver. (If r contains any information about the quantum input $|\xi\rangle$ the strong no-cloning theorem [41] would prevent the original state from being recovered perfectly, even if the channel were noiseless. However, r might contain information on how the input $|\xi\rangle$ is coded.) She also sends the remaining quantum state through χ as encoded state $|\zeta_r\rangle$. The channel maps $|\zeta_r\rangle$ onto $|\eta_{ri}\rangle$ for a noise syndrome i.

Consider the unitary transformation Bob uses for decoding in the case of some value of the classical data r for which the decoding is successful and without loss of generality name this case $r=0$. (For a code which corrects with asymptotically perfect fidelity there may be some cases of r for which the correction does not work.) We also consider error syndrome i which is successfully corrected by U_d. We have

$$U_d(r=0)(|\eta_{0i}\rangle\otimes|\mathbf{0}\rangle)=|\xi\rangle\otimes|a_i\rangle. \tag{58}$$

(For our choice of i the final $|a_i\rangle$ state can without loss of generality be taken to be $|\mathbf{0}\rangle$ in an appropriately sized Hilbert space.) Applying $U_d^{-1}(r=0)$ gives

$$U_d^{-1}(r=0)(|\xi\rangle\otimes|\mathbf{0}\rangle)=|\eta_{0i}\rangle\otimes|\mathbf{0}\rangle. \tag{59}$$

There must exist another unitary operation U_s which rotates $|\eta_{0i}\rangle$ into the noiseless coded vector $|\zeta_0\rangle$. Thus,

$$U_sU_d^{-1}(r=0)(|\xi\rangle\otimes|\mathbf{0}\rangle)=|\zeta_0\rangle\otimes|\mathbf{0}\rangle. \tag{60}$$

In other words, $U_sU_d^{-1}(r=0)$ takes $|\xi\rangle$ into $|\zeta_0\rangle$ along with some ancillary inputs and outputs always in a standard $|\mathbf{0}\rangle$ state. Therefore $U_sU_d^{-1}(r=0)$ is a good encoder. Since this encoder always results in the correct code vector corresponding to classical data $r=0$ this data need not be sent to Bob at all, as he will have anticipated it. Thus, $U_sU_d^{-1}(r=0)$ and U_d form a code needing no classical side channel.

It may happen that for a large block code which only error-corrects to some high fidelity ($|\langle\xi|\xi_f\rangle|>1-\epsilon$ where $|\xi_f\rangle$ is the final output of the decoder) that *no* case is cor-

rected perfectly. Then the coded states produced by $U_s U_d^{-1}(r=0)$ will be imperfect. After transmission through the noisy channel and correction by U_d the final output will then be less perfect than in the original code. Nevertheless, because of unitarity it is clear that as $\epsilon \to 0$ the fidelity of this code will also approach unity.

Thus any protocol using classical one-way data transmission to supplement a quantum channel can be converted into a protocol in which the classical transmission is unnecessary and with the same capacity $Q=Q_1$. We have also now shown that the encoding stage is unitary, in the sense that no extra classical or quantum results accumulate in Alice's laboratory.

If the error syndrome $i=0$, corresponding to no error, is decoded with high fidelity by U_d then U_s can be taken to be the identity. Thus, the encoding and decoding transformations can in this case be written in a form where $U_e=U_d^{-1}$, a fact independently shown by Knill and Laflamme [40]. If the $i=0$ error syndrome is not decoded with high fidelity by U_d [42] then the encoder cannot be the inverse of the decoder. The proof is simple: $U_e(|\xi\rangle \otimes |\mathbf{0}\rangle)=|\zeta\rangle$ (where we have dropped the r subscripts since it has been proven the classical data is never needed) and therefore $U_e^{-1}|\zeta\rangle=(|\xi\rangle \otimes |\mathbf{0}\rangle)$. Thus U_e^{-1} decodes the noiseless coded vectors $|\zeta\rangle$ which is exactly what U_d has been assumed not to do.

B. Additivity of perfect and imperfect quantum channel capacities

Consider a channel of capacity $Q>0$ supplemented by a perfect channel of capacity 1. Suppose the imperfect channel is used n times and the perfect channel is used m times. We will call the maximum number of qubits transmitted through the channels in this case T. If the capacity of this joint channel is additive then $T=T_a=Qn+m$.

Suppose the number of qubits transmitted is superadditive, i.e., $T>T_a$. From the definition of noisy channel capacity we know that we can use an imperfect channel t times to simulate a perfect channel being used m times where $Qt=m$. We now use the imperfect channel a total $n+t$ times and we can transmit T qubits through this two-part use of the imperfect channel. But $T>T_a=Qn+m$ so

$$T>Qn+Qt. \tag{61}$$

The capacity of this channel is $Q'=T/(n+t)$. Using Eq. (61) we can write

$$Q'=\frac{T}{n+t}>\frac{Qn+Qt}{n+t}=Q. \tag{62}$$

A capacity of $Q'>Q$ has been achieved using only the original imperfect channel whose capacity was Q. This cannot be so.

C. QECC → 1-EPP proving $\forall_M \quad D_1(M) \geq Q(\hat{\chi}(M))$

To demonstrate this inequality (cf. Fig. 14) we use bipartite mixed states M in place of the standard maximally entangled states (Φ^+) to teleport n qubits from Alice to Bob. This teleportation defines a certain noisy channel $\hat{\chi}(M)$, so

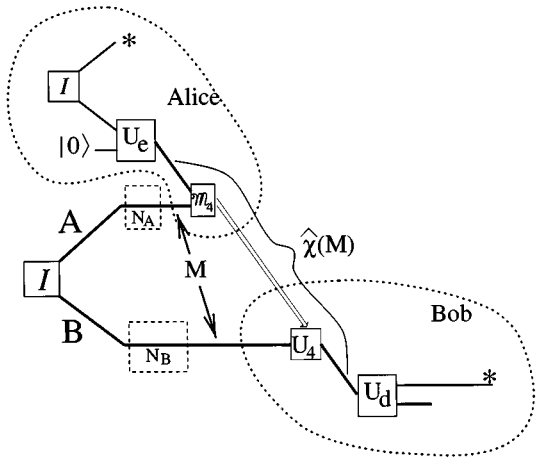

FIG. 14. A QECC can be transformed into a 1-EPP. Teleporting (M_4, U_4) via a mixed state M defines the noisy channel $\hat{\chi}(M)$. If a quantum error-correcting code $\{U_e, U_d\}$ can correct the errors in this channel, the code and channel can be used to share pure entanglement between Alice and Bob (*). This establishes inequality (52), viz., $\forall_M D_1(M) \geq Q(\hat{\chi}(M))$.

designated on the center right of the figure. Alice prepares n qubits to be teleported through this channel by applying the encoding transformation U_e of a QECC to m halves of EPR pairs which she generates in her laboratory (upper left) at I and to $n-m$ ancillas in the standard $|0\rangle$ state. The resulting quantum-encoded n qubits are teleported to Bob at lower right through the noisy channel. There Bob applies the decoding transformation U_d. If the code can successfully correct the errors introduced by the noisy teleportation, then the result is that Alice and Bob share m time-separated EPR pairs (*). Indeed, the whole figure can be regarded as a one-way purification protocol whereby Alice and Bob prepare m good EPR pairs from n of the initial mixed states M, using a QECC of rate $Q=m/n$ able to correct errors in the noisy quantum channel $\hat{\chi}(M)$. Thus $D_1(M)$ must be at least as great as the rate $Q(\hat{\chi}(M))$ of the best QECC able to achieve reliable quantum transmission through $\hat{\chi}(M)$.

D. 1-EPP → QECC proving $\forall_\chi \quad D_1(\hat{M}(\chi)) \leq Q(\chi)$

In the same style as the preceding section, we establish the second inequality by exhibiting an explicit protocol. The object is to show that, given the existence of a 1-EPP acting on the mixed state $\hat{M}(\chi)$ obtained from quantum channel χ, Alice can successfully transmit arbitrary quantum states $|\xi\rangle$ to Bob. The capacity Q of this quantum channel is the same as D_1 for the 1-EPP; this establishes that the capacity of χ is at least as good as the D_1 of the corresponding 1-EPP.

In fact, this protocol just involves the application of quantum teleportation [5] mentioned in the Introduction. In Fig. 15 we show more explicitly the necessary construction, which has already been touched on in Figs. 3 and 4. Alice and Bob are connected by channel χ. Alice arranges to share the bipartite mixed state $\hat{M}(\chi)$ with Bob by passing halves (the B particles) of maximally entangled states (Φ^+) from source I through χ to Bob. Then Alice and Bob partake in the 1-EPP protocol. We have represented this procedure

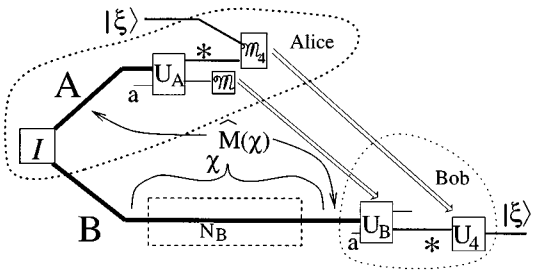

FIG. 15. A 1-EPP can be transformed into a QECC. Given χ, Alice creates mixed states $\hat{M}(\chi)$ by passing halves of entangled states Φ^+ from source I through the channel. Alice and Bob perform a 1-EPP resulting in perfectly entangled states (*) which are then used to teleport $|\xi\rangle$ safely to Bob, completing a QECC.

somewhat more generally than is necessary for the hashing-type procedures shown earlier, or for the finite-block protocols to be derived below. We simply indicate that they must preform two operations U_A and U_B, and that Alice will perform some measurements $\mathcal{M}$ and pass the results to Bob. The measurements which Bob would perform in the hashing protocol are understood to be incorporated in U_B. Also, we have accounted for the possibility that either Alice or Bob might employ an ancilla a for some of their processing operations.

By hypothesis, this protocol leaves Alice and Bob with nD_1 maximally entangled states (*). They then may use this resource to teleport nD_1 unknown quantum bits in the state $|\xi\rangle$. Thus, the net effect is that Alice and Bob, using channel χ supplemented by one-way classical communication, have a means of reliably transmitting quantum data, with capacity $D_1(\hat{M}(\chi))$. This is exactly a QECC on χ with a one-way classical side channel. However, Eq. (50) (proven in Sec. V A) states that the same capacity can be obtained without the use of classical communication. Thus, the ultimate capacity Q of channel χ must be at least as great. This establishes the inequality.

VI. SIMPLE QUANTUM ERROR-CORRECTING CODES

For most of the remainder of this paper, we will exploit the equivalence which we have established between 1-EPP on $\hat{M}(\chi)$ and a QECC on χ. We note that when the 1-EPP has the property that the unitary transformations U_B and U_4 performed by Bob can be done ''in place'' (i.e., no ancilla qubits need to be introduced; see Fig. 3), the 1-EPP can be transformed into a particularly simple style of QECC, exactly like the schemes which have been introduced by Shor [9] and have now been extended by many others [10–16], which are also all done ''in place.'' As we have seen in Figs. 14 and 15, some versions of 1-EPP and QECC may require ancilla a for their implementation.

The proof of the correspondence between the in-place 1-EPP and in-place QECC is immediate, following Sec. V D. The 1-EPP is used to make a QECC as in Fig. 15. The unitary transformations U_B and U_4 performed by Bob are combined as a U_d and U_d is performed in place by assumption. Thus $U_e = U_s U_d^{-1}$ (see Sec. V A) can also be done in place.

As a simple consequence of this result, the one-way hashing protocol of Sec. III B 3 can be reinterpreted as an explicit error-correction code, and indeed it does the same kind of job as the recent quantum error-correction schemes based on linear-code theory of Calderbank and Shor [10] and Steane [11]: in the limit of large qubit block size n, it protects an arbitrary state in a 2^m-dimensional Hilbert space from noise. We note that the hashing protocol actually does somewhat better than the linear-code schemes. $D_1(\hat{M}(\chi))$, and therefore $Q(\chi)$ [see Eq. (54)], is higher for hashing than for the linear-code scheme, as shown in Figs. 8 and 9.

We will make further contact with this other work on error-correction coding in finite blocks by showing how finite blocks of EPR pairs can be purified in the presence of noise which only affects a finite number of the Bell states. When transformed into an error-correcting code, this becomes a procedure for recovering from a finite number of qubit errors, as in Shor's procedure in which one qubit, coded into nine qubits, is safe from any error on a single qubit. We develop efficient numerical strategies based on the Bell-state approach which look for new coding schemes of this type, and in fact we find a code which does the same job as Shor's using only five EPR pairs.

A. Another derivation of a QECC from a restricted 1-EPP

Another way to derive the in-place QECC from the in-place 1-EPP is to exploit the symmetry between measurement and preparation in quantum mechanics. Here we will restrict our attention to noise models which are one-sided (i.e., N_A absent in Fig. 3), or *effectively* one-sided. An important case where the noise is effectively one-sided is when the mixed state M obtained in Fig. 5 is Bell diagonal, i.e., has the form of W [Eq. (29)]. We can say that, subjected to this noise, the pure Bell state is taken to an ensemble of each of the four Bell states, with some probabilities. Using the notation of Sec. III B 1 these are p_{00}, p_{01}, p_{10}, and p_{11}:

$$|\Phi^+\rangle \rightarrow \{\sqrt{p_{00}}|\Phi^+\rangle, \sqrt{p_{10}}|\Phi^-\rangle, \sqrt{p_{01}}|\Psi^+\rangle, \sqrt{p_{11}}|\Psi^-\rangle\} = \{R_{mn}|\Phi^+\rangle\}. \tag{63}$$

(Here R_{mn} are proportional to the operators $\{I, \sigma_x, \sigma_y, \sigma_z\}$ of Table I.) It is easy to show that the same mixed state could be obtained if the B particles were subjected to a generalized depolarizing channel, and N_A were absent. More generally, we require that $N_{A,B}$ be such that the resulting M could be obtainable from some channel χ; $M = \hat{M}(\chi)$ for some χ. This is a fairly obvious restriction to make, since we are planning on defining a QECC on this effective quantum channel χ. Note also that, since the twirling of Sec. III A (item 1) converts any bipartite mixed state into a Werner state, for some purposes *any* noise can be made effectively one-sided.

We will now show that under these conditions, the operations performed by Alice in Fig. 15 can be greatly simplified. Consider the joint state of the A and B particles after Alice has applied the unitary transformation U_1 of Fig. 3 as part of the purification protocol, but before the one-sided noise N_B has acted on the B particles. The joint state is still a pure, maximally entangled state. For convenience, we assume that the source I produces Φ^+ Bell states. (If it produced another type of Bell state, some additional simple rotations can be

inserted in the derivation we are about to give.) The initial product of n Bell states may be written

$$|\Phi\rangle_i = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle_A |x\rangle_B . \tag{64}$$

After the application of the unitary transformation U_1 to Alice's particles, the new state of the system is

$$|\Phi\rangle_f = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} \sum_{y=0}^{2^n-1} (U_1)_{x,y} |y\rangle_A |x\rangle_B . \tag{65}$$

But notice that by a simple change of the dummy indices, this state can be rewritten

$$|\Phi\rangle_f = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} \sum_{y=0}^{2^n-1} |x\rangle_A (U_1^T)_{x,y} |y\rangle_B . \tag{66}$$

That is, the unitary transformation applied to the A particles is completely equivalent to the same operation (transposed) applied to the B particles.

Alice's tasks in the 1-EPP protocol are thus reduced to making one-particle measurements $\mathcal{M}$ on $n-m$ of the A particles, making Bell measurements $\mathcal{M}_4$ between the m qubits $|\xi\rangle$ to be protected and her remaining m particles (as in quantum teleportation [5]), and applying U_1^T to the B particles before sending them, along with her classical measurement results, to Bob. (Recall from the Introduction that m is the yield of good singlets from the purification protocol.)

However, the $n-m$ one-particle measurements $\mathcal{M}$ can be eliminated entirely. We use the property of Φ^+ states that if one of the particles is measured to be $|0\rangle$ or $|1\rangle$ in the z basis, then the other particle is ''collapsed'' into the same state [1,2]. So, rather than creating $n-m$ entangled states at I, Alice simply prepares $n-m$ qubits in a definite state and sends them directly into the U_1^T operation. To mimic the randomness of the measurement $\mathcal{M}$, Alice might do $n-m$ coin flips to decide what the prepared state of these B particles will be, and send this classical data on to Bob. But this is unnecessary, since by hypothesis, the 1-EPP always yields perfect entangled pairs (*), no matter what the values of the $\mathcal{M}$ measurements were. So, Alice and Bob may as well preagree on some particular definite set of values (e.g., all 0's), and Alice will always preset those B particles to that state [43].

The only A particles remaining in the protocol at this point are the m particles forming the halves of perfect EPR pairs with Bob, and which are immediately used for teleportation to Bob. But we note that, following the usual rules of teleportation, the measurement $\mathcal{M}_4$ causes the corresponding B particles, immediately after their creation at source I, to be in the state $|\xi\rangle$ (if the measurement outcome were 00), or a rotated version, $\sigma_{x,y,z}|\xi\rangle$ (for the other measurement outcomes). Again, the protocol should succeed no matter what the value of this measurement; therefore, if Alice and Bob preagree that this classical data should be taken to have the value 00, then Alice can eliminate the A particles entirely, eliminate the preparation I of entangled states, and simply feed in the $|\xi\rangle$ states directly as B particles into the U_1^T transformation. (Bob also does the U_4 operation of Fig. 3 appropriate for 00, namely, a no-op.)

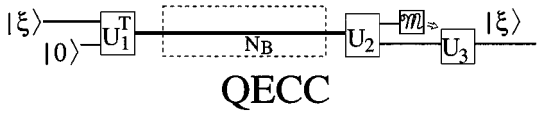

FIG. 16. The one-way purification protocol of Fig. 4 may be transformed into the quantum-error-correcting-code protocol shown here. In a QECC, an arbitrary quantum state $|\xi\rangle$, along with some qubits which are originally set to $|0\rangle$, are encoded in such a way by U_1^T that, after being subjected to errors N_B, decoding U_2 followed by measurement $\mathcal{M}$, followed by final rotation U_3, permits an exact reconstruction of the original state $|\xi\rangle$.

Finally we step back to see the effect that this series of transformations has produced, as summarized in Fig. 16. All use of bipartite states I, and the corresponding A particles, has been eliminated, along with all the measurement results transmitted to Bob. The net effect is that Alice has taken the m-qubit unknown quantum state $|\xi\rangle$ along with $n-m$ ''blank'' qubits, processed them with U_1^T, and sent them on channel χ to Bob. He is able to use his half of the protocol, without any additional classical messages, to reconstruct $|\xi\rangle$. This, of course, is precisely the in-place QECC that we want.

B. Finite block-size purification and error correcting codes

We have now shown that Bell-state purification procedures can be mapped directly into quantum error-correcting codes. This gives an alternative way to look for quantum error-correction procedures within the purification approach. This can be both analytically and computationally useful. In fact, we can take over everything which we obtained via the hashing protocol of Sec. III B 3, in which Alice and Bob perform a sequence of unilateral and bilateral unitary operations to transform their bipartite state from one collection of Bell states to another, in order to gain information about the errors to which their particles have been subjected.

In this section we will show that this approach can also be used to do purification, and thus error correction, in small, finite blocks of qubits, in the spirit of much of the other recent work on QECC [8–16]. In these procedures the object is slightly different than in the protocols which employ asymptotically large block sizes: Here, we wish purify a finite block of n EPR pairs, of which no more than t have interacted with the environment (i.e., been subjected to noise). The end result is to be $m<n$ maximally entangled pairs, for which $F=1$ exactly. The explicit result we present below will be for $n=5$, $m=1$, and $t=1$. This protocol thus has the same capability as the one recently reported by Laflamme *et al.* [12], although the quantum network which we derive below is simpler in some respects. We are still investigating the extent to which our two protocols are equivalent.

The general approach will be the same as in Sec. III; however, our earlier emphasis was on error correction in asymptotically large blocks of states. To deal with the finite-block case, we will need a few small but important modifications (see Fig. 17).

(1) There will again be a set $\mathcal{L}$ of possible collections of Bell states after the action of the noise N_B; but rather than

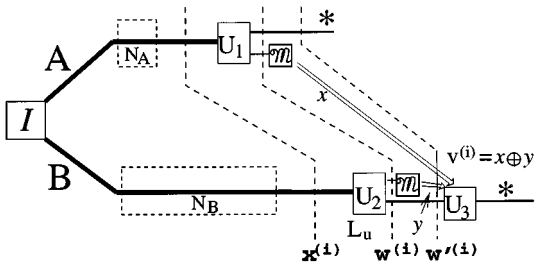

FIG. 17. The 1-EPP of Fig. 3 marked with the notation used in this section.

being a ''likely set'' defined by the fidelity of the channel, we will characterize the noise by a promise that the number of errors cannot exceed a certain number t. Cases with $t+1$ errors are not just deemed to have low probability; they are declared to be disallowed, following Shor [9].

(2) The set $\mathcal{L}$ will have a definite, finite size; if the size of the Bell state block is n and the number of erroneous Bell states to be corrected is t, then the size of the set is [13]

$$S=\sum_{p=0}^{t} 3^p \binom{n}{p}. \tag{67}$$

Borrowing the traditional language of error correction, each member of the set, indexed by i, $1\leqslant i\leqslant S$, defines an *error syndrome*. The ''3'' in Eq. (67) corresponds to the number of possible incorrect Bell states occurring in the evolution of Eq. (63): there is either a phase error ($\Phi^+\rightarrow\Phi^-$), an amplitude error ($\Phi^+\rightarrow\Psi^+$), or both ($\Phi^+\rightarrow\Psi^-$) [11,13]. It has been noted [10,13] that correcting these three types of error is sufficient to correct any arbitrary noise to which the quantum state is subjected, which we prove in Appendix B.

(3) The object of the error correction is slightly different than in Sec. III; in the earlier case it was to find a protocol where the fidelity of the remaining EPR pairs approached unity asymptotically as $n\rightarrow\infty$. In the finite-block case, the object is to find a protocol such that the fidelity attains exactly 100%, that is, m good EPR pairs are guaranteed to be recoverable from the original set of n Bell states for every single one of the S error syndromes.

Let us emphasize again that, in the purification language which we have developed, the quantum error correction problem has been turned into an entirely classical exercise: given a set of n Bell states, we use the operations of item 2 in Sec. III A to create a classical Boolean function which maps these Bell states onto others such that, for all S of the error syndromes, the first m Bell states are always the same when the measurement results on the remaining $n-m$ Bell states are the same.

We will develop this informal statement of the problem in a more formal mathematical language. First, recall the code which we introduced for the Bell states in item 5 of Sec. III A in which, for example, the collection of Bell states $\Phi^+\Phi^-\Phi^+$ is coded as the six-bit word 001000. As in our hashing-protocol discussion (Sec. III B 3), we denote such words by $x^{(i)}$, where the superscript i denotes the word appropriate for the ith error syndrome. These words have $2n$ bits, and we will sometimes denote by $x_k^{(i)}$ the kth bit of the word.

Alice and Bob subject $x^{(i)}$ to the unitary transformations U_1 and U_2. They are confined to performing sequences of the unilateral and bilateral operations introduced in Table I. In particular, they can do either (1) a bilateral XOR, which flips the low (right) bit of the target iff the low bit of the source is 1, and flips the high (left) bit of the source iff the high bit of the target is 1; (2) a bilateral $\pi/2$ rotation B_y of both spins in a pair about the y axis, which interchanges the high and low bits; (3) a unilateral (by either Alice or Bob) π rotation σ_z of one spin about the z-axis, which complements the low bit; or (4) a composite operation $\sigma_x B_x$, where the σ_x operation is unilateral and the B_x is bilateral; the simple net effect of this sequence of operations is to flip the low bit iff the high bit is one.

It is easy to show that with these four operations, Alice and Bob can do anything which they can do with the full set of operations in Table I. In our classical representation, the effect of such a sequence of operations is to apply a classical Boolean function L_u to $x^{(i)}$, yielding a string $w^{(i)}$:

$$w^{(i)}=L_u(x^{(i)}). \tag{68}$$

We use the symbol L_u for this function because, with the operations that Alice and Bob have at their disposal, L_u is constrained to be a linear, reversible Boolean function. This is easy to show for the sequences of the four operations given above. Note, however, that not all linear reversible Boolean functions are obtainable with this repertoire. A linear Boolean function [44] can be written as a matrix equation

$$w^{(i)}=M_u x^{(i)}+b. \tag{69}$$

Here the matrix M and the vector b are Boolean valued ($\in\{0,1\}$), and addition is defined modulo 2. Reversibility adds an additional constraint: $\det(M)=1$ (modulo 2). In a moment we will write down the condition which the set of $w^{(i)}$ must satisfy in order for purification to succeed.

The next step of purification is a measurement $\mathcal{M}$ of $n-m$ of the Bell states. As discussed in item 5 of Sec. III A, after learning Alice's measurement result, Bob can deduce the low bit of each of the measured Bell states. If we write these measurement results for error syndrome i as another Boolean word $v^{(i)}$ (of length $n-m$), the measurement can be expressed as another linear Boolean function:

$$v^{(i)}=M_m w^{(i)}. \tag{70}$$

The matrix elements of M_m are

$$(M_m)_{kl}=\delta_{k,2(m+k)}. \tag{71}$$

The state of the remaining unmeasured Bell states is coded in a truncated word w' of length $2m$:

$$w'^{(i)}=(w_1 w_2\cdots w_{2m})^{(i)}. \tag{72}$$

We now have all the machinery to state the condition for a successful purification. The object is to perform a final rotation U_3 on the state coded by w' and restore it, for every error syndrome i, to the state $00\cdots0$. Whatever w' is, such a restoring U_3 is always available to Bob; for each Bell state, he does the Pauli rotations:

TABLE III. Possible initial Bell states and the resulting final state after the gate array of Fig. 18 has been applied.

i	Initial state $x^{(i)}$					Final state $w^{(i)}$					Measurement result $v^{(i)}$			
1	00	00	00	00	00	00	00	00	00	01	0	0	0	1
2	01	00	00	00	00	01	00	00	01	01	0	0	1	1
3	10	00	00	00	00	10	01	00	00	01	1	0	0	1
4	11	00	00	00	00	11	01	00	01	01	1	0	1	1
5	00	01	00	00	00	00	01	00	00	00	1	0	0	0
6	00	10	00	00	00	01	10	01	00	01	0	1	0	1
7	00	11	00	00	00	01	11	01	00	00	1	1	0	0
8	00	00	01	00	00	10	00	11	11	01	0	1	1	1
9	00	00	10	00	00	00	00	01	00	00	0	1	0	0
10	00	00	11	00	00	10	00	10	11	00	0	0	1	0
11	00	00	00	01	00	10	01	01	10	01	1	1	0	1
12	00	00	00	10	00	00	00	01	01	00	0	1	1	0
13	00	00	00	11	00	10	01	00	11	00	1	0	1	0
14	00	00	00	00	01	00	00	00	00	00	0	0	0	0
15	00	00	00	00	10	01	11	11	01	11	1	1	1	1
16	00	00	00	00	11	01	11	11	01	10	1	1	1	0

Bell state	U_3transformation
00	I(do nothing)
01	σ_z
10	σ_x
11	σ_y.

(73)

But Bob must know which of these four rotations to apply to each of the remaining m Bell states. The only information he has on which of them to perform are the bits of the measurement vector $v^{(i)}$. This information will be sufficient, if for every error syndrome which produces a distinct w',v is distinct; in this case, Bob will know exactly which final rotation U_3 to apply.

This, then, is our final condition for successful purification. In more mathematical language, we require an operation L_u for which

$$\forall_{i,j} \quad w'^{(i)} \neq w'^{(j)} \Rightarrow v^{(i)} \neq v^{(j)}. \tag{74}$$

We will shortly show the results of a search for L_u which satisfy Eq. (74).

But first, we touch a point which has been raised in the recent literature [10–13]: Bob will obviously know which rotation U_3 to apply if from the measurement he learns the precise error syndrome, that is if for each error syndrome the measurement outcome is distinct. This ''condition for learning all the errors'' may be stated mathematically in a way parallel to Eq. (74):

$$\forall_{i,j} \quad i \neq j \Rightarrow v^{(i)} \neq v^{(j)}. \tag{75}$$

This condition is obviously *sufficient* for successful error correction; however, it is more restrictive than Eq. (74), and it is not a *necessary* condition. If Eq. (75) *were* a necessary condition for error correction, then a comparison of the number of possible distinct measurements $v^{(i)}$ with the number of error syndromes S leads [13,12] to a restriction on the block size in which a certain number of errors can be corrected:

$$S = \sum_{p=0}^{t} 3^p \binom{n}{p} \leqslant 2^{n-m}. \tag{76}$$

It is this bound which is attained, asymptotically, by the hashing and breeding protocols above. However, Eq. (74) puts no obvious restriction on the block size in which error correction can succeed, suggesting that the bound Eq. (76) can actually be exceeded. For example, if the transformation L_u were permitted to be any arbitrary Boolean function, then it would be capable of setting $w' = 00\cdots0$ for every syndrome i, in which case *no* error correction measurements v would be needed.

However, L_u is very strongly constrained in addition to being a linear, reversible Boolean function, and we are left uncertain to what degree the bound Eq. (76) may be violated. For the small cases which we have explored below, in which one Bell state is restored from single-qubit errors ($m=1$, $t=1$), we find that the bound of Eq. (76) is *not* exceeded. All solutions which we find which satisfy Eq. (74) also happen to identify every error syndrome uniquely [Eq. (75)]. The present work, therefore, does not demonstrate that Eq. (74) actually leads to more power error-correction schemes than Eq. (75). However, Shor and Smolin [35] have recently exhibited a family of new protocols which, at least asymptotically for large n, exceed the bound Eq. (76) by a small but finite amount.

C. Monte Carlo results for finite-block purification protocols

For the single-error ($t=1$), single-purified-state ($m=1$) case, we have performed a Monte Carlo computer search for unitary transformations U_1 and U_2. The program first tabulates the $x^{(i)}$ for all the allowed error syndromes i, as shown

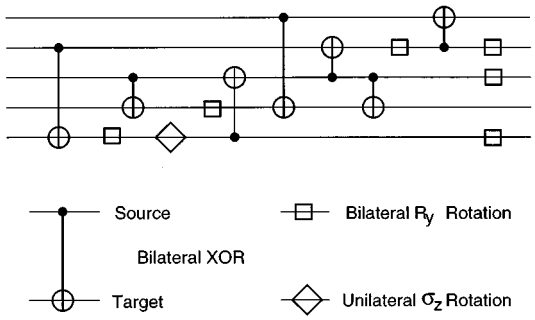

FIG. 18. The quantum gate array, determined by our computer search, which protects one qubit from single-bit errors in a block of five. ''Bilateral'' and ''unilateral'' refer to whether both Alice and Bob, or only Alice (or Bob), perform the indicated steps in the 2-EPP; in the QECC version, it corresponds to whether the operation is done in both coding and decoding, or in just the coding (or decoding) operations. All but the first qubit are measured at the end.

in Table III. (For the case of $t=1$ there are $S=3n+1$ error syndromes, since either of the n Bell states could suffer three types of error, plus one for the no-error case.) The program then randomly selects one of the four basic operations enumerated above, and randomly selects a Bell state or pair of Bell states to which to apply the operation. The program then checks whether the resulting set of states $w^{(i)}$ satisfies the error-correction condition of Eq. (74). If the answer is no, then the program repeats the procedure, adding another random operation. If the answer is yes, the program saves the list of operations, and starts over, seeking a shorter solution. Two ''shortness'' criteria were explored: fewest total operations, and fewest total BXOR's (since two-bit operations could be the more difficult ones to implement in a physical apparatus [32]).

A simple argument akin to the one of Sec. IV shows that error correction in a block of 2 ($t=1$, $m=1$, $n=2$) is impossible. We performed an extensive search for $n=3$ and $n=4$ codes; it would not be possible to detect the complete error syndrome for these cases [Eq. (76)], but it would appear *a priori* possible to satisfy Eq. (74). Nevertheless, no solutions were found, strongly suggesting that, for this case, $n=5$ is the best block code possible [12]. Knill and Laflamme have recently proved this [40].

Our search found many solutions for $n=5$ with similar numbers of quantum gate operations. The minimal network which was eventually found was one with 11 operations, six of which were BXOR's. Here we present a complete analysis of a slightly different solution, which involves 12 operations, seven of which are BXOR's. The gate array for this solution is shown in Fig. 18. The complete action of U_1 and U_2 produced by this quantum network is given in Table III.

Note that, as indicated above, this code not only satisfies the actual error-correction criterion Eq. (74), but it also satisfies the stronger condition Eq. (75); all the error syndromes are distinguished by the measurement results $v^{(i)}$.

It is interesting to note, as a check, that the tabulated transformation is indeed a reversible, linear Boolean operation. The reader may readily confirm that the results of Table III are obtained from the linear transformation Eq. (69), with

$$M_u=\begin{pmatrix}
1&0&0&0&0&1&0&1&0&0\\
0&1&1&0&0&0&0&0&1&0\\
0&0&1&0&0&0&0&0&1&0\\
1&0&0&1&0&0&0&1&1&0\\
0&0&0&0&0&1&0&0&1&0\\
0&0&1&0&1&1&1&1&1&0\\
0&0&0&0&0&1&0&1&0&0\\
0&0&0&0&0&1&1&0&1&0\\
0&1&0&0&0&0&0&0&1&0\\
0&0&0&1&1&0&1&0&0&1
\end{pmatrix} \tag{77}$$

and

$$b=(0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1). \tag{78}$$

D. Alternative conditions for successful quantum error correction code

While all of our work has involved deriving QECC's using the 1-EPP construction, it is possible, and instructive, to formulate the conditions for a good error correcting code directly in the QECC language. As Shor first showed [9], in this language the requirements become a set of constraints which the subspace into which the quantum bits are encoded must satisfy. In the course of our work we derived a set of general conditions for the case of error-correcting a single bit ($m=1$). They are quite similar to conditions which other workers have formulated recently [13,45]. Knill and Laflamme have recently obtained the same condition [40].

We will assume that only one qubit is to be protected, but the generalization to multiple qubits is straightforward. Suppose a qubit is encoded (by U_1^T in Fig. 16) as a state

$$|\xi\rangle=\alpha|v_0\rangle+\beta|v_1\rangle, \tag{79}$$

where α and β are arbitrary except for the normalization condition

$$|\alpha|^2+|\beta|^2=1, \tag{80}$$

and $|v_0\rangle$ and $|v_1\rangle$ are two basis vectors in the high-dimensional Hilbert space of the quantum memory block. Can $|v_0\rangle$ and $|v_1\rangle$ be chosen such that, after the quantum state is subjected to Werner-type errors, the original quantum state can still be perfectly reconstituted as the state of a single qubit,

$$|\xi_f\rangle=\alpha|0\rangle+\beta|1\rangle? \tag{81}$$

We shall derive the conditions which $|v_0\rangle$ and $|v_1\rangle$ must satisfy in order for this to be true.

We specify the action of the noise as a mapping of the original quantum state into an ensemble of unnormalized state vectors given by applying the linear operators R_i to the original state vector:

$$|\xi\rangle\rightarrow\{R_i|\xi\rangle\}. \tag{82}$$

For each error syndrome i there is an (unnormalized) operator R_i specifying the effect of the noise, as in Eq. (63). For

single-bit errors, the R_i's are just proportional to a σ_x, σ_y, or σ_z operator applied to one of the quantum-memory qubits, as discussed below. Two-bit errors would involve operators like $R_i = \sigma^{\alpha}_{x,y,z}\sigma^{\beta}_{x,y,z}$ applied to two different qubits α and β, and so forth. Equivalently to Eq. (82), the effect of the noise N_B in Fig. 16 can be expressed as a ensemble of normalized state vectors $|\xi_i\rangle$ with their associated probabilities p_i:

$$|\xi\rangle \rightarrow \{p_i, |\xi_i\rangle\} = \left\{ \langle \xi | R_i^\dagger R_i | \xi \rangle, \frac{R_i|\xi\rangle}{\sqrt{\langle \xi | R_i^\dagger R_i | \xi \rangle}} \right\}. \quad (83)$$

The Werner noise can be set up so that the p_i's are the probabilities that the environment ''measures'' the ith outcome of a pointer or ancilla space. We can evaluate the probability p_i (for the ith outcome of these measurements) for the state Eq. (79) using the expression in Eq. (83):

$$p_i = (\alpha^*, \beta^*) \times \begin{pmatrix} \langle v_0|R_i^\dagger R_i|v_0\rangle & \langle v_0|R_i^\dagger R_i|v_1\rangle \\ \langle v_1|R_i^\dagger R_i|v_0\rangle & \langle v_1|R_i^\dagger R_i|v_1\rangle \end{pmatrix} \times \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (84)$$

We have used the linearity of the operators R_i. The matrix notation used in Eq. (84) will prove useful in a moment.

The first, necessary condition which must be satisfied in order that the state may be reconstituted as in Eq. (81) is that the environment producing the Werner noise can acquire no information about the initial quantum state by doing this ancilla measurement. This will be true so long as p_i in Eq. (84) is not a function of the state vector coefficients α and β. It may be noted that the right hand side of Eq. (84) has the form of the expectation value of a 2×2 Hermitian operator in the state $(\alpha,\beta)^T$. It is a well-known theorem of linear algebra that such an operator can only have an expectation value independent of the state vector $(\alpha,\beta)^T$ iff the Hermitian operator is proportional to the identity operator. This gives us the first two conditions that the state vector may be recovered exactly: $\forall_i$,

$$\langle v_0|R_i^\dagger R_i|v_0\rangle = \langle v_1|R_i^\dagger R_i|v_1\rangle = p_i,$$
$$\langle v_1|R_i^\dagger R_i|v_0\rangle = 0. \quad (85)$$

If this condition is satisfied, then the ensemble of state vectors in Eq. (82) can be written in the simplified form:

$$\alpha|v_0\rangle + \beta|v_1\rangle \rightarrow \left\{ p_i, \frac{\alpha R_i|v_0\rangle + \beta R_i|v_1\rangle}{\sqrt{p_i}} \right\}. \quad (86)$$

Now, given that the environment learns nothing from the measurement, a further, sufficient condition is that there exist a unitary transformation (U_2) which takes each of the state vectors of Eq. (86) to a vector of the form:

$$\frac{1}{\sqrt{\langle v_0|R_i^\dagger R_i|v_0\rangle}}(\alpha R_i|v_0\rangle + \beta R_i|v_1\rangle) \rightarrow (\alpha|0\rangle + \beta|1\rangle)|a_i\rangle. \quad (87)$$

Here $|a_i\rangle$ is a normalized state vector of all the qubits excluding the one which will contain the final state Eq. (81). Because of unitarity, the angle between any two state vectors must be preserved. Taking the dot product of the state vectors resulting from two different syndromes i and j, and equating the result before and after the unitary operation gives

$$\frac{1}{\sqrt{\langle v_0|R_i^\dagger R_i|v_0\rangle}\sqrt{\langle v_0|R_j^\dagger R_j|v_0\rangle}} \times (\alpha^*, \beta^*) \times \begin{pmatrix} \langle v_0|R_i^\dagger R_j|v_0\rangle & \langle v_0|R_i^\dagger R_j|v_1\rangle \\ \langle v_1|R_i^\dagger R_j|v_0\rangle & \langle v_1|R_i^\dagger R_j|v_1\rangle \end{pmatrix} \times \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = |\alpha|^2\langle a_i|a_j\rangle + |\beta|^2\langle a_i|a_j\rangle = \langle a_i|a_j\rangle. \quad (88)$$

In the last part we have used the normalization condition to eliminate α and β. Now, since the right-hand side of Eq. (88), and the prefactor of the left hand side, are independent of α and β, so must be the expectation value of the 2×2 Hermitian operator. We again conclude that this Hermitian operator must be proportional to the identity operator, and this gives the final necessary and sufficient conditions [46] for successful storage of the quantum data: $\forall_{i,j}$,

$$\langle v_0|R_i^\dagger R_j|v_0\rangle = \langle v_1|R_i^\dagger R_j|v_1\rangle, \quad (89)$$

$$\langle v_1|R_i^\dagger R_j|v_0\rangle = 0. \quad (90)$$

For the specific five-qubit code described above, we found (by another, simple computer calculation) that the two basis vectors of Eq. (79) are

$$|v_0\rangle \propto (-|00000\rangle - |11000\rangle - |01100\rangle - |00110\rangle - |00011\rangle - |10001\rangle + |10010\rangle + |10100\rangle + |01001\rangle + |01010\rangle + |00101\rangle + |11110\rangle + |11101\rangle + |11011\rangle + |10111\rangle + |01111\rangle), \quad (91)$$

i.e., a superposition of all even-parity kets, with particular signs, and

$$|v_1\rangle = (\text{the corresponding vector with 0 and 1 interchanged}). \quad (92)$$

It is easy to confirm that this pair of vectors satisfies the conditions Eqs. (89) and (90). It is interesting to note that these two vectors do *not* span the same two-dimensional subspace as the ones recently reported by Laflamme *et al.* [12]; but it has recently been shown that they are related to one another by one bit rotations [47].

E. Implications of error-correction conditions on channel capacity

Knill and Laflamme [40] have used the error correction conditions [Eqs. (89) and (90)] to provide a stronger upper bound for Q and D_1 than the one of Sec. IV by showing that $D_1=0$ when $F=0.75$. We indicate this on Figs. 8 and 9, using our channel-additivity result of Sec. V B to extend this to the linear bound shown. Their proof is as follows: write the coded qubit basis states [cf. Eqs. (91) and (92)] as

$$|v_i\rangle=\sum_x \alpha_x^i|x\rangle=\sum_{y:z} \alpha_{y:z}^i|y:z\rangle. \tag{93}$$

Here x stands for an n bit binary number, and $y:z$ stands for a partitioning of x into a $2t$-bit substring y and an $(n-2t)$-bit substring z. (The partitioning may be arbitrary, and need not be into the least significant and most significant bits.) Knill and Laflamme then consider the reduced density matrices on the y and the z spaces:

$$\rho_{n-2t}^i=\sum_{y,z_1,z_2} \alpha_{y:z_1}^i \alpha_{y:z_2}^{i*}|z_1\rangle\langle z_2|, \tag{94}$$

$$\rho_{2t}^i=\sum_{y_1,y_2,z} \alpha_{y_1:z}^i \alpha_{y_2:z}^{i*}|y_1\rangle\langle y_2|. \tag{95}$$

Knill and Laflamme then prove two operator equations. First:

$$\rho_{n-2t}^0\rho_{n-2t}^1=0. \tag{96}$$

This is proved by using the condition for a successful error-correction code [Eq. (90)], where the linear operator R_i operates on a set of t bits, and R_j operates on a different set of t bits. (These R's should be taken as projection operators in this proof.) Likewise, by applying Eq. (89) with the same operators R_i and R_j, they prove

$$\rho_{2t}^0=\rho_{2t}^1. \tag{97}$$

These two equations give a contradiction when the two substrings are of the same size, because it says that reduced matrices are simultaneously orthogonal and identical. This says that no code can exist if $2t=n-2t$, which corresponds to $F=1-t/n=0.75$. As a bonus, these results give an interesting insight into the behavior of coded states: no measurement on $2t$ qubits can reveal anything about whether a 0 or a 1 is encoded, while there exists a measurement on $n-2t$ qubits which will distinguish with certainty a coded 0 from a coded 1.

This result shows that the lowest fidelity Werner channel with finite capacity must have $F>0.75$. Call that fidelity F_0. Consider a channel with fidelity F between F_0 and 1. The capacity of this channel is no greater than that of a composite channel consisting of a perfect channel used a fraction $(F-F_0)/(1-F_0)$ of the time and a channel with fidelity F_0 used $(1-F)/(1-F_0)$ of the time because the first channel is the same as the composite channel provided one is unaware of whether the fidelity is 1 or F_0 on any particular use of the channel. (This construction is akin to that of Sec. IV.) By the channel additivity argument of Sec. V B the capacity of the composite channel, which bounds the capacity of the fidelity F channel, cannot exceed $(F-F_0)/(1-F_0)$. Since F_0 cannot be below 0.75 we obtain the straight-line bound

$$Q=D_1\leqslant 4F-3, \tag{98}$$

as shown in Figs. 8 and 9.

VII. DISCUSSION AND CONCLUSIONS

There has been an immense amount of recent activity and progress in the theory of quantum error-correcting codes, including block codes with some error-correction capacities in blocks of two [16], three [13,14], and four [16]. Codes which completely correct single-bit errors have now been reported for block sizes of five as in the present work [12], seven [11], eight [15], and nine [9]; this is in addition to the work using linear-code theory of families of codes which work up to arbitrarily large block sizes [10,11]. A variety of subsidiary criteria have been introduced, such as correcting only phase errors, maintaining constant energy in the coded state, and correction by a generalized watchdogging process. Much of this work can be expressed in entanglement purification language, in some cases more simply.

Our results highlight the different uses to which a quantum channel may be put. When a noisy quantum channel is used for classical communication, the goal—by optimal choice of preparations at the sending end, measurements at the receiving end, and classical error-correction techniques—is to maximize the throughput of reliable classical information. When used for this purpose, a simple depolarizing channel from Alice to Bob has a positive classical capacity $C>0$ provided it is less than 100% depolarizing. Adding a parallel classical side channel to the depolarizing quantum channel would increase the classical capacity of the combination by exactly the capacity of the classical side channel.

When the same depolarizing channel is used in connection with a QECC or EPP to transmit unknown quantum states or share entanglement, its quantum capacity Q is positive only if the depolarization probability is sufficiently small ($<1/3$), and this capacity is not increased at all by adjoining a parallel classical side channel. On the other hand, an additional classical back channel, from Bob to Alice, does enhance the quantum capacity, making it positive for all depolarization probabilities less than 2/3.

It is instructive to compare our results to the simpler theory of noiseless quantum channels and pure maximally entangled states. There the transmission of an intact two-state quantum system or qubit (say, from Alice to Bob) is a very strong primitive, which can be used to accomplish other weaker actions, in particular the undirected sharing of an ebit of entanglement between Alice and Bob, or the directed transmission of a bit of classical information from Alice to Bob. (These two weaker uses to which a qubit can be put are

mutually exclusive, in the sense that k qubits cannot be used simultaneously to share ℓ ebits between Alice and Bob *and* to transmit m classical bits from Alice to Bob if $\ell+m>k$ [48]).

A noisy quantum channel χ, if it is not too noisy, can similarly be used, in conjunction with QECC's, for the reliable transmission of unknown quantum states, the reliable sharing of entanglement, or the reliable transmission of classical information. Its capacity for the first two tasks, which we call the quantum capacity $Q(\chi)$, is a lower bound on its capacity $C(\chi)$ for the third task, which is the channel's conventional classical capacity.

Most error-correction protocols are designed to deal with error processes that act independently on each qubit, or affect only a bounded number of qubits within a block. A quite different error model arises in quantum cryptography, where the goal is to transmit qubits, or share pure ebits, in such a way as to shield them from entanglement with a malicious adversary. Traditionally one grants this adversary the ability to listen to all classical communications between the protagonists Alice and Bob, and to interact with the quantum data in a highly correlated way designed to defeat their error-correction or entanglement-purification protocol. It is not yet known whether protocols can be developed to deal successfully with such an adversarial environment.

Even for the simple error models which introduce no entanglement between the message qubits, there are still a wide range of open questions. As Fig. 8 has shown, we still do not know what the attainable yield is for a given channel fidelity; but we are hopeful that the upper and lower bounds we have presented can be moved towards one another, for both one-way and two-way protocols.

Improving the lower bounds is relatively straightforward, as it simply involves construction of protocols with higher yields. An important step towards this has been the realization that it is not necessary to identify the entire error syndrome to successfully purify. This has permitted the lower bound for one-way protocols (and thus for QECC's) to be raised slightly above the D_H curve of Fig. 8 (see Ref. [35]).

Improvement of the upper bounds is more problematical. For two-way protocols, we presently have no insight into how this bound can be lowered below E. Characterizing D_1, D_2, and E for all mixed states would be a great achievement [49], but even that would not necessarily provide a complete theory of mixed state entanglement. Such a theory ought to describe, for any two bipartite states M and M', the asymptotic yield with which state M' can be prepared from state M by local operations, with or without classical communication. In general, the most efficient preparation would probably not proceed by distilling pure entanglement out of M', then using it to prepare M; it is even conceivable that there might be incomparable pairs of states, M and M' such that neither could be prepared from the other with positive yield.

Surprisingly, basic questions about even the classical capacity of quantum channels remain open. For example, it is not known whether the classical capacity of two parallel quantum channels can be increased by entangling their inputs.

For us, all of this suggests that, even 70 years after its establishment, we still are only beginning to understand the full implications of the quantum theory. Its capacity to store, transmit, and manipulate information is clearly different from anything which was envisioned in the classical world. It still remains to be seen whether the present surge of interest in quantum error correction will enable the great potential power of quantum computation to be realized, but it is clearly a step in this direction.

ACKNOWLEDGMENTS

We are especially grateful to Peter Shor for many important insights which he has imparted to us during the course of this work. We thank G. Brassard, R. Cleve, A. Ekert, R. Jozsa, M. Knill, R. Laflamme, R. Landauer, C. Macchiavello, S. Popescu, and B. Schumacher for helpful discussions.

APPENDIX A: IMPLEMENTATION OF RANDOM BILATERAL ROTATION

In this appendix we show how an arbitrary density matrix of two particles can be brought into the Werner form by making a random selection, with uniform probabilities, from a set of 12 operations $\{U_i\}$ which involve identical rotations on each of the two particles. [Thus, the rotations U_i are members of a particular SU(2) subset of SU(4).] After such a set of rotations the density matrix is transformed into an arithmetic average of the rotated matrices:

$$M_T=\frac{1}{N}\sum_{i=1}^{N} U_i^\dagger M U_i . \tag{A1}$$

N will be 12 in the example we are about to give. The 4×4 density matrix M, expressed in the Bell basis, has three parts which behave in different ways under rotation: (1) the diagonal singlet (Ψ^-) matrix element, which transforms as a scalar; (2) three singlet-triplet matrix elements, which transform as a vector under rotation; and (3) the 3×3 triplet block, which transforms as a second-rank symmetric tensor. In the desired Werner form the vector part of the density matrix is zero, and the symmetric second-rank tensor part is proportional to the identity.

The mathematics of this problem is the same as that which describes the tensor properties of a large collection of molecules as would occur in a liquid, glass, or solid. In the case of a liquid, all possible orientations of the molecules occur. Because of the orientational averaging [mathematically equivalent to Eq. (A1), where the sum runs over all SU(2) operations], vector quantities become zero (e.g., the net electric dipole moment of the liquid is zero), while second-rank tensor quantities become proportional to the identity (e.g., the liquid's dielectric response is isotropic) [50].

But following the molecular-physics analogy further, we know that crystals, in which the molecular units only assume a discrete set of orientations, can also be optically isotropic and nonpolar. It is also well known that only cubic crystals have sufficiently high symmetry to be isotropic. This suggests that if the sum in Eq. (A1) is over the discrete subgroup of SU(2) corresponding to the symmetry operations of a tet-

TABLE IV. Modification of part of Table I, including the phase changes of the Bell states.

		Source			
		Ψ^-	Φ^-	Φ^+	Ψ^+
Bilateral $\pi/2$ rotations:	I	Ψ^-	Φ^-	Φ^+	Ψ^+
	B_x	Ψ^-	Φ^-	$i\Psi^+$	$i\Phi^+$
	B_y	Ψ^-	$-\Psi^+$	Φ^+	Φ^-
	B_z	Ψ^-	$i\Phi^+$	$i\Phi^-$	Ψ^+

rahedron (the simplest object with cubic symmetry), then the desired Werner state will result; and this turns out to be the case.

The bilateral rotations $B_{x,y,z}$ introduced in Sec. III B 3 are the appropriate starting point for building up the desired set of operations. In fact they correspond to fourfold rotations of a cube about the x, y, and z axes. This is not evident from their action on Bell states as shown in Table I where they appear to correspond to twofold operations. This is because this table does not show the effect of the B rotations on the phase of the Bell states. Phases are not required in the purification protocols described in the text, because the density matrix in all these cases is already assumed to be diagonal, so that the phases do not appear. But for the present analysis they do, so we repeat the table with phases in Table IV.

When presented in this way, it is evident that these operations are fourfold (that is, $B_i^4=I$), and indeed, they are the generators of the 24-element group of rotations of a cube, known as the group O in crystallography [50]. (It is also isomorphic to S_4, the permutation group of four objects.)

Now, as mentioned above, only the rotations which leave a tetrahedron invariant are necessary to make the density matrix isotropic. This is a 12-element subgroup of O known as T (which is isomorphic to A_4, the group of all even permutations of four objects). Written in terms of the B_i's, these 12 operations are

$$\{U_i\}=\begin{matrix} I(\text{identity}) \\ B_xB_x \\ B_yB_y \\ B_zB_z \\ B_xB_y \\ B_yB_z \\ B_zB_x \\ B_yB_x \\ B_xB_yB_xB_y \\ B_yB_zB_yB_z \\ B_zB_xB_zB_x \\ B_yB_xB_yB_x\,. \end{matrix} \quad M\to W_F \tag{A2}$$

It is easily confirmed by direct calculation, using Table IV, that this set of 12 $\{U_i\}$, when applied to a general density matrix M in Eq. (A1), results in a Werner density matrix W_F of Eq. (17).

There are a couple of special cases in which the set of rotations can be made simpler. If it is only required that the state M be taken to some Bell-diagonal state W [Eq. (29)], then a smaller subset, corresponding to the orthorhombic crystal group D_2 (an abelian four-element group) may be used:

$$\{U_i\}=\begin{matrix} I \\ B_xB_x \\ B_yB_y \\ B_zB_z\,. \end{matrix} \quad M\to W \tag{A3}$$

Finally there is another special case, which arises in some of our purification protocols, in which the density matrix W is already diagonal in the Bell basis, but is not isotropic (i.e., the triplet matrix elements are different from one another). To carry W into W_F, the discrete group in Eq. (A1) can be again be reduced, in this case to the three-element group with the elements

$$\{U_i\}=\begin{matrix} I \\ B_xB_xB_xB_y \\ B_xB_xB_xB_z\,. \end{matrix} \quad W\to W_F \tag{A4}$$

One further feature of any set $\{U_i\}$ that takes the density matrix to the isotropic form W_F, which can be used to simplify the set, is that the modified set $\{RU_i\}$, for any bilateral rotation R, also results in a Werner density matrix W_F in Eq. (A1). Since the density matrix is already isotropic, any additional rotation R leaves it isotropic. (A cubic crystal has the same dielectric properties no matter how it is rotated.) For example, if we take $R=B_x$, the three operations of Eq. (A4) take the form

$$\{U_i\}=\begin{matrix} B_x \\ B_y \\ B_z\,. \end{matrix} \quad W\to W_F \tag{A5}$$

APPENDIX B: GENERAL-NOISE ERROR CORRECTION

In this appendix we present an argument, based on twirling, that correcting amplitude and phase errors corrects every possible error. We have derived finite-block purifications under the assumption that the pairs which are affected by the environment are subject to errors of the Werner type, in which the Bell state evolves into a classical mixture of Bell states [see Eq. (63)]. But the most general effect which noise can have on a Bell state appears very different from the Werner noise model, and is characterized by the 4×4 density matrix M into which a standard Bell state Φ^+ evolves (see Fig. 5). Many additional parameters besides the fidelity

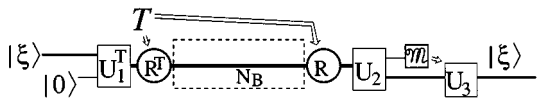

FIG. 19. If the state is subject to the initial and final rotations R^T and R (the ''twirl'' T) in the QECC of Fig. 16, then the action of the noise N_B is guaranteed to be of a simple form in which only three types of errors, amplitude, phase, or amplitude-and-phase, can occur on each qubit [13]; this corresponds to the Werner mixed state W_F in the purification picture. As described in the text, for finite-block error correction the QECC protocol will succeed even if the twirl T is not performed.

$F=\langle\Phi^+|M|\Phi^+\rangle$ are required for the specification of this general error model. A general 4×4 density matrix of course requires 15 real parameters for its specification. However, not all of these parameters define distinct errors, since any change of basis by Alice or Bob cannot essentially change the situation (in particular, the ability to purify EPR pairs cannot be changed). This says that six parameters, those involved in two different SU(2) changes of basis, are irrelevant. But this still leaves nine parameters which are required to fully specify the most general independent-error model [51]. How then does correction of just amplitude, phase, and both, deal with all of these possible noise conditions, characterized by nine continuous parameters?

To show this we will again introduce the ''twirl'' of Fig. 5, although in the end it will be removed again. Recall that any density matrix is transformed into one of the Werner type by the random twirl. (See item 5 of Sec. III A for the method of twirling the Φ^+ state.) Thus, if twirling is inserted as shown in Fig. 19, or in the corresponding places in Fig. 3, then the channel is converted to the Werner type, and the error correction criteria we describe in Sec. VI will work.

But let us consider the action of the twirl in more detail. Let us personify the twirl action T in Fig. 19 (or in the corresponding purification protocol of Fig. 3, as in Fig. 5) by saying that an agent (''Tom'') performs the twirl for the n bits by randomly choosing n times from among one of 12 bilateral rotations tabulated in Appendix A. Tom makes a record of which of these 12^n actions he has taken; he does not, however, reveal this record to Alice or Bob. Without this record, but with a knowledge that Tom has performed this action, Alice and Bob conclude that the density matrix of the degraded pairs has the Werner form. They proceed to use the protocol they have developed to purify m EPR pairs perfectly. Now, suppose that after this has been done, Tom reveals to Alice and Bob the twirl record which he has heretofore kept secret. At this point, Alice and Bob now have a revised knowledge of the state of the particle pairs which entered their purification protocol; in fact, they now know that the density matrix is just some particular rotated version of the non-Werner density matrix in which the environment leaves the EPR pairs. Nevertheless, this does not change the fact that the purification protocol has succeeded. Indeed, we must conclude that it succeeds for each of the 12^n possible values of Tom's record, and in particular it succeeds even in the case that each of Tom's n rotations was the identity operation. Thus, the purification protocol works on the original non-Werner errors, even if Tom and his twirling is completely removed. This completes the desired proof, and we will thus develop protocols for correcting Werner type errors, Eq. (63), keeping in mind their applicability to the more general case.

A slight extension of the above arguments shows that asymptotic large-block purification schemes such as our hashing protocol of Sec. III B 3 are also capable of correcting for non-Werner error. Consider a non-Bell-diagonal product density matrix of n particles, $\mathbf{M}=(M)^n$, whose fidelity is such that, after twirling, it can be successfully purified, resulting in entangled states whose final fidelity with respect to perfect singlets approaches 1 in the limit $n\to\infty$. The hashing protocol produces truly perfect singlets of unit fidelity for a likely set $\mathcal{L}$ of error syndromes containing nearly all the probability. This means that we can write $\mathbf{M}=(1-\epsilon)\mathbf{M}'+\epsilon\delta\mathbf{M}$, where $\mathbf{M}'$ can be purified with exactly 100% final fidelity. By the above arguments, $\mathbf{M}'$ can be successfully purified even if twirling is not performed. Since $\epsilon\to0$ as $n\to\infty$, the original state $\mathbf{M}$ will also be purified to fidelity approaching 1, even without twirling.

[1] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935) [reprinted in *Quantum Theory and Measurement*, edited by J. A. Wheeler and W. Z. Zurek (Princeton University Press, Princeton, 1983)].

[2] J. S. Bell, Physics **1**, 195 (1964) [reprinted in J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 1987), p. 14].

[3] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1980).

[4] J. F. Clauser and A. Shimony, Rep. Prog. Phys. **41**, 1981 (1978); A. Aspect, P. Grangier, and G. Roger, Phys. Rev. Lett. **47**, 460 (1981); P. G. Kwiat *et al.*, *ibid.* **75**, 4337 (1995).

[5] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

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

[7] C. H. Bennett and S. J. Wiesner, Phys. Rev. Lett. **69**, 2881 (1992).

[8] I. L. Chuang and R. Laflamme, Report No. quant-ph/9511003; I. L. Chuang and Y. Yamamoto, Phys. Rev. Lett. **76**, 4281 (1996).

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

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

[11] A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996); Phys. Rev. A (to be published); Report No. quant-ph/9601029.

[12] R. Laflamme, C. Miquel, J.-P. Paz, and W. H. Zurek, Phys. Rev. Lett. **77**, 198 (1996); Report No. quant-ph/9602019.

[13] A. Ekert and C. Machiavello, Phys. Rev. Lett. **77**, 2585 (1996); Report No. quant-ph/9602022.

[14] S. L. Braunstein, Report No. quant-ph/9603024.

[15] M. B. Plenio, V. Vedral, and P. L. Knight, Phys. Rev. A (to be published); Report No. quant-ph/9603022.

[16] L. Vaidman, L. Goldenberg, and S. Wiesner, Report No. quant-ph/9603031.

[17] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Phys. Rev. Lett. **76**, 722 (1996).
[18] Systems need not interact directly in order to become entangled. For example, two systems A and B that have never interacted can become entangled indirectly, by a sequence of interactions in which A becomes entangled with C and B with D, followed by a joint measurement on C and D, followed by unitary operation on A or B depending on the result of the joint measurement. This is an example of quantum teleportation [5].
[19] N. Gisin, Phys. Lett. A **154**, 201 (1991).
[20] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).
[21] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).
[22] S. Popescu, Phys. Rev. Lett. **72**, 797 (1994).
[23] S. Popescu, Phys. Rev. Lett. **74**, 2619 (1995).
[24] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978), esp. p. 237.
[25] This is an application of the Schmitt decomposition; see A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1993), p. 123f.
[26] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).
[27] M. Horodecki, P. Horodecki, and R. Horodecki, Report No. quant-ph/9605038.
[28] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).
[29] Alice and Bob should choose a basis for the Bell states which maximizes the fully entangled fraction f, Eq. (19).
[30] Reference [17] states incorrectly that a set of four discrete rotations is sufficient to simulate a random bilateral rotation; in fact, 12 are both necessary and sufficient. See Appendix A.
[31] This follows from the fact that ensembles with the same density matrix are indistinguishable by measurements performed on members of the ensembles. (See, e.g., Ref. [25], p. 75.) (Such measurements could include executing a purification protocol followed by testing the quality of the resulting purified pairs.)
[32] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Phys. Rev. A **52**, 3457 (1995).
[33] C. H. Bennett, Phys. Today **48**(10), 27 (1996).
[34] C. Macchiavello (private communication) pointed out that the yield of the recurrence protocol could be increased by substituting a deterministic B_x operation for the random twirl T'. See D. Deutsch *et al.*, Phys. Rev. Lett. **77**, 2818 (1996).
[35] P. Shor and J. A. Smolin, Report No. quant-ph/9604006.
[36] This purification strategy was worked out in a discussion with Richard Jozsa.
[37] M. Horodecki, P. Horodecki, and R. Horodecki, Report No. quant-ph/9607009.
[38] C. H. Bennett, G. Brassard, C. Crépeau, and U. Mauer, IEEE Trans. Inf. Theory **41**, 1915 (1995).
[39] W. K. Wootters and W. Zurek, Nature (London) **299**, 802 (1982).
[40] E. Knill and R. Laflamme, Report No. quant-ph/9604034.
[41] C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).
[42] For a large-block code which only error-corrects to some high fidelity, the code may not need to properly treat the no-error case because this case is highly improbable. Such a case may also arise for certain peculiar block error models from which the no-error case is excluded.
[43] If the 1-EPP is not perfect, but only produces entangled states with some high fidelity $1-\epsilon$, then it is obvious that there is at least one set of values for the $\mathcal{M}$ measurements for which the final fidelity is at least $F=1-\epsilon$, since F is the fidelity averaged over all possible measurement results. It is this set which Alice and Bob should preagree to use for the state preparation of these B particles.
[44] D. Coppersmith and E. Grossman, SIAM J. Appl. Math. **29**, 624 (1975).
[45] H. Mabuchi and P. Zoller, Phys. Rev. Lett. **76**, 3108 (1996).
[46] We note that the conditions for the ''error prevention'' (or ''error watchdogging'') protocol of Ref. [16] are a very simple subset of the conditions which we have derived for error correction: ''error prevention'' requires that Eqs. (89) and (90) be satisfied not for all i and j, but just for those i and j which refer to errors involving the same qubit. Thus it is obvious that every QECC is an ''error prevention'' scheme, but not vice versa.
[47] D. P. DiVincenzo and P. W. Shor, Phys. Rev. Lett (to be published); Report No. quant-ph/9605031.
[48] To show that k qubits cannot be used to share ℓ ebits *and* transmit m classical bits if $\ell+m>k$, suppose the contrary, and let the ℓ ebits so shared be used for superdense coding [7]. If that were done, the initial k qubits, plus ℓ subsequent qubits used to convey the treated EPR spins to Bob in the second stage of superdense coding, would together suffice to transmit from Alice to Bob a total of $2\ell+m>k+\ell$ classical bits. This is impossible, as it would imply that the intermediate quantum system, consisting of the k qubits initially transmitted plus the ℓ qubits subsequently sent during superdense coding, had more reliably distinguishable states than the $2^{k+\ell}$ dimensions of its Hilbert space.
[49] An important step has been taken in this characterization of D_2 and E by Horodecki *et al.* [37], who show that for mixed state M of a pair of qubits, $D_2(M)=0$ if and only if $E(M)=0$.
[50] See, e.g., M. Tinkham, *Group Theory and Quantum Mechanics* (Prentice-Hall, Englewood Cliffs, NJ, 1964).
[51] In the Bell basis, this restriction to nine parameters is achieved by making the matrix elements $\langle\Phi^+|M|\Phi^-\rangle$, $\langle\Phi^+|M|\Psi^-\rangle$,$\langle\Phi^-|M|\Psi^+\rangle$, and $\langle\Psi^+|M|\Psi^-\rangle$ purely real, and making $\langle\Phi^+|M|\Psi^+\rangle$ and $\langle\Phi^-|M|\Psi^-\rangle$ purely imaginary. This corresponds to making the reduced density matrices ρ_A and ρ_B diagonal, and making additional phase adjustments (z-axis rotations: see Ref. [32]) to the A and the B particles.