To the Chairman of Examiners for Part III Mathematics.

Dear Sir,

I enclose the Part III essay of Tobias S. Mansuripur.

Signed.....

James Norris

Director of Studies

Entanglement Manipulation Under Finite Resources

Tobias S. Mansuripur

Churchill College

University of Cambridge

I declare that this essay is work done as part of the Part III Examination. I have read and understood the Statement on Plagiarism for Part III and Graduate Courses issued by the Faculty of Mathematics, and have abided by it. This essay is the result of my own work, and except where explicitly stated otherwise, only includes material undertaken since the publication of the list of essay titles, and includes nothing which was performed in collaboration. No part of this essay has been submitted, or is concurrently being submitted, for any degree, diploma or similar qualification at any university or similar institution.

Signed.....

Tobias Mansuripur

Room X03

Churchill College

Cambridge CB3 0DS

Entanglement Manipulation Under Finite Resources

1. INTRODUCTION

The quantum theory has completely reshaped our understanding of the laws of physics. Many of the ideas it gave birth to have no classical analog: particles can exist in superpositions of states, measurement of a system disturbs its free evolution and yields an inherently probabilistic result, and the position and momentum of a particle cannot be simultaneously known to arbitrary accuracy. These conclusions all have dramatic consequences down to the level of a single particle. Many regard the signature effect of quantum mechanics, however, as the strong nonlocal correlations discovered by Einstein, Podolsky, and Rosen between two non-interacting particles that have interacted in the past. Such particles are referred to as entangled, meaning that the full state of the bipartite system cannot be decomposed into a tensor product of the single particle states. Take, for example, the singlet state of two spin-1/2 particles,

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle), \tag{1}$$

where $|0\rangle$ and $|1\rangle$ are the single particle states representing spin-up and spin-down (say, along the z-axis). Once the entanglement is created, the two particles can be physically separated, and the results of local measurements by two distinct observers of the spin along any axis will be perfectly anticorrelated. It is not difficult to entangle two particles when they are in the same vicinity. For example, a measurement of the total angular momentum of the two particles will cause the state to collapse into the singlet whenever the result of the measurement is $l = 0$. However, it is impossible to nonlocally (i.e. by using operations which can only act on the single particle Hilbert spaces rather than the combined space) prepare an entangled state from two previously unentangled particles .

The goal of quantum information theory is to understand the implications of quantum mechanics for the storage and transmission of information. It has been discovered that entanglement serves a very important role in this regard. An entangled pair shared between two parties serves as a resource for information transmission that has no classical analog. For example, the faithful transmission of a qubit can be accomplished in two ways: 1) the straightforward way, by physically sending the qubit from one party to another, or 2) using quantum teleportation, in which the consumption of one entangled pair, supplemented with the transfer of two bits of classical information, accomplishes the same goal. The inverse problem, of sending two bits of classical information with only one use of a quantum channel,

can also be accomplished if the two parties initially share an entangled pair, using the superdense coding protocol.

Entanglement is not an all-or-nothing resource; it can assume a continuum of values. Consider the state

$$|\Phi_\theta\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle. \tag{2}$$

Quantitative measurements of pure state entanglement will be discussed in greater detail in Sec. 2.1, but for now it will suffice to know that the state $|\Phi_\theta\rangle$ is maximally entangled (like the aforementioned singlet) only if $\theta = \pi/4$, unentangled if $\theta = 0$ or $\pi/2$, and partially entangled for any other value of θ. The importance of maximally entangled states is easily demonstrated by attempting quantum teleportation when the shared resource is in the state $|\Phi_\theta\rangle$. If Alice's goal is to transmit an arbitrary qubit in the state $\alpha|0\rangle + \beta|1\rangle$, Bob will instead receive (disregarding normalization) either the state $\alpha\cos\theta|0\rangle + \beta\sin\theta|1\rangle$ or $\alpha\sin\theta|0\rangle + \beta\cos\theta|1\rangle$ (and know with certainty which of the two he receives), after Alice has sent the result of her measurement and he has applied the fix-up operation. Unless $\theta = \pi/4$, it is clear that noisy transmission will result. Even if Bob knows the value of θ, he cannot maneuver his qubit into the intended state. Faithful transmission is only possible if Alice and Bob begin with a maximally entangled pair; these states are the most pure form of the 'entanglement resource.'

Therefore, the problem of how to faithfully transmit a quantum state is closely related to the problem of how to get Alice and Bob to share halves of a maximally entangled pair. Ideally, Alice could prepare such a state in her laboratory, and send Bob one of the particles through a noiseless channel. Of course, if they had access to a noiseless channel, then none of this trickery would be necessary in the first place! Presumably, the channel introduces some noise into the system, and by the time Bob receives his qubit, the pair is no longer maximally entangled, but perhaps only partially entangled, or even in a mixed state ρ_{AB}. Alice can use the noisy channel to send as many qubits as she wishes, say n of them, so that the shared state is in general given by $\rho_{AB}^{\otimes n}$. We seek a method by which Alice and Bob, using local operations and classical communication (LOCC), can distill the entanglement present in $\rho_{AB}^{\otimes n}$ into m maximally entangled states, such as singlets $|\Psi^-\rangle^{\otimes m}$. The yield of the process is given by m/n, and is referred to as the distillable entanglement. In this paper, we focus mainly on the asymptotic case, by allowing $n \to \infty$. Because the distilled states can be used for teleportation in place of a noiseless quantum channel, entanglement distillation

is intimately related to quantum error correction.

The remainder of the paper is organized as follows. In Sec. 2, we address the situation where Alice and Bob share partially entangled pure states. We explain the quantitative measure of entanglement for pure states and will prove various upper bounds of the distillable entanglement of a state. Then, two purification procedures a described: a non-asymptotic method–called the Procrustean method–which can distill entanglement under finite resources, and an efficient asymptotic method. These results are mostly a recapitulation of the ideas presented in [1]. In Sec. 3, we generalize these ideas for mixed states. Again, entanglement measures are discussed, and explicitly derived for the special case of Bell-diagonal mixed states. Then, three distillation protocols are discussed: the recurrence method, the universal hashing method, and a method for purifying non-Bell diagonal mixed states. Finally, the relationship between entanglement distillation and quantum error correction is explored. This second half of the paper is mostly a recapitulation of the results presented in [2].

2. PURE STATES

2.1. Entanglement measures

When presented with two different bipartite pure states, it is not immediately obvious how to compare the amount of entanglement in them. One well-known measure of entanglement of a bipartite pure state is its Schmidt number N_{Sch}, defined as the number of non-zero coefficients λ_i in the Schmidt decomposition of the state

$$|\Upsilon\rangle = \sum_{i=1}^{2} \sqrt{\lambda_i} |i_A\rangle \otimes |i_B\rangle, \tag{3}$$

where $\{|i_A\rangle\}$ ($\{|i_B\rangle\}$) forms an orthonormal basis for $\mathcal{H}_A$ ($\mathcal{H}_B$). This measure is not robust enough for our purposes: both partially and maximally entangled states have $N_{Sch} = 2$.

A better approach to determining how much entanglement a state $|\Upsilon\rangle$ contains is to consider the goal: to allow Alice and Bob to operate on the state $|\Upsilon\rangle^{\otimes n}$ using LOCC in order to prepare a new state that will allow them to faithfully teleport some number of qubits. The number of qubits that they can teleport is equal to the number of maximally entangled states that they can distill, say m, from the initial shared state. Thus, by converting $|\Upsilon\rangle$ into m

maximally entangled states, we have found a common currency by which to compare $|\Upsilon\rangle$ to other bipartite states. We call the number m/n the 'distillable entanglement' D of the state $|\Upsilon\rangle$, and it is to be taken in the asymptotic sense $n \to \infty$. Conversely, we could imagine a situation where Alice and Bob are provided with maximally entangled states to begin with, and asked to prepare the state $|\Upsilon\rangle^{\otimes n}$ using LOCC. This is known as entanglement dilution. The minimum number of pairs they require, say m', represents a sort of cost of creating the state $|\Upsilon\rangle$, and we refer to the number m'/n as the 'entanglement of formation' E_F.

Towards this end, it will turn out to be helpful to define the quantity

$$E(|\Upsilon\rangle) = S(\rho_A) = S(\rho_B), \tag{4}$$

where $S(\rho) = -\mathrm{Tr}(\rho \log_2 \rho)$ is the von Neumann entropy, and $\rho_A = \mathrm{Tr}_B |\Upsilon\rangle\langle\Upsilon|$ is the reduced density matrix seen by Alice (and likewise for ρ_B). For now, we will refer to the quantity E simply as the 'entanglement,' without specifying its relation to the distillable entanglement or the entanglement of formation (though eventually we will show that they are all equal). We motivate the definition of E by enumerating four desirable properties.

1) The entanglement of a pair of qubits ranges from 0, for an unentangled state, to 1, for a maximally entangled pair. For convenience, we define an *ebit* as the amount of entanglement present in a state for which $E = 1$.

2) The entanglement of separable systems is additive. This follows from the multiplicativity of the partial trace and the additivity of the von Neumann entropy over tensor products. Consider two bipartite states $|\Psi\rangle$ and $|\Phi\rangle$:

$$\begin{aligned} E(|\Psi\rangle \otimes |\Phi\rangle) &= S\left(\mathrm{Tr}_B(|\Psi\rangle\langle\Psi| \otimes |\Phi\rangle\langle\Phi|)\right) \\ &= S\left(\mathrm{Tr}_B|\Psi\rangle\langle\Psi| \otimes \mathrm{Tr}_B|\Phi\rangle\langle\Phi|\right) \\ &= E(|\Psi\rangle) + E(|\Phi\rangle). \end{aligned} \tag{5}$$

Thus, if $E = E(|\Upsilon\rangle)$, then $E(|\Upsilon\rangle^{\otimes n}) = nE$.

3) Using only local unitary operations, Alice and Bob cannot change E. In other words, E is conserved under operations of the form $U = U_A \otimes U_B$ where U_A and U_B are unitary. Unitary operations can only change the eigenvectors, but not the eigenvalues of the reduced density matrix in Alice's possession. The result follows because the von Neumann entropy is a function only of the eigenvalues of a matrix.

4) Using nonunitary operations, Alice and Bob cannot increase the expectation value of E. There are two types of nonunitary operations to consider: measurement, which maps pure states to pure states, and discarding of a subsystem, which maps pure states to mixed states. We will address the first case in the following theorem, and delay the second result until the full treatment of mixed state entanglement measures in Sec. 3.1.

Theorem 1: Suppose Alice and Bob share the state $|\Upsilon\rangle$, and Alice performs a measurement on her subsystem which will collapse the full system to the state $|\Upsilon_k\rangle$ with probability p_k. Then the expected entanglement of the final states is no greater than the entanglement of the original state:

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

Proof. Suppose Alice and Bob agree beforehand that Alice will perform some measurement at a specified time, say noon. Before noon, the state of Bob's qubits is simply given by $\rho_B = \mathrm{Tr}_A|\Upsilon\rangle\langle\Upsilon|$. Immediately after noon, Bob knows that Alice has performed the measurement, and he knows the probabilities p_k of each of the possible resulting states $|\Upsilon_k\rangle$, but he does not know Alice's measurement result. Therefore, his new density matrix ρ'_B is the weighted average of all the reduced density matrices ρ_k that he could now possess, where $\rho_k = \mathrm{Tr}_A|\Upsilon_k\rangle\langle\Upsilon_k|$. However, ρ_B must equal ρ'_B, otherwise Bob could perform some measurement on his system immediately after noon and conceivably learn something about the result of Alice's measurement, in violation of the principle that information cannot travel faster than light speed. Hence, we can write

$$\rho_B = \sum_k p_k \rho_k. \tag{7}$$

Taking the von Neumann entropy of both sides leads to

$$S(\rho_B) = S\left(\sum_k p_k \rho_k\right) \geq \sum_k p_k S(\rho_k) \tag{8}$$

where the inequality is a consequence of the concavity of the entropy. The left-hand side is the entanglement of the original state $|\Upsilon\rangle$, and the right-hand side is the expected entanglement of the post-measurement state. QED.

One corollary of the theorem is that the distillable entanglement can never be greater than the entanglement of formation,

$$D \leq E_F. \tag{9}$$

To prove this by contradiction, assume that $D > E_F$. This means, given a state $|\Upsilon\rangle^{\otimes n}$, we can distill m singlets. But we can then recreate the state $|\Upsilon\rangle^{\otimes n}$ using only m' singlets, where $m' < m$, leaving us with the original state we began with, as well as an excess of $m - m'$ singlets. Since we can *reliably* repeat the process, we can continue to generate an unlimited number of singlets, in clear violation of the above theorem. Hence, E_F must upper bound D. (Note: it is possible to find an *unreliable* method which can occasionally, with luck, result in a final state that has more entanglement than the initial state, as will be demonstrated by the Procrustean method. This does not violate the above theorem, because the *expectation* of the final entanglement is still less than the original entanglement. The distillable entanglement is defined in the asymptotic sense, where luck is negligible, and results of measurements tend toward the expected value.)

Another important corollary is that the entanglement of formation can never be smaller than this quantity E:

$$E_F \geq E. \tag{10}$$

This follows easily because any reliable procedure which begins with an amount of entanglement E_F and creates a state with entanglement E, where $E > E_F$, would violate the theorem.

2.2. Entanglement of formation

Because E_F is an upper bound on D, it will serve us well to look into the process of entanglement dilution before moving on to distillation. As mentioned before, Alice and Bob are now asked to prepare the state $|\Upsilon\rangle^{\otimes n}$ using a minimal number of singlets. If the minimum number of pairs they require is m, then $E_F = m/n$. One way to accomplish this is for Alice to prepare the state $|\Upsilon\rangle^{\otimes n}$ in her laboratory, which requires $2n$ qubits. Let us say that at this point, the state lives in the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_C$, where A is the system of n qubits that Alice intends to keep for herself, and C comprises the n qubits intended for Bob. Then, she teleports the system C to Bob, using n singlets which they already share. In this case, $m/n = 1$, which is always an upper bound of E_F and D, so this particular protocol has not helped us learn anything.

An improved protocol calls for Alice to apply Schumacher compression (i.e. projection onto the typical subspace) to the state of system C after she has prepared it. This results

in a state $|\Upsilon'\rangle$ whose entanglement fidelity with $|\Upsilon\rangle^{\otimes n}$ goes to 1 for large enough n. Most importantly, the compressed state lives in a space of dimension $2^{n(S(\rho_C)+\epsilon)}$, the dimension of the typical subspace, where $\rho_C = \mathrm{Tr}_A|\Upsilon\rangle\langle\Upsilon|$ and ϵ has the usual meaning associated with an 'ϵ-typical sequence,' and can be made as small as desired. Now, system C still consists of n qubits–no qubits have suddenly disappeared. However, because we have compressed the dimension of the Hilbert space, it is possible to apply unitary operations to C and concentrate all of the information from the state into $nS(\rho_C)$ qubits, leaving $n - nS(\rho_C)$ unentangled qubits which can be discarded. (An explicit example of this is worked out in Sec. 2.4.1 for a similar case, which is why we gloss over the details here.) Then, Alice needs to use only $m = nS(\rho_C)$ singlets to teleport the information to Bob. Bob's system B has now assumed the compressed form of C, so $\rho_B = \rho_C$. To decompress, he appends $n - nS(\rho_C)$ qubits to his system and applies the inverse unitary transformations. In the end, Alice and Bob share the state $|\Upsilon'\rangle$, which is nearly indistinguishable from the intended state $|\Upsilon\rangle^{\otimes n}$, so we declare that they have succeeded. For this protocol, we have achieved $m/n = S(\rho_B) = E(|\Upsilon\rangle)$.

We must stress that we have described only one particular dilution protocol and obtained a value for m/n. There is usually no guarantee that a more efficient protocol does not exist. In this case, however, we see that we have actually reached equality in Eq. (10). Therefore, it is certain that we have found the most efficient protocol, and we have succeeded in finding the entanglement of formation

$$E_F(|\Upsilon\rangle) = E(|\Upsilon\rangle) \tag{11}$$

for all bipartite pure states $|\Upsilon\rangle$. Henceforth, when dealing with pure states, we may use E and E_F interchangeably.

2.3. Procrustean method of purification

Suppose Alice and Bob each possess one half of an entangled pair in the state $|\Phi_\theta\rangle$ of Eq. (2). The entanglement of formation of this state is $E_F = H(\cos^2\theta)$ (plotted as the top curve in Fig. 1) where H is the binary entropy function given by $H(x) = -x\log(x) - (1-x)\log(1-x)$. Furthermore, let us assume that at least one of them knows the value of θ. This is a special case, for which there exists a method of converting a single partially entangled pair into a maximally entangled pair, albeit probabilistically, and rather inefficiently when

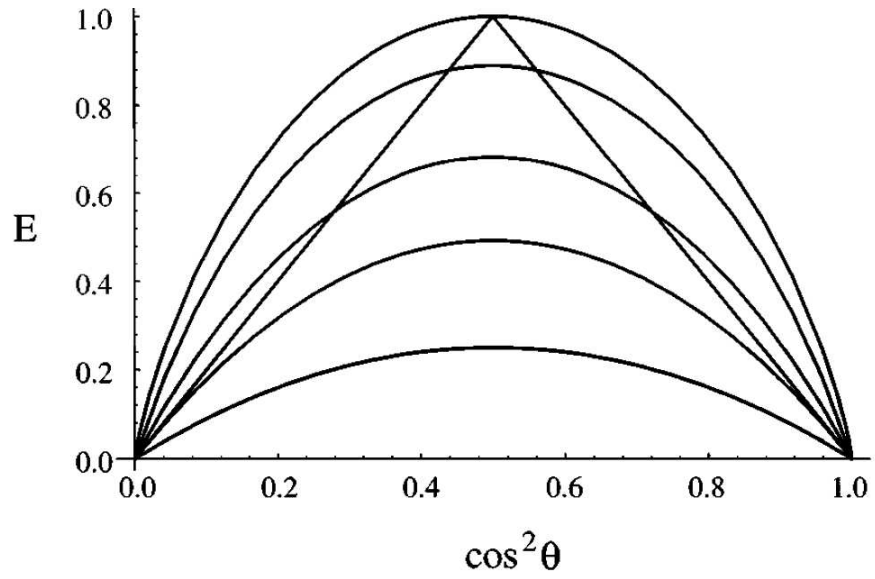


FIG. 1: Reprinted from [1]. Efficiency of pure state distillation procedures. The top curve is the entanglement of formation of $|\Phi_\theta\rangle$, a theoretical upper bound on the distillable entanglement. The yield of distilled entanglement from the Schmidt projection applied to n= 32, 8, 4, and 2 is plotted on successively lower smooth curves. The Λ-shaped curve gives the expected yield of distilled entanglement using the Procrustean method.

compared to the asymptotic method treated in Sec. 2.4. However, it is a very instructive case to begin with for two reasons. First of all, it demonstrates that Alice and Bob do not need to have multiple copies of a state $|\Psi\rangle^{\otimes n}$ in order to distill entanglement; they can succeed some percentage of the time with only finite resources. Secondly, the method is a demonstration of 'gambling' with entanglement, whereby Alice and Bob can occasionally increase E, in spite of the decreasing nature of the expectation of E. The goal of the procedure is to equalize the probability amplitudes of the two terms in Eq. (2), by 'cutting off' part of the larger probability amplitude. Hence the relation to Procrustes, leaving us with a rather brutal name for quite a beautiful procedure.

Rather than present a purely mathematical description of the procedure in terms of abstract unitaries and measurement operators, we will describe the table-top setup (Fig. 2) that can be used to perform the distillation, which provides physical insight as well as clarifies the mathematics. Therefore, it is helpful to think in terms of optics, so let $|0\rangle$ represent the vertical and $|1\rangle$ the horizontal polarization state of a photon. Let us assume for now that

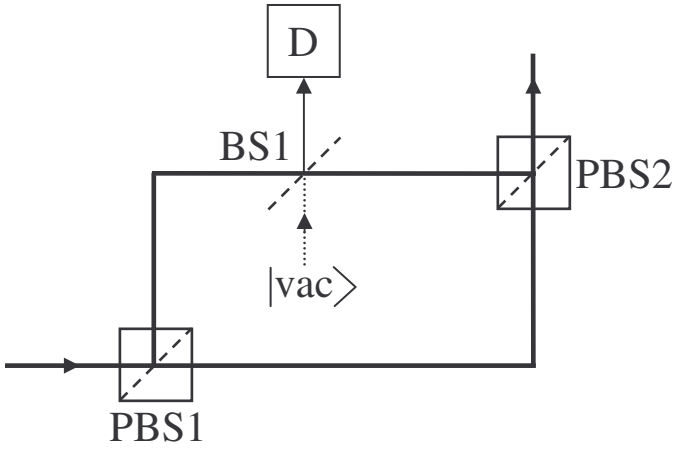


FIG. 2: Modified from [3]. Set-up of hypothetical experiment used to apply the Procrustean method of entanglement distillation to partially entangled photons in the state $|\Phi_\theta\rangle$. PBS1 is a polarizing beam splitter which passes the horizontal polarization component and deflects the vertical component by 90°. BS1 is a partial beam splitter with transmission coefficient $t = \tan\theta$ used to attenuate the probability amplitude of the vertical component (assuming $0 \leq \theta \leq \pi/4$). PBS2 recombines the polarization components. If the detector D registers a photon, all entanglement is lost. If not, the output of PBS2 is a maximally entangled state.

$0 < \theta < \pi/4$, so that the state $|00\rangle$ has a greater probability amplitude than $|11\rangle$. The goal is to somehow reduce the probability amplitude of $|00\rangle$ from $\cos\theta$ down to $\sin\theta$, so that what remains is a maximally entangled state. By virtue of the entanglement between the two photons, only one of the parties needs to perform the perform the procedure, so let us nominate Bob. All of the operations that Bob will perform will be of the form $I \otimes U$, where I is the identity (acting on Alice's qubit), and U is a unitary which acts on Bob's qubits and any ancillas he chooses to introduce.

First, Bob passes his photon through a polarizing beam splitter (PBS1) which passes the horizontal component of the photon through and deflects the vertical component by 90°. This is useful because it does not alter the quantum state, yet will allow us to operate separately on the horizontal and vertical components. Next, Bob passes the vertical component of the original photon through a partial beam splitter (BS1) with real transmission and reflection coefficients t and r, such that $t^2 + r^2 = 1$. To understand the effect of this operation mathematically, we must realize that the beam splitter works probabilistically when a single

photon approaches it. Either the photon gets passed through with probability amplitude t, or gets reflected with amplitude r. In the latter case, the result is effectively a photon in a 'new' Hilbert space. To keep track of this photon, we needed to introduce an ancilla before BS1, initialized in the vacuum state $|\text{vac}\rangle$, so that our initial state is in fact $|\Phi_\theta\rangle \otimes |\text{vac}\rangle$. Then the beam splitter can be modeled as a unitary U such that

$$U(|0\rangle_B \otimes |\text{vac}\rangle) = t|0\rangle_B \otimes |\text{vac}\rangle + r|\text{vac}\rangle_B \otimes |0\rangle, \tag{12}$$

and U behaves as the identity when operating on any other state. The components of the original photon are recombined at a second polarizing beam splitter (PBS2), so that the full effect of the apparatus is

$$|\Phi_\theta\rangle \otimes |\text{vac}\rangle \rightarrow r\cos\theta|0\rangle_A \otimes |\text{vac}\rangle_B \otimes |0\rangle + (t\cos\theta|00\rangle_{AB} + \sin\theta|11\rangle_{AB}) \otimes |\text{vac}\rangle. \tag{13}$$

Looking at the second term on the right, the probability amplitudes of $|00\rangle$ and $|11\rangle$ will be equal if Bob chooses the transmission coefficient $t = \tan\theta$ (this step makes explicit use Bob's prior knowledge of θ). With this choice, the final state $|\Phi'_\theta\rangle$ is

$$|\Phi'_\theta\rangle = \sqrt{\cos(2\theta)}|0\rangle_A \otimes |\text{vac}\rangle_B \otimes |0\rangle + \sqrt{2}\sin\theta|\Phi^+\rangle_{AB} \otimes |\text{vac}\rangle \tag{14}$$

where

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{15}$$

is a maximally entangled pair. The final step is to place a photon detector (D) in the path of the reflected beam from BS1 to measure the state of the ancilla. If the detector clicks, then the photon was indeed reflected, and the full state collapses to an unentangled state. In this case, Bob tells Alice to discard her qubit, and all the entanglement has been lost. However, with probability $2\sin^2\theta$ the ancilla is in the state $|\text{vac}\rangle$, in which case Bob's photon emerges from the experiment unabsorbed, and more importantly, in a maximally entangled state $|\Phi^+\rangle$ with Alice's photon. The yield of the Procrustean method, defined as the expected number of distilled maximally entangled pairs per input pair, is therefore equal to $2\sin^2\theta$ for $0 \leq \theta \leq \pi/4$. If $\pi/4 \leq \theta \leq \pi/2$, then BS1 is placed in the path of the horizontal component, and by similar reasoning the yield is equal to $2\cos^2\theta$. The yield is plotted as the Λ-shaped curve in Fig. 1.

In essence, the Procrustean method aims to attenuate the probability amplitude of one component, at the risk of losing everything. It was originally implemented as a means

of distinguishing non-orthogonal quantum states [3]: by attenuating one component, two non-orthogonal states can be made orthogonal and then easily distinguished. However, this cannot be done with perfect reliability for the same reason that the entanglement is occasionally completely lost. The procedure we have outlined corresponds mathematically to a two-outcome POVM, and provides a neat illustration of the usually abstract feature that a POVM can be implemented by appending ancillas and using unitary operations.

2.4. Asymptotic purification

To find the most efficient method of purifying partially entangled pure states, we go to the asymptotic case by allowing Alice and Bob to share the state $|\Phi_\theta\rangle^{\otimes n}$. This method has the advantage of not requiring knowledge of θ. Before treating the general case, we will work an example for $n = 2$.

2.4.1. Worked example for $n = 2$

(Notation: when we use the shortcut notation to avoid writing the tensor product symbol, as in $|01\rangle$, we typically mean that the first entry corresponds to the state of Alice's qubit and the second entry to Bob's qubit. To be explicit, we could include this information in a subscript as in $|01\rangle_{AB}$. When dealing with multiple pairs of qubits, it is often helpful to isolate all of Alice's qubits in one ket, and all of Bob's in another, so that $|01\rangle_{AB} \otimes |01\rangle_{AB} \equiv |00\rangle_A \otimes |11\rangle_B$. In cases of ambiguity, subscripts will be used to make this clear.)

Alice and Bob share the state

$$
\begin{aligned}
|\Phi_\theta\rangle^{\otimes 2} &= (\cos\theta|00\rangle_{AB} + \sin\theta|11\rangle_{AB})^{\otimes 2} && (16)\\
&= \cos^2\theta(|00\rangle_A \otimes |00\rangle_B) && (17)\\
&\quad + \cos\theta\sin\theta(|01\rangle_A \otimes |01\rangle_B + |10\rangle_A \otimes |10\rangle_B) && (18)\\
&\quad + \sin^2\theta(|11\rangle_A \otimes |11\rangle_B). && (19)
\end{aligned}
$$

Notice that $N_{Sch} = 4$, but not all of the Schmidt coefficients are equal, which is why the state is only partially entangled. Now, Alice makes a measurement of the z-component of the cumulative spin of all of her particles. Because the z-component of spin is additive over multiple electrons, this measurement effectively tells her how many of her particles are

spin-down, without specifying which particles in particular. Let the number of spin-down particles be k, then the measurement result can be either $k = 0, 1,$ or 2. If $k = 0$ or 2, then the state will collapse to line 18 or 19 (prior to renormalization), and the resulting state has zero entanglement. (The probability of these results will go to zero as $n \to \infty$.) With probability

$$p_{k=1} = 2\cos^2\theta \sin^2\theta \tag{20}$$

the measurement result is $k = 1$, and the state collapses to

$$|\Phi_\theta\rangle^{\otimes n} \to |\Phi_{k=1}\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle_A \otimes |01\rangle_B + |10\rangle_A \otimes |10\rangle_B\right). \tag{21}$$

For this state, clearly $N_{Sch} = 2$, and both Schmidt coefficients are equal. The same can be said of the state

$$|\Phi'\rangle = \frac{1}{\sqrt{2}}\left(|\downarrow\downarrow\rangle_{AB} + |\uparrow\uparrow\rangle_{AB}\right) \otimes |\psi\rangle_A \otimes |\phi\rangle_B \tag{22}$$

$$= \frac{1}{\sqrt{2}}\left(|\downarrow \psi\rangle_A \otimes |\downarrow \phi\rangle_B + |\uparrow \psi\rangle_A \otimes |\uparrow \phi\rangle_B\right), \tag{23}$$

where $|\uparrow\rangle$ and $|\downarrow\rangle$ are states of opposite spin along an arbitrary axis, and $|\psi\rangle$ and $|\phi\rangle$ are arbitrary single particle states. The state $|\Phi'\rangle$ has the advantage of being expressed as a maximally entangled pair along with unentangled qubits. Now, any two pure states of a composite system AB with identical Schmidt coefficients can be transformed into each other using an operator of the form $U_A \otimes U_B$, i.e. local unitary operations. (We will demonstrate the explicit transformation in this example; the generalization for arbitrary states and dimensions follows because a unitary operator always maps between two orthonormal bases, in this case the Schmidt bases.) We want to transform $|\Phi_{k=1}\rangle$ into $|\Phi'\rangle$ so that all of the entanglement is concentrated in a single pair. To make this easier, choose $|\downarrow\rangle = |0\rangle$, $|\uparrow\rangle = |1\rangle$, and $|\psi\rangle = |\phi\rangle = |0\rangle$. Then

$$|\Phi'\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle_A \otimes |00\rangle_B + |10\rangle_A \otimes |10\rangle_B\right). \tag{24}$$

By comparison of Eqs. (21) and (24), Alice and Bob both need to perform the same operation U on their qubits which satisfies

$$U|01\rangle = |00\rangle$$

$$U|10\rangle = |10\rangle.$$

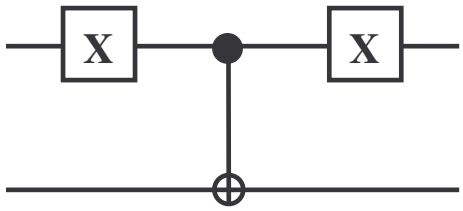


FIG. 3: Quantum circuit employed separately by both Alice and Bob to implement the operation U on their qubits used to transform $|\Phi_{k=1}\rangle$ into one maximally entangled pair and two unentangled qubits.

The action of U on the remaining two basis states is irrelevant, but to preserve unitarity it must map one orthogonal basis to another orthogonal basis, so let us simply choose

$$\begin{aligned} U|00\rangle &= |01\rangle \\ U|11\rangle &= |11\rangle. \end{aligned}$$

Thus, the matrix representation of U is

$$U = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{25}$$

and can be implemented using NOT-gates (X) and controlled-NOT gates with the quantum circuit in Fig. 3. The yield of the protocol is defined as the expected number of distilled maximally entangled pairs divided by the number of input pairs:

$$\text{yield}_n = \frac{\langle \text{distilled pairs} \rangle}{n}. \tag{26}$$

In this example, the numerator is equal to $p_{k=1}$ since one maximally entangled pair is distilled whenever the measurement result is $k = 1$, and no pairs are distilled otherwise. Therefore, the yield of this protocol is

$$\text{yield}_{n=2} = \cos^2\theta \sin^2\theta, \tag{27}$$

which is plotted as the lowest curve in Fig. (1). The vertical axis of Fig. (1) is measured in E because the yield can equivalently be thought of as the distilled entanglement per input pair.

2.4.2. General case

Alice and Bob share the state

$$|\Phi_\theta\rangle^{\otimes n} = (\cos\theta|00\rangle_{AB} + \sin\theta|11\rangle_{AB})^{\otimes n}\,. \tag{28}$$

Their goal is to distill m maximally entangled states, so the final state will have the form

$$|\Phi'\rangle = \left(\frac{|00\rangle_{AB} + |11\rangle_{AB}}{\sqrt{2}}\right)^{\otimes m} \otimes \prod_{i=1}^{n-m} \left(|\psi_i\rangle_A \otimes |\phi_i\rangle_B\right), \tag{29}$$

where the $|\psi_i\rangle$ and $|\phi_i\rangle$ are single particle states which can be discarded. For this state, $N_{Sch} = 2^m$ and all of the Schmidt coefficients are equal. Our protocol will demonstrate how Alice and Bob manipulate $|\Phi_\theta\rangle^{\otimes n}$ into satisfying these two criteria, after which $|\Phi'\rangle$ can be created using local unitary operators, as demonstrated in the worked example. The protocol consists of two main steps: first, equalize the Schmidt coefficients and second, force the Schmidt number to be a power of 2.

When $|\Phi_\theta\rangle^{\otimes n}$ is binomially expanded, the terms are naturally grouped into $n+1$ distinct subspaces, each with coefficient $\cos^{n-k}(\theta)\sin^k(\theta)$, where k ranges from 0 to n and serves as a label for each subspace. There is also a physical interpretation of k: a given state of Alice's (or Bob's) qubits belongs to subspace k only if it contains exactly k spin-down particles. Thus, each k-space comprises $\binom{n}{k}$ orthogonal states, each of which has the same probability amplitude. We can then achieve the first criterion–equal Schmidt coefficients–simply by projecting into a k-space, which occurs when Alice (or Bob, or both) measures the z-component of the total spin. We refer to this measurement as the Schmidt projection. The probability of outcome k is

$$p_k = \binom{n}{k} \left(\cos^2\theta\right)^{n-k} \left(\sin^2\theta\right)^k, \tag{30}$$

and the resulting state $|\Phi_k\rangle$ has Schmidt number $N_{Sch,k} = \binom{n}{k}$. It is easy to show that the state $|\Phi_k\rangle$ is characterized by entanglement

$$E(|\Phi_k\rangle) \equiv E_k = \log_2 N_{Sch,k}, \tag{31}$$

which upper bounds the number of distillable pairs.

Of course, it is nonsensical to speak of distilling a non-integer number of pairs, which is why $N_{Sch,k}$ must be made into a power of two. This is done by projecting the support of

$|\Phi_k\rangle$ onto a subspace of dimension 2^m. For example, suppose we started with $n = 3$ partially entangled pairs and Schmidt projected into the $k = 1$ subspace. The resulting state has $N_{Sch,k=1} = 3$ and is given by

$$|\Phi_{k=1}^{n=3}\rangle = \frac{1}{\sqrt{3}} \left(|001\rangle_A \otimes |001\rangle_B + |010\rangle_A \otimes |010\rangle_B + |100\rangle_A \otimes |100\rangle_B \right). \tag{32}$$

Of the three states in the superposition, Alice need only choose any two of them to project onto. For instance, she can perform the measurement characterized by the projection operators

$$P_{2^m} = ({}_A|001\rangle\langle 001|_A + {}_A|010\rangle\langle 010|_A) \otimes I_B \tag{33}$$

$$P_{ex} = {}_A|100\rangle\langle 100|_A \otimes I_B \tag{34}$$

where P_{2^m} projects onto the desired subspace, and P_{ex} projects onto the excess states. Of course, if $N_{Sch,k}$ is nowhere near a power of two, then there is a good chance that the result of this measurement is to project into the excess space, in which case most or all of the entanglement is lost. The probability of successfully projecting into the 2^m-dimensional subspace is given by

$$p_{2^m} = \langle \Phi_k | P_{2^m} | \Phi_k \rangle = \frac{2^m}{N_{Sch,k}} = \frac{1}{1+\varepsilon} \tag{35}$$

where we have expressed the Schmidt number as $N_{Sch,k} = 2^m(1+\varepsilon)$, $0 \leq \varepsilon < 1$. Furthermore, even if this successful projection is achieved, an amount of entanglement $\Delta E = \log_2(1+\varepsilon)$ is thrown away. To ensure both a high success rate of projection into the correct subspace and minimal waste of entanglement requires that ε be made very small. However, it seems we have little control over ε, since the Schmidt projection simply hands us a state with $N_{Sch,k} = \binom{n}{k}$, which is not necessarily close to a power of two.

There is a clever method to deal with this problem. Rather than give Alice and Bob only one state $|\Phi_\theta\rangle^{\otimes n}$, let them repeat the Schmidt projection step on l copies of the state. Each projection will in general yield different values of k: k_1, k_2,..., k_l. The total state $|\Phi_l\rangle$ is now the tensor product of all the residual states $|\Phi_k\rangle$,

$$|\Phi_l\rangle = |\Phi_{k_1}\rangle \otimes |\Phi_{k_2}\rangle \otimes \cdots \otimes |\Phi_{k_l}\rangle. \tag{36}$$

The multiplicity of the Schmidt number over tensor products means that

$$N_{Sch,l} = N_{Sch,k_1} N_{Sch,k_2} \cdots N_{Sch,k_l} \tag{37}$$

and furthermore, all of the Schmidt coefficients of $|\Phi_l\rangle$ will still remain equal. This procedure is continued until $N_{Sch,l}$ lies between 2^m and $2^m(1+\varepsilon)$, where ε can be made arbitrarily close to zero. (That $N_{Sch,l}$ will satisfy this criterion for some l will be demonstrated in Sec. 2.4.3.) At this point, Alice and Bob will each have nl qubits. Alice performs the projection onto the 2^m-dimensional subspace, which will succeed with probability $p_{2^m} = 1/(1+\varepsilon) \simeq 1-\varepsilon$. Moreover, no entanglement is thrown away as $\varepsilon \to 0$.

2.4.3. Efficiency

For convenience, let us denote the entanglement of the state $|\Phi_\theta\rangle$ as

$$E(|\Phi_\theta\rangle) \equiv E_\theta = H(\cos^2\theta). \tag{38}$$

Alice and Bob start with entanglement nE_θ in dilute form in the state $|\Phi_\theta\rangle^{\otimes n}$. The Schmidt projection results in state $|\Phi_k\rangle$, having entanglement E_k, with probability p_k. Using Eqs. (30) and (31), the expected value of entanglement is therefore

$$\langle E_k\rangle = \sum_{k=0}^{n} \binom{n}{k} (\cos^2\theta)^{n-k} (\sin^2\theta)^k \log_2 N_{Sch,k}. \tag{39}$$

The yield of concentrated entanglement from the Schmidt projection, defined as $\langle E_k\rangle/n$, is plotted in Fig. 1 for $n = 32$, 8, 4 and 2 as the four lower smooth curves, where successively lower curves correspond to smaller values of n. The yield is seen to approach the theoretical limit E_F, plotted as the top curve, as n increases. As $n \to \infty$, we will show that that the Schmidt projection results in E_k arbitrarily close to nE_θ with probability greater than $1-\delta$. In other words, no entanglement is lost. The Schmidt projection is asymptotically perfectly efficient.

Each of Alice's qubits is in the mixed state

$$\rho_A = \begin{pmatrix} \cos^2\theta & 0 \\ 0 & \sin^2\theta \end{pmatrix}. \tag{40}$$

Hence, she can treat the full state of her qubits as a superposition of all n-bit sequences of 0s and 1s, where within each sequence, the probability of a 0 is equal to $\cos^2\theta$ and the probability of a 1 is given by $\sin^2\theta$. Each sequence corresponds to a direct product state of her qubits, so we can apply an important result from the well-known Typical Subspace

Theorem: the probability of projecting onto the 'ϵ-typical subspace' is greater than $1-\delta$, for arbitrarily small δ. Here, the 'ϵ-typical subspace' is the subspace spanned by all states associated with a sequence which occurs with probability between

$$2^{-n(H(\cos^2\theta)+\epsilon)} \leq p(\epsilon\text{-typical sequence}) \leq 2^{-n(H(\cos^2\theta)-\epsilon)}. \tag{41}$$

Note that the ϵ-typical states will in general not be confined to a single k-space. Also, if a state within one k-space is typical, then all of the states within that k-space are typical, because each state is equiprobable.

By the Typical Subspace Theorem, the Schmidt projection will project into a k-space whose states are ϵ-typical–in other words, a k-space which is a subspace of the typical subspace–with probability greater than $1-\delta$. It remains to determine which k-spaces are subspaces of the typical subspace. The subspace $k = n\sin^2\theta$ consists of states which occur with probability $2^{-nH(\cos^2\theta)}$, hence it always belongs to the typical subspace for any ϵ. To determine which other k-spaces belong to the typical subspace, let us vary k about the value $k = n\sin^2\theta$ by defining

$$k_\pm = n(\sin^2\theta \pm \epsilon'). \tag{42}$$

The probability of a sequence belonging to the $k_\pm$-space is given by

$$\begin{aligned} p(\text{sequence } \in k_\pm\text{-space}) &= (\cos^2\theta)^{n-n(\sin^2\theta\pm\epsilon')}(\sin^2\theta)^{n(\sin^2\theta\pm\epsilon')} \\ &= 2^{-n\left(H(\cos^2\theta)\mp\epsilon'\log_2\tan^2\theta\right)}. \end{aligned} \tag{43}$$

Therefore, if we choose $\epsilon = |\epsilon'\log_2\tan^2\theta|$, then a sequence is ϵ-typical if and only if it belongs to a k-space which obeys

$$n(\sin^2\theta - \epsilon') \leq k \leq n(\sin^2\theta + \epsilon). \tag{44}$$

Since ϵ and ϵ' are proportional, as one goes to zero so does the other. Putting it all together, the Schmidt projection will project into a k-space obeying (44) with probability greater than $1-\delta$. In other words, k can be made arbitrarily close to $n\sin^2\theta$ with high probability.

The entanglement in the residual state $|\Phi_k\rangle$ is just a function of k, and can therefore be

made arbitrarily close to

$$\begin{aligned} E_{k=n\sin^2\theta} &= \log_2 \binom{n}{n\sin^2\theta} && (45)\\ &= \log_2(n!) - \log_2((n\cos^2\theta)!) - \log_2((n\sin^2\theta)!) && (46)\\ &\simeq nH(\cos^2\theta) && (47)\\ &= nE_\theta && (48) \end{aligned}$$

with probability greater than $1-\delta$, where Stirling's formula $\ln n! = n\ln n - n$ has been used to reach the third line. This is the same amount of entanglement that was present in the original state $|\Phi_\theta\rangle^{\otimes n}$. Thus, we have proved our claim that the Schmidt projection is asymptotically perfectly efficient.

It remains to show that the projection onto a subspace of dimension 2^m conserves entanglement asymptotically. In the previous section it was shown that the state $|\Phi_l\rangle$ given by Eq. (36) has Schmidt number $N_{Sch,l}$ given by Eq. (37). It was claimed that $N_{Sch,l}$ would, by continuing to increase l, eventually satisfy

$$N_{Sch,l} = 2^m(1+\varepsilon) \tag{49}$$

for some m and arbitrarily small, positive ε. We now prove this claim.

By taking the logarithm of both sides of Eq. (49), we can rewrite the necessary condition as

$$\log_2 N_{Sch,l} - m \leq \log_2(1+\varepsilon), \tag{50}$$

where m is the greatest integer less than $\log_2 N_{Sch,l}$. In other words, we only care about the fractional part (the mantissa) of the logarithm, which we will denote by

$$Z_l = \mathrm{frac}(\log_2 N_{Sch,l}). \tag{51}$$

When the Schmidt number evolves

$$N_{Sch,l} \to N_{Sch,l+1} = N_{Sch,l}N_{Sch,k_{l+1}}, \tag{52}$$

the effect on Z_l is

$$Z_l \to Z_{l+1} = \mathrm{frac}(Z_l + \log_2 N_{Sch,k_{l+1}}). \tag{53}$$

Visually, we can view the evolution of Z_l as a random walk on the interval $[0,1)$ with wrap-around, with step sizes $\log_2 N_{Sch,k}$ drawn from the distribution in Eq. (30). We need to

show that this walk will visit the interval $[0, \varepsilon']$ with certainty after some l steps, where $\varepsilon' = \log_2(1+\varepsilon)$. All of our step sizes are either 0 or irrational, because the base-2 logarithm of any number which is not a power of two is irrational. Let us first consider the non-random walk where each step size is given by the same irrational number x. First, it is true that the same point will never be visited twice, else x could be expressed as a rational number. Now, let us suppose that this walk never visits the interval $[0, \varepsilon']$. If that were true, then the walk could by extension never visit the intervals $[zx, zx + \varepsilon']$ for any integer z (taking the fractional part of the quantities inside this interval is implied). For each z, a set of finite length is excised from the set of possible locations that the walk can visit. (These sets will never fully overlap each other because the the walk never visits the same point twice, i.e. $zx = z'x \iff z = z'$.) Eventually, the entire unit interval is excised, meaning that the walk cannot visit *any* point, which is clearly a contradiction. The same reasoning works when the step size is chosen from a distribution of irrationals. Thus, the walk will, for some l, visit the interval $[0, \varepsilon']$ with certainty, and the perfect efficiency of the projection follows from the discussion at the end of Sec. 2.4.2.

Because the Schmidt projection and the P_{2^m} projection do not waste any entanglement E as $n \to \infty$, we have demonstrated a protocol which begins with entanglement nE_θ and produces the maximum number of maximally entangled pairs, $m = nE_\theta$. The distillable entanglement of $|\Phi_\theta\rangle$ is therefore $D = E_\theta$. From our previous results, we have now demonstrated that

$$E = E_F = D \tag{54}$$

for pure states. With regard to entanglement manipulation, the entanglement E of a pure state is the single number that tells us everything we might wish to know.

3. MIXED STATES

3.1. Entanglement measures

Finding a suitable measure of entanglement for mixed states is slightly more complicated than for pure states. The reason is simply that a mixed state can be characterized by several different ensembles $\mathcal{E}$ of pure states, each of which gives rise to exactly the same density matrix, and hence exactly the same measurement statistics. For example, the completely

mixed state of two qubits

$$\rho_{AB} = \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix} \tag{55}$$

could be prepared by mixing the unentangled states $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$ with equal probabilities (call this ensemble $\mathcal{E}_0$), or equivalently by mixing the four maximally entangled Bell states

$$|\Psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle) \tag{56}$$

$$|\Phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle) \tag{57}$$

with equal probabilities (call this the Bell ensemble $\mathcal{E}_B$). While the former ensemble clearly has zero distillable entanglement, the latter costs one full ebit to prepare, and this difference will not be captured by a simple function of the density matrix alone, as was possible with pure states $E = S(\rho_A)$. We must be wary of the difference between ensembles and mixed states, and begin with two definitions to clarify this.

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

This definition is well motivated for the following reason. Suppose Alice and Bob wish to prepare a particular ensemble $\mathcal{E}$. Using the dilution protocol discussed for pure states, they can create each state $|\Upsilon_i\rangle$ using $E(|\Upsilon_i\rangle)$ shared maximally entangled pairs (again, asymptotically). To create the full ensemble, they must create each state $|\Upsilon_i\rangle$ with a frequency p_i, so the required entanglement approaches the ensemble average of the entanglement of the pure states. (In fact, we have avoided a tricky subtlety which is exposed and discussed near the end of Sec. 3.2.2.)

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

To justify this as a useful measure of entanglement, we need to demonstrate that the quantity $E(M)$ (or rather, its expectation) cannot be increased under LOCC, in analogy with Theorem 1 for pure states. Otherwise, one mixture could be reliably transformed

into another mixture with greater entanglement, violating the principle that entanglement cannot be created locally; hence, $E(M)$ would be useless as a measure of the entanglement of formation. To prove this theorem, we must first prove the proposition introduced in Sec. 2.1 that discarding a subsystem cannot increase the entanglement.

Theorem 2: Suppose Alice and Bob share the state $|\Upsilon\rangle$ as before, but we simply imagine Alice's qubits to be subdivided into two systems A and C, which are together entangled with Bob's system B. Let Alice discard system C, so that the resulting mixed state is given by $M = \mathrm{Tr}_C|\Upsilon\rangle\langle\Upsilon|$. Then $E(M) \leq E(|\Upsilon\rangle)$; the mixed state M can always be formed from an ensemble requiring less entanglement than originally present in $|\Upsilon\rangle$.

Proof: In Theorem 1, Bob knew that Alice's measurement collapsed the state $|\Upsilon\rangle$ to $|\Upsilon_k\rangle$ with probability p_k, so his density matrix after the measurement was given by the ensemble average of the reduced density matrices $\rho_k = \mathrm{Tr}_A|\Upsilon_k\rangle\langle\Upsilon_k|$. This case is slightly different because Alice's disposal action generates a mixed state M rather than a pure state. Still, we can consider *any* ensemble $\mathcal{E} = \{p_k, |\Upsilon_k\rangle\}$ that gives rise to M as a viable ensemble, so Bob's density matrix ρ'_B after Alice's action is again given by the weighted average of the ρ_k. Again, $\rho_B = \rho'_B$ (otherwise a superluminal channel would exist), and applying the concavity of the Shannon entropy yields the same result as in Theorem 1:

$$E(|\Upsilon\rangle) \geq \sum_k p_k S(\rho_k). \tag{58}$$

Now, the right-hand side is just the entanglement of the ensemble $E(\mathcal{E})$. This result applies for any ensemble, in particular the minimum entanglement ensemble, so

$$E(|\Upsilon\rangle) \geq E(M). \tag{59}$$

QED.

Using Theorems 1 and 2, we can now justify the utility of $E(M)$ as a measure of entanglement by proving the following theorem.

Theorem 3: Let Alice subject a bipartite mixed state M to a local operation which results in mixed state M_k with probability p_k. Then the expected entanglement of the residual state is no greater than the entanglement of formation of the original state:

$$\sum_k p_k E(M_k) \leq E(M). \tag{60}$$

Proof: First, note that the theorem covers both the cases of measurement and disposal of a subsystem. In the latter situation, there will only be one value of k which occurs with unit probability, since there is no uncertainty about the resulting state M_k: it is simply the partial trace of M over the discarded subsystem.

All Alice and Bob know is that they possess a mixed state M; they do not know which particular ensemble they have. Let as assume for the moment that they are actually in possession of the minimum entanglement ensemble $\mathcal{E} = \{p_j, \Upsilon_j\}$ (ignoring the ket notation for convenience). We will return to the implications of this assumption shortly. Alice's operation will collapse the state Υ_j to Υ_{jk} with probability $p_{k|j}$. (Note: if Alice's operation is a measurement, then the states Υ_{jk} are unique. If Alice discards a subsystem, then there is some freedom in choosing the states Υ_{jk}, but this does not affect the argument we make.) If the measurement result is k, the resulting density matrix M_k is

$$M_k = \sum_j p_{j|k} |\Upsilon_{jk}\rangle\langle\Upsilon_{jk}| \tag{61}$$

(by the measurement postulate of quantum mechanics) where $p_{j|k}$ is the conditional probability of having begun with the state Υ_j, given that the measurement result is k. This equation conveniently provides one possible ensemble $\mathcal{E}' = \{p_{j|k}, \Upsilon_{jk}\}$, though not necessarily of minimal entanglement, of states realizing the mixed state M_k. For any ensemble $\mathcal{E}'$ realizing the mixed state M_k,

$$E(M_k) \leq E(\mathcal{E}') \tag{62}$$

follows directly from the definition of $E(M_k)$ as the minimal entanglement of formation over all ensembles $\mathcal{E}'$. Using the particular $\mathcal{E}'$ at hand, we find

$$E(M_k) \leq \sum_j p_{j|k} E(\Upsilon_{jk}). \tag{63}$$

We will use this result in a moment.

Applying Theorems 1 and 2 to each pure state Υ_j yields the inequality

$$\sum_k p_{k|j} E(\Upsilon_{jk}) \leq E(\Upsilon_j) \tag{64}$$

for each j, where $p_{k|j}$ is the conditional probability of measuring k, given that the initial state drawn from the ensemble was in fact Υ_j. Multiplying by p_j and summing over j gives

$$\sum_{j,k} p_j p_{k|j} E(\Upsilon_{jk}) \leq \sum_j p_j E(\Upsilon_j) = E(M), \tag{65}$$

where the equality follows because $\mathcal{E}$ was chosen to be the minimum entanglement ensemble. Using Bayes' theorem

$$p_{j,k} = p_j p_{k|j} = p_k p_{j|k} \tag{66}$$

we can rewrite Eq. (65) as

$$\sum_k p_k \left(\sum_j p_{j|k} E(\Upsilon_{jk}) \right) \leq E(M). \tag{67}$$

We can substitute for the term in parentheses using the upper bound of $E(M_k)$ from Eq. (63), which leads to the result

$$\sum_k p_k E(M_k) \leq E(M). \tag{68}$$

QED.

We set out to prove that Alice and Bob cannot increase the expected value of $E(M)$ by LOCC. Theorem 3 considers only a single operation by Alice. However, suppose Alice performs a measurement and communicates the result to Bob. Now, Alice and Bob are in possession of another generally mixed state, so they are in the same position as before. One of them can perform a measurement and relate the result to the other party, and so on. Any operation done by LOCC can be modeled in this way. Since the expected entanglement of formation of the mixed state resulting from each measurement must obey Theorem 3, we conclude that no operation done by LOCC can increase the expected value of the entanglement of formation $E(M)$ of a mixed state. A corollary of this theorem is that the entanglement of formation $E(M)$ upper bounds the distillable entanglement $D(M)$ of the mixed state M,

$$D(M) \leq E(M), \tag{69}$$

which follows by reasoning analogous to that used in the proof of the corresponding corollary for pure states (Eq. (9)).

There is one objection that can be made to the above corollary which merits discussion. So far, we have assumed a scenario in which Alice and Bob both know the density matrix M which characterizes their shared mixed state state, without knowing the particular ensemble that they share. Let us consider the situation where Alice and Bob are knowledgeable about the ensemble they share. For example, a third party ('Charlie') may prepare an equal mixture of the four Bell states and randomly choose one pair to share between Alice and

Bob, so that they possess the Bell ensemble $\mathcal{E}_B$ mentioned at the beginning of the section. The density matrix M is the completely mixed state (Eq. (55)), for which $E(M) = 0$ since the minimal entanglement ensemble $\mathcal{E}_0$ consists of completely unentangled states. That no entanglement can be distilled from the ensemble $\mathcal{E}_0$ is obvious, since entanglement cannot be created locally and there is none to begin with. However, it seems possible that Alice and Bob may be able to distill entanglement from the ensemble $\mathcal{E}_B$, since they already share one full ebit. Now we can state the objection: why should the distillable entanglement be upper bounded by $E(M)$, the entanglement of the most economical ensemble giving rise to M, rather than some function of the actual ensemble giving rise to M?

The answer is simple: a quantum state is fully characterized by its density matrix. Because the postulates of quantum mechanics can be written completely in terms of density matrices, all the results of possible measurements and operations depend only on the density matrix M, not the ensemble $\mathcal{E}$. In other words, two ensembles which give rise to the same density matrix are *completely indistinguishable*. Therefore, no entanglement can be distilled from $\mathcal{E}_B$ simply for the reason that no entanglement can be distilled from $\mathcal{E}_0$. We can be as creative as we like and dream up a distillation protocol consisting of complicated measurements and ancillas and unitary operations. No matter, the protocol will maneuver the Bell ensemble $\mathcal{E}_B$ into the same mixed state as it would the ensemble $\mathcal{E}_0$. Since we know we can never distill a maximally entangled pair from $\mathcal{E}_0$, we know we can never distill a maximally entangled pair from $\mathcal{E}_B$. Therefore, the upper bound of the distillable entanglement of some ensemble $\mathcal{E}$ is not determined by the entanglement of formation $E(\mathcal{E})$ of the ensemble itself, but rather the entanglement of formation $E(M)$ of the most economical ensemble which gives rise to the same mixed state M. Incidentally, this argument justifies the assumption made in the proof of Theorem 3 that Alice and Bob were in possession of the minimal entanglement ensemble $\mathcal{E} = \{p_j, \Upsilon_j\}$. All that matters is M, so we could have chosen any ensemble. The choice of the minimal entanglement ensemble happened to be convenient in proving the desired result.

3.2. Entanglement of formation of mixed states

3.2.1. Entanglement of formation of pure states, revisited

We already know the entanglement of formation of a pure state is given by the von Neumann entropy of its reduced density matrix. We reapproach the problem here in a way that will provide us with significant insight into the problem of finding the entanglement of formation for mixed states.

Consider an arbitrary pure state $|\Upsilon\rangle$ of a system of two spin-1/2 particles. We have previously spent our time analyzing such a state expressed in the Schmidt basis. Now, we consider a basis $\{|e_j\rangle\}$ closely related to the Bell basis,

$$\begin{aligned} |e_1\rangle &= |\Phi^+\rangle \\ |e_2\rangle &= i|\Phi^-\rangle \\ |e_3\rangle &= i|\Psi^+\rangle \\ |e_4\rangle &= |\Psi^-\rangle, \end{aligned} \tag{70}$$

so that

$$|\Upsilon\rangle = \sum_{j=1}^{4} \alpha_j |e_j\rangle. \tag{71}$$

Let us calculate the entanglement of formation $E(|\Upsilon\rangle) = S(\rho_A)$, where $\rho_A = \mathrm{Tr}_B |\Upsilon\rangle\langle\Upsilon|$. The calculation is straightforward but algebraically tedious, so we will take it in steps. By expanding $\rho = |\Upsilon\rangle\langle\Upsilon|$ and taking the partial trace over system B, we find

$$\rho_A = \begin{pmatrix} a-b & -c+id \\ -c-id & a+b \end{pmatrix} \tag{72}$$

where

$$\begin{aligned} a &= 1/2 \\ b &= \mathrm{Im}(\alpha_2\alpha_1^*) + \mathrm{Im}(\alpha_3\alpha_4^*) \\ c &= \mathrm{Im}(\alpha_3\alpha_1^*) + \mathrm{Im}(\alpha_4\alpha_2^*) \\ d &= \mathrm{Im}(\alpha_4\alpha_1^*) + \mathrm{Im}(\alpha_2\alpha_3^*). \end{aligned} \tag{73}$$

The eigenvalues of ρ_A are given by

$$\lambda_\pm = a \pm \sqrt{b^2 + c^2 + d^2}. \tag{74}$$

It remains to solve for $b^2+c^2+d^2$, which, by judiciously adding and subtracting the quantity $(|\alpha_1|^4 + |\alpha_2|^4 + |\alpha_3|^4 + |\alpha_4|^4)$, can be expressed as

$$b^2 + c^2 + d^2 = \frac{1}{4}\left(1 - \left|\sum_j \alpha_j^2\right|^2\right). \tag{75}$$

The von Neumann entropy of ρ_A is equal to the Shannon entropy of its eigenvalues, so the entanglement of $|\Upsilon\rangle$ is given by

$$E(|\Upsilon\rangle) = H\left(\frac{1+\sqrt{1-C^2}}{2}\right), \tag{76}$$

where H is the binary entropy function and

$$C \equiv \left|\sum_j \alpha_j^2\right|. \tag{77}$$

(Caution: in the definition of C, the complex numbers α_j are squared, not their moduli.)

Equation (76) is a formula for the entanglement of a pure state in terms of its coefficients α_j. However, we previously had a much simpler formula for the same quantity in terms of the Schmidt coefficients, $E = H(\cos^2\theta)$. Our true gain will come from analyzing the quantity C. Since C and E range from 0 to 1, and both are monotonically increasing functions of each other, C is itself a useful measure of entanglement. Thinking of complex numbers as vectors in the complex plane, note that C is just the magnitude of the sum of the four complex vectors α_j^2. The normalization constraint $\sum_j |\alpha_j|^2 = 1$ requires that the lengths of the vectors α_j^2 sum to one. Figure 4 shows how different sets of coefficients $\{\alpha_j\}$ give rise to different values for C by visually summing the complex numbers $\{\alpha_j^2\}$. For example, the values $\{\alpha_j^2\}$ giving rise to Fig. 4(a) yield $C < 1$, so these values correspond to some partially entangled state.

Claim: Every maximally entangled state can be written, up to an overall phase factor, as a real linear combination of the $|e_j\rangle$'s. In other words, all the coefficients α_j are real, once a common phase has been factored out. (Had we used the Bell basis rather than the $\{|e_j\rangle\}$ basis, this claim would not be possible.)

Proof: We write the complex coefficients α_j in polar form as $\alpha_j = |\alpha_j|\exp(i\phi_j)$, so that the vectors α_j^2 are given by $\alpha_j^2 = |\alpha_j|^2\exp(i2\phi_j)$. A maximally entangled state is characterized by $E = 1$, hence $C = 1$. Now, $C = 1$ is achieved if and only if the vectors α_j^2 all point along the same ray as in Fig. 4(b); that is, the complex phases $2\phi_j$ must all be equal,

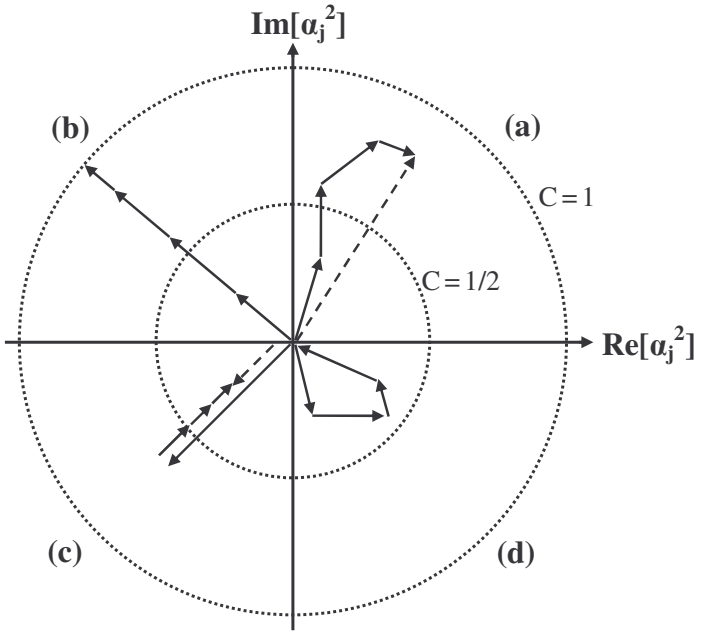


FIG. 4: Four sets of values for α_j^2 and the corresponding C quantity, organized by quadrant for convenience. The complex numbers α_j^2 are represented by solid vectors. Where needed, the dashed vectors represent $\sum_j \alpha_j^2$; C is given by the length of these vectors. (a) An arbitrary partially entangled state. (b) A maximally entangled state requires that the α_j^2 all have the same phase. (c) If $|\alpha_j|^2$ exceeds $\frac{1}{2}$ for one of the α_j, the state is guaranteed to be partially entangled. (d) If $|\alpha_j|^2$ is less than $\frac{1}{2}$ for all of the α_j, it is always possible to choose the phases ϕ_j such that $C = 0$.

modulus 2π. Equivalently, we can say that the phases ϕ_j are all equal, modulus π. We define $\phi = \phi_j \bmod(\pi)$. If we factor the common phase $\exp(i\phi)$ out of each coefficient α_j, the remaining factor is either $|\alpha_j|$ or $-|\alpha_j|$, real in either case. QED.

Another important scenario is demonstrated in Fig. 4(c). Suppose that $|\alpha_1|^2 > \frac{1}{2}$. By the normalization condition, $|\alpha_2|^2 + |\alpha_3|^2 + |\alpha_4|^2 = 1 - |\alpha_1|^2 < \frac{1}{2}$, so there is no way to force C to zero. To minimize C, we must choose α_2^2, α_3^2, and α_4^2 to perfectly anti-align with the vector α_1^2, which yields the lower bound

$$C \geq |\alpha_1|^2 - (1 - |\alpha_1|^2) = 2|\alpha_1|^2 - 1. \tag{78}$$

Substituting into Eq. (76) yields a lower bound on the entanglement

$$E(|\Upsilon\rangle) \geq H\left(\tfrac{1}{2} + \sqrt{|\alpha_1|^2(1 - |\alpha_1|^2)}\right) \tag{79}$$

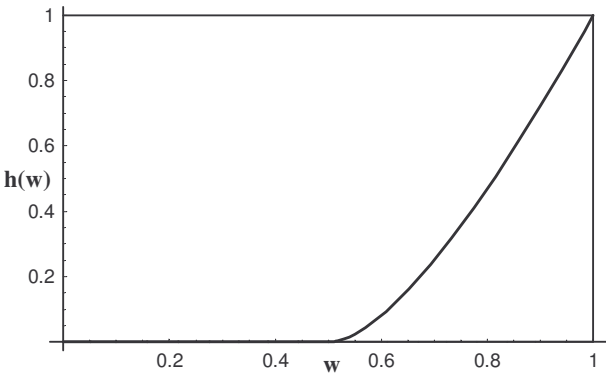


FIG. 5: A plot of the piecewise function $h(w)$, which upper bounds the entanglement of formation of a pure state $|\Upsilon\rangle$ given $w = |\langle e|\Upsilon\rangle|^2$.

for the case $|\alpha_1|^2 > \frac{1}{2}$. Of course, our decision to use the component α_1 was arbitrary; the result is stronger if we allow ourselves to use the largest component α_j when evaluating the bound. Indeed, the bound can be made stronger yet if we realize that the basis $\{|e_j\rangle\}$ was chosen rather arbitrarily. Any basis $\{|e'_j\rangle\}$ which upholds the property that every maximally entangled state can be written as a real linear combination of the basis vectors (up to a universal phase) would work just as well. (Formally, $|e'_j\rangle = \sum_k R_{jk}|e_k\rangle$ for any real, orthogonal matrix R.) Thus, we can substitute for α_1 the component of $|\Upsilon\rangle$ along any $|e'_j\rangle$. It is easy to see that by a suitable choice of R, the basis vector $|e'_j\rangle$ can be made into *any* maximally entangled state $|e\rangle$. Therefore, let us define the quantity $w = |\langle e|\Upsilon\rangle|^2$. We also introduce the function

$$h(x) = \begin{cases} H(\frac{1}{2} + \sqrt{x(1-x)}) & \text{for } x \geq \frac{1}{2}, \\ 0 & \text{for } x < \frac{1}{2}. \end{cases} \tag{80}$$

The bound (79) can be restated as

$$E(|\Upsilon\rangle) \geq h(w). \tag{81}$$

To achieve the strongest bound, w should be maximized over all $|e\rangle$. The function $h(x)$ is plotted in Fig. 5. Note that it is convex.

3.2.2. Lower bound on the entanglement of formation of mixed states

The lower bound in Eq. (81) for $E(|\Upsilon\rangle)$ can be used to easily prove a lower bound on the entanglement of formation $E(M)$ for mixed states. This is a powerful result, because proving anything for mixed states is usually difficult.

Take any mixed state M and consider an arbitrary ensemble $\mathcal{E} = \{p_k, |\Upsilon_k\rangle\}$ which gives rise to M. For each pure state $|\Upsilon_k\rangle$, we define $w_k = |\langle e|\Upsilon_k\rangle|^2$. (Note that $|e\rangle$ is chosen to be the same for all k, for reasons that will become clear.) The entanglement of the ensemble $\mathcal{E}$ is bounded below by

$$E(\mathcal{E}) = \sum_k p_k E(|\Upsilon_k\rangle) \geq \sum_k p_k h(w_k) \geq h\left(\sum_k p_k w_k\right), \tag{82}$$

where the first inequality follows from Eq. (81) and the second from the convexity of $h(x)$. Now, this bound is truly only useful if it can be expressed in terms of the mixed state M rather than the particulars of the ensemble $\mathcal{E}$. Fortunately, the argument of h

$$\sum_k p_k w_k = \sum_k p_k \langle e|\Upsilon_k\rangle\langle\Upsilon_k|e\rangle = \langle e|\left(\sum_k p_k |\Upsilon_k\rangle\langle\Upsilon_k|\right)|e\rangle = \langle e|M|e\rangle \tag{83}$$

is a function of M. (Had we allowed different choices of the entangled state $|e_k\rangle$ for each k, we could not have factored $\langle e|$ and $|e\rangle$ out of the sum to reach the second equality in Eq. (83).) We obtain the strongest bound by maximizing $\langle e|M|e\rangle$ over all entangled states $|e\rangle$. This quantity is so useful that we define the 'fully entangled fraction' of a mixed state M by

$$f(M) \equiv \max\langle e|M|e\rangle, \tag{84}$$

where the maximization is taken over all fully entangled states $|e\rangle$. Since Eq. (82) holds for all ensembles $\mathcal{E}$, in particular it holds for the minimal entanglement ensemble. Thus, we obtain a lower bound on the entanglement of formation $E(M)$ for any mixed state:

$$E(M) \geq h[f(M)]. \tag{85}$$

(Caveat: This lower bound is not entirely true in all situations, and some remarks on the subtleties involved in the definitions of $E(\mathcal{E})$ and $E(M)$ are required to understand why. Recall that whenever we talk about the entanglement of formation of a pure state $|\Upsilon_k\rangle$, we mean it in the asymptotic sense, in which we create the state $|\Upsilon_k\rangle^{\otimes n}$ using m maximally

entangled pairs, and say that the entanglement of formation is given by m/n in the limit $n \to \infty$. Now consider the preparation of an ensemble $\mathcal{E} = \{p_k, |\Upsilon_k\rangle\}$. How do we create such an ensemble asymptotically? Let us start by saying we are going to create a set of n pairs of qubits. Of these n pairs, $p_k n$ of them will be in the state $|\Upsilon_k\rangle$. Alice therefore prepares the state $|\Upsilon_k\rangle^{\otimes p_k n}$ in her laboratory, and teleports the appropriate half of the qubits to Bob using $p_k n E(|\Phi_k\rangle)$ maximally entangled pairs. Once she has done this for each k, she will have used up a number of maximally entangled pairs m given by

$$m = n \sum_k p_k E(|\Upsilon_k\rangle). \tag{86}$$

Now, what can we say about the state that Alice and Bob share? Suppose that Alice and Bob somehow forget, in a sudden amnesic episode, to which particular state $|\Upsilon_k\rangle$ each of their n qubits belongs. Then, all they know is that they share n pairs, where each pair could be in the state $|\Upsilon_k\rangle$ with probability p_k. In other words, they have created the state $\rho^{\otimes n}$, where ρ is the density matrix corresponding to the mixed state M that we are interested in. Since they have created n copies of ρ using m maximally entangled pairs, with m given by Eq. (86), we declare that the entanglement of formation of the ensemble is given by

$$E(\mathcal{E}) = \sum_k p_k E(|\Upsilon_k\rangle). \tag{87}$$

Of course, this is just the definition that we used for $E(\mathcal{E})$, and the above argument is the motivation for that definition. There is nothing wrong with defining $E(\mathcal{E})$ in the way that we have. But we must observe that the definition was 'derived,' in a sense, from the particular dilution protocol described above: Alice and Bob prepare the the state $\rho^{\otimes n}$ by preparing each state $|\Upsilon_k\rangle$ in the ensemble pair-by-pair. Still, it is just a definition, and we really have not ascribed too much physical significance to the quantity $E(\mathcal{E})$.

However, we have ascribed a significant amount of physical significance to $E(M)$, which is supposed to be the minimal amount of entanglement necessary to asymptotically create the mixed state M. The flaw in our reasoning is now simple to pinpoint: we defined $E(M)$ as the minimum value of $E(\mathcal{E})$ over all ensembles $\mathcal{E}$ which realize the mixed state M. The problem is that this assumes we are forced to use the pair-by-pair protocol for creating $\rho^{\otimes n}$. One could perhaps imagine a more efficient protocol which created the state $\rho^{\otimes n}$ all at once rather than pair by pair. In that case, it was conjectured and subsequently proved [4], though the proof is beyond the scope of this paper, that the entanglement of formation is

given by

$$E(\rho) = \lim_{n\to\infty} \frac{E(\rho^{\otimes n})}{n}. \tag{88}$$

Still, there is no easy way to calculate $E(\rho^{\otimes n})$, and there remained the nagging suspicion that the pair-by-pair protocol could be the most efficient protocol for creating any state $\rho^{\otimes n}$; in other words, that E was additive over tensor products. This remained an open problem until 2008, when Hastings' counterexample to the additivity conjecture of the Holevo capacity for quantum channels simultaneously disproved the additivity conjecture of the entanglement of formation over tensor products, because the two problems were directly related. Thus, there exist some states ρ for which the pair-by-pair protocol consumes more than the minimum amount of entanglement necessary to create $\rho^{\otimes n}$.

Having raised this objection, we will go back to treating $E(M)$, defined as the minimization over all $E(\mathcal{E})$, as the true entanglement of formation of the mixed state M, but the reader is urged to always remember this caveat. In any case, $E(M)$ still serves as an upper bound on the distillable entanglement $D(M)$; if a more efficient dilution protocol were discovered, it would serve only to lower $E(M)$.)

We have spent much of our time complaining about the difficulty of calculating the entanglement of formation of a mixed state, because it requires a difficult minimization over all viable ensembles. Now, at least we have proven a lower bound on the entanglement of formation, but the bound is useful only if we can find the fully entangled fraction $f(m)$, which requires a maximization over all fully entangled states. Luckily, there is simple procedure to accomplish this.

Claim: The fully entangled fraction $f(M) = \max\langle|e|M|e\rangle$ of a mixed state M is given by the largest eigenvalue of the matrix $\mathrm{Re}(M)$, when M is expressed in the $\{|e_j\rangle\}$ basis of Eq. (70) (or indeed, any basis which is related to $\{|e_j\rangle\}$ by an orthogonal transformation).

Proof: When expressed in the $\{|e_j\rangle\}$ basis, a pure state is maximally entangled if and only if each of its vector components is real, as proven before. So the maximization over all fully entangled states $|e\rangle$ can be restated as the maximization over all real vectors $|e\rangle$. We split M into the sum of a real matrix and a purely imaginary matrix $M = \mathrm{Re}(M) + \mathrm{Im}(M)$. M is Hermitian by virtue of being a density matrix, and it is easy to see that $\mathrm{Re}(M)$ and $\mathrm{Im}(M)$ must also be Hermitian. The quantity of interest is given by

$$\langle e|M|e\rangle = \langle e|\mathrm{Re}(M)|e\rangle + \langle e|\mathrm{Im}(M)|e\rangle. \tag{89}$$

The second term on the right is equal to zero for any real $|e\rangle$ and purely imaginary, Hermitian matrix $\mathrm{Im}(M)$; one way to check this for a 4×4 matrix is simply to do the matrix multiplication with arbitrary components. Turning our attention to the first term, note that $\mathrm{Re}(M)$ will have strictly real eigenvectors, from which we form the orthonormal basis $\{|e'_j\rangle\}$, and associated real eigenvalues m_j. We can express any $|e\rangle$ in this basis as $|e\rangle = \sum_j \alpha_j |e'_j\rangle$ using only real coefficients α_j. Then,

$$\langle e|\mathrm{Re}(M)|e\rangle = \langle e| \sum_j m_j \alpha_j |e'_j\rangle = \sum_j \alpha_j^2 m_j, \tag{90}$$

which is just a weighted mean of the eigenvalues m_j, since $\sum_j \alpha_j^2 = 1$ by the normalization condition. Clearly, the maximum value of this weighted average, over all possible choices of $\{\alpha_j\}$, will just be the largest eigenvalue m_j. QED.

The bound of Eq. (85) is actually achieved, meaning $E(M) = h[f(M)]$, in cases where M is a pure state or a Bell-diagonal mixture. We demonstrate this fact for pure states here, and delay the second result until the following section.

Through local unitary operations, any pure state $|\Upsilon\rangle$ can be transformed into $|\Phi_\theta\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle$ (provided any phases are absorbed into the definitions of the states $|0\rangle$ and $|1\rangle$). Since such operations do not change the entanglement of the state, we only need to show that the bound is achieved for $|\Phi_\theta\rangle$. For this state, $M = |\Phi_\theta\rangle\langle\Phi_\theta|$, it is self-evident that f is achieved by choosing $|e\rangle = |\Phi^+\rangle$, which results in $f(M) = |\langle\Phi^+|\Phi_\theta\rangle|^2 = \frac{1}{2} + \sin\theta\cos\theta$. By plugging into Eq. (80), one finds that $h(\frac{1}{2} + \sin\theta\cos\theta) = H(\cos^2\theta)$, which we know is equal to the entanglement $E(|\Phi_\theta\rangle)$. Therefore, the bound is achieved for all pure states.

3.2.3. Entanglement of formation for mixtures of Bell states

The lower bound on the entanglement of formation given by Eq. (85) offers a means of searching for the true entanglement of formation. The strategy is as follows. Given a mixed state M, we can choose any ensemble $\mathcal{E}$ which realizes the mixed state and calculate $E(\mathcal{E})$. If $E(\mathcal{E})$ happens to equal $h[f(M)]$, then the lower bound assures us that $\mathcal{E}$ must be a minimal entanglement ensemble for the mixed state M. In that case, we can confidently declare $E(M) = h[f(M)]$. Granted, a little luck is involved in guessing the correct ensemble $\mathcal{E}$. It turns out that we can do this for Bell-diagonal mixtures.

Consider the mixed state

$$W = \sum_{j=1}^{4} p_j |e_j\rangle\langle e_j|. \tag{91}$$

(Note: the exact same mixed state is achieved with a similar ensemble consisting of Bell states rather than $|e_j\rangle$'s, so there is no problem in referring to this as a Bell-diagonal mixture. It is convenient to continue to use the $\{|e_j\rangle\}$ basis for calculations, but the results apply to mixtures of Bell states, or more generally mixtures of any four orthonormal maximally entangled states.) This state W can be generated by an equal mixture of the following eight states:

$$\sqrt{p_1}e^{i\theta_1/2}|e_1\rangle \pm \sqrt{p_2}e^{i\theta_2/2}|e_2\rangle \pm \sqrt{p_3}e^{i\theta_3/2}|e_3\rangle \pm \sqrt{p_4}e^{i\theta_4/2}|e_4\rangle, \tag{92}$$

for arbitrary θ_j, as can be checked by direct calculation. This ensemble, call it $\mathcal{E}$, has the nice property that

$$C = \left| \sum_j p_j e^{i\theta_j} \right| \tag{93}$$

takes the same value for each of the eight states. In other words, each state has the same entanglement of formation. The entanglement of formation for the ensemble, given by the ensemble average of the entanglements of the pure states, is therefore equal to the entanglement of formation of the pure states:

$$E(\mathcal{E}) = H[\tfrac{1}{2}(1 + \sqrt{1 - C^2})], \tag{94}$$

with C given by Eq. (93). We consider two separate cases.

Case 1: None of the p_j are greater than $\frac{1}{2}$. Then the fully entangled fraction $f < \frac{1}{2}$, since f is given by the largest of the p_j when the mixture is Bell-diagonal. Hence, the bound tells us that $E(W) \geq h[f(W)] = 0$ (see Eq. (80)). Since none of the p_j are greater than $\frac{1}{2}$, it is always possible to choose the θ_j such that $\sum_j p_j e^{i\theta_j} = 0$. (This is easiest to see visually, as in Fig. 4(d); given four line segments of fixed length, it is always possible to arrange the segments in a quadrilateral, provided that the sum of the lengths of any three segments is always greater than the length of the fourth.) By choosing these particular θ_j, we see from Eq. (93) that $C = 0$. By Eq. (94), $E(\mathcal{E}) = H(1) = 0$. Therefore, $E(W) = 0 = h[f(W)]$; the bound is achieved.

Case 2: One of the p_j is greater than $\frac{1}{2}$; choose p_1 without loss of generality. Then $f(W) = p_1$, and the bound tells us that $E(W) \geq h(p_1)$. We can no longer choose the θ_j as

in Case 1; instead, we choose $\theta_1 = 0$ and $\theta_2 = \theta_3 = \theta_4 = \pi$ so that our ensemble consists of the eight states

$$\sqrt{p_1}|e_1\rangle + i(\pm\sqrt{p_2}|e_2\rangle \pm \sqrt{p_3}|e_3\rangle \pm \sqrt{p_4}|e_4\rangle). \tag{95}$$

Now,

$$C = p_1 - p_2 - p_3 - p_4 = 2p_1 - 1 \tag{96}$$

which leads to

$$E(\mathcal{E}) = H[\tfrac{1}{2} + \sqrt{p_1(1-p_1)}] = h(p_1) = h[f(W)], \tag{97}$$

so we have again succeeded in achieving the bound. Taken together, these two cases imply that for any Bell-diagonal mixture W,

$$E(W) = h[f(W)]. \tag{98}$$

We have shown that the bound in Eq. (85) is achieved for both pure states and Bell-diagonal mixed states. It is not necessarily achieved for a general mixed state M. Consider the state

$$M = \tfrac{1}{2}|00\rangle\langle 00| + \tfrac{1}{2}|\Psi^+\rangle\langle\Psi^+|. \tag{99}$$

It is easy to check that $f(M) = \frac{1}{2}$, so the bound tells us only that $E(M) \geq 0$. We will show in Sec. 3.3.4 that $D(M) > 0$: some entanglement can be reliably distilled from M. Since $E(M)$ always upper bounds $D(M)$, we obtain $E(M) > 0$. Hence, the bound is not always achieved.

We conclude this section with a concrete example which illustrates the sometimes non-intuitive nature of the entanglement of formation. Consider the Werner states W_F defined by

$$W_F = F|\Psi^-\rangle\langle\Psi^-| + \frac{1-F}{3}\left(|\Psi^+\rangle\langle\Psi^+| + |\Phi^+\rangle\langle\Phi^+| + |\Phi^-\rangle\langle\Phi^-|\right) \tag{100}$$

which constitute a subspace of all Bell-diagonal mixed states. The Werner states depend only on a single parameter F, the proportion of the singlet state in the ensemble, and the remainder of the ensemble is divided up evenly among the three triplet states. There are several ways to view the preparation of this mixed state. We could say that it is drawn from an ensemble which consists of F parts singlet and $(1-F)/3$ of each of the triplets; this ensemble would cost one full ebit to prepare. A more economical ensemble can be found by exploiting the similarity between the subspace spanned by the triplet states and

the completely mixed state: we could draw from from an ensemble of $x = (4F-1)/3$ parts pure singlet and $1-x$ parts the totally mixed state. This would only cost $(4F-1)/3$ ebits to prepare. For example, consider the state $W_{5/8}$. (This state is frequently analyzed when discussing information transmission tasks, because it results from sending one qubit of a pure singlet pair $|\Psi^-\rangle$ through a 50% depolarizing channel, i.e. a channel which faithfully transmits a qubit's state with probability $\frac{1}{2}$ and replaces the qubit with the totally mixed state the other half of the time.) By directly implementing the second ensemble, the $W_{5/8}$ state can be prepared using 0.5 ebits. For this state, however, $f(W_F) = F = 5/8$, so Eq. (98) assures us that $E(W_{5/8}) = h(5/8) = 0.117$ ebits, which can be achieved using the ensemble given in Eq. (95).

3.3. Entanglement distillation for mixed states

We now concern ourselves with the prospect of distilling entanglement from Bell-diagonal mixed states. (At this point, we will leave the $\{|e_j\rangle\}$ basis behind and stick with the more familiar Bell basis of Eqs. (56) and (57).) This is not quite so restrictive as it may seem, since there is a local operation (the 'twirl,' discussed in Sec. 3.3.1) which converts any mixed state M into a Bell-diagonal Werner state W_F. The protocols will rely on either one or two-way classical communication between Alice and Bob, and we define the distillation yield from these procedures by D_1 and D_2, respectively. Because one-way communication is just a subset of two-way communication (i.e. Alice can choose to ignore the messages from Bob),

$$D_1(M) \leq D_2(M) \leq E(M), \tag{101}$$

where the second inequality follows for the usual reasons. Since we are narrowing our focus to Bell-diagonal mixtures, we can use the expression $E(M) = h[f(M)]$ from the previous section as an upper bound on the distillable entanglement.

3.3.1. Basics

The distillation protocols will be much easier to understand if we spend some time discussing the types of quantum operations that Alice and Bob will employ, along with some useful notation.

(1) By restricting our focus to Bell-diagonal mixed states, we can imagine our ensemble to consist purely of different proportions of Bell pairs (by the previous argument that two ensembles which generate the same mixed state are completely indistinguishable). Therefore, the most useful operations to employ in a given protocol are the ones which map Bell states onto each other, because it will allow us to think in terms of Bell states throughout the entire protocol. Only a small subset of all possible quantum operations on a two-qubit system satisfy this criterion. These operations can be either *bilateral*, meaning both Alice and Bob operate on their half of the Bell pair to generate another Bell state, or *unilateral*, meaning only one of them performs an operation. We will demonstrate that the relevant single-pair operations are: (1) unilateral rotations by π radians about each spatial axis, corresponding to the Pauli matrices σ_x, σ_y, and σ_z, and (2) bilateral rotations by $\pi/2$ radians, which we will denote B_x, B_y, and B_z. We also need to consider the possibility of an operation acting over multiple pairs, and we will find that (3) the bilateral two-qubit XOR operation (BXOR) maps Bell states onto each other.

Consider the single-particle (passive) SU(2) rotation by angle ϕ about the axis $\hat{n}$,

$$\mathcal{D}(\hat{n},\phi) = \begin{pmatrix} \cos\frac{\phi}{2} + in_z \sin\frac{\phi}{2} & (in_x + n_y)\sin\frac{\phi}{2} \\ (in_x - n_y)\sin\frac{\phi}{2} & \cos\frac{\phi}{2} - in_z\sin\frac{\phi}{2} \end{pmatrix}. \tag{102}$$

By choosing $\phi = \pi$, the rotations about each of the three axes are given by the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{103}$$

up to a universal phase factor. This phase factor rarely concerns us, because none of the measurements we perform in our distillation protocols are affected by the relative phases between the pairs of the ensemble. For practical purposes, therefore, it is fine to use the Pauli matrices as representations of π rotations in SU(2). To see that the unilateral application of a Pauli matrix maps Bell states onto each other, we consider one example:

$$\begin{aligned} (I \otimes \sigma_y)|\Phi^+\rangle &= (I \otimes \sigma_y)\tfrac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle) \\ &= \tfrac{1}{\sqrt{2}}(|0\rangle \otimes i|1\rangle + |1\rangle \otimes (-i)|0\rangle) \\ &= i|\Psi^-\rangle \\ &\simeq |\Psi^-\rangle, \end{aligned}$$

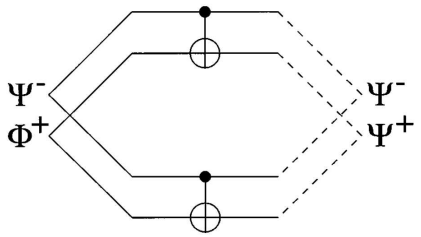


FIG. 6: Reprinted from [2]. The BXOR operation. Given a source pair and a target pair, Alice and Bob both subject their source qubit (solid dot) and target qubit (crossed circle) to a CNOT gate. In this example, the source $|\Psi^-\rangle$ and target $|\Phi^+\rangle$ are mapped to $|\Psi^-\rangle$ and $|\Psi^+\rangle$, respectively.

where the last line again follows because the universal phase is irrelevant for our purposes. Thus, the unilateral σ_y takes $|\Phi^+\rangle \rightarrow |\Psi^-\rangle$ (no matter which party performs rotation). The results for all possible unilateral π rotations are presented at the top of Table I.

Next, we consider the rotations about each axis by $\pi/2$ radians:

$$\mathcal{D}_{\hat{x},\frac{\pi}{2}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}, \mathcal{D}_{\hat{y},\frac{\pi}{2}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \mathcal{D}_{\hat{z},\frac{\pi}{2}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix}. \tag{104}$$

These rotations only map Bell states to Bell states when applied bilaterally. For example,

$$\begin{aligned} B_z|\Phi^+\rangle &= (\mathcal{D}_{\hat{z},\frac{\pi}{2}} \otimes \mathcal{D}_{\hat{z},\frac{\pi}{2}}) \tfrac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle) \\ &= \tfrac{1}{2\sqrt{2}}[(1+i)|0\rangle \otimes (1+i)|0\rangle + (1-i)|1\rangle \otimes (1-i)|1\rangle] \\ &= i|\Phi^-\rangle \\ &\simeq |\Phi^-\rangle \end{aligned}$$

The results for all possible bilateral $\pi/2$ rotations are presented in the middle of Table I. Note that the singlet state $|\Psi^-\rangle$ is curiously invariant under B_x, B_y, and B_z; this feature will make the 'twirl' operation possible.

In the above operations, Alice and Bob each only operate on one qubit. By contrast, the BXOR operation takes a source (control) pair and a target pair and applies a controlled-NOT

Unilateral π Rotations:

	source				
	Ψ^-	Φ^-	Φ^+	Ψ^+	
I	Ψ^-	Φ^-	Φ^+	Ψ^+	
σ_x	Φ^-	Ψ^-	Ψ^+	Φ^+	
σ_y	Φ^+	Ψ^+	Ψ^-	Φ^-	
σ_z	Ψ^+	Φ^+	Φ^-	Ψ^-	

Bilateral $\pi/2$ Rotations:

	source				
	Ψ^-	Φ^-	Φ^+	Ψ^+	
I	Ψ^-	Φ^-	Φ^+	Ψ^+	
B_x	Ψ^-	Φ^-	Ψ^+	Φ^+	
B_y	Ψ^-	Ψ^+	Φ^+	Φ^-	
B_z	Ψ^-	Φ^+	Φ^-	Ψ^+	

Bilateral XOR:

	source				
target	Ψ^-	Φ^-	Φ^+	Ψ^+	
	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)

TABLE I: Reproduced from [2]. The three types of operations which map Bell states to Bell states, and are therefore useful in distillation protocols for Bell-diagonal mixtures. The BXOR operation is the only one which acts on two pairs simultaneously; accordingly, each entry of the BXOR table identifies what happens to the source state (top line) and what happens to the target state (bottom line).

gate at both Alice's and Bob's end (Fig. 6). For example,

$$\begin{aligned}
\text{BXOR}\left[|\Psi^-\rangle \otimes |\Phi^+\rangle\right] &= \text{BXOR}\left[\tfrac{1}{2}\left((|01\rangle_{AB} - |10\rangle_{AB}) \otimes (|00\rangle_{AB} + |11\rangle_{AB})\right)\right] \\
&= \text{BXOR}\left[\tfrac{1}{2}(|00\rangle_A \otimes |10\rangle_B + |01\rangle_A \otimes |11\rangle_B - |10\rangle_A \otimes |00\rangle_B - |11\rangle_A \otimes |01\rangle_B)\right] \\
&= \tfrac{1}{2}(|00\rangle_A \otimes |11\rangle_B + |01\rangle_A \otimes |10\rangle_B - |11\rangle_A \otimes |00\rangle_B - |10\rangle_A \otimes |01\rangle_B) \\
&= \tfrac{1}{2}\left((|01\rangle_{AB} - |10\rangle_{AB}) \otimes (|01\rangle_{AB} + |10\rangle_{AB})\right) \\
&= |\Psi^-\rangle \otimes |\Psi^+\rangle.
\end{aligned}$$

The results for all possible BXOR operations are presented at the bottom of Table I.

(2) As mentioned before, any mixed state M can be transformed into a Werner state W_F by an irreversible operation T known as the 'twirl.' In particular, the action of T on a general mixed state density matrix M (Hermitian, Trace=1) expressed in the basis $\{|\Psi^-\rangle, |\Phi^-\rangle, |\Phi^+\rangle, |\Psi^+\rangle\}$ is given by

$$T\begin{pmatrix} M_{00} & M_{01} & M_{02} & M_{03} \\ M_{10} & M_{11} & M_{12} & M_{13} \\ M_{20} & M_{21} & M_{22} & M_{23} \\ M_{30} & M_{31} & M_{32} & M_{33} \end{pmatrix} = \begin{pmatrix} M_{00} & 0 & 0 & 0 \\ 0 & (1-M_{00})/3 & 0 & 0 \\ 0 & 0 & (1-M_{00})/3 & 0 \\ 0 & 0 & 0 & (1-M_{00})/3 \end{pmatrix}. \quad (105)$$

Recall from Table I that $|\Psi^-\rangle$ is invariant under the bilateral rotations B_x, B_y, and B_z. In fact the result is more general: the singlet state is invariant under *any* bilateral rotation, by any angle about any axis. Therefore, twirling can be achieved by randomly selecting an independent SU(2) rotation for each impure pair and–without knowing which rotation has been selected–applying it to both members of the pair. This operation will preserve the matrix element $M_{00} = \langle\Psi^-|M|\Psi^-\rangle$ and randomize the rest of the mixed state, thereby achieving Eq. (105). In fact, it is not necessary to randomly select from a continuum of SU(2) rotations; a discrete twirl achieves the same effect when the rotation is randomly

chosen from the following set of 12 operations:

$$\{U_i\} = \begin{array}{c} I \\ B_xB_x \\ B_yB_y \\ B_zB_z \\ B_xB_y \\ B_yB_z \\ B_zB_x \\ B_yB_x \\ B_xB_yB_xB_y \\ B_yB_zB_yB_z \\ B_zB_xB_zB_x \\ B_yB_xB_yB_x. \end{array} \tag{106}$$

(These are the 12 symmetry operations of the tetrahedron, but the analogy will not be discussed further.) It is easy to check that the discrete twirl achieves the Werner state,

$$W_F = \frac{1}{12}\sum_{i=1}^{12} U_i^\dagger M U_i, \tag{107}$$

with $F = M_{00}$. Caution: Table I cannot be used to deduce the matrix forms of B_x, B_y, and B_z because the discarded phase information does play a role in this calculation. The phase information is included in Table II. For example,

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

TABLE II: Reproduced from [2]. The effect of bilateral $\pi/2$ rotations on Bell states, including the phase changes. The phases are only important when computing the effect of the discrete twirl in Eq. (107).

$$B_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & i & 0 \end{pmatrix} \tag{108}$$

can be read off of the second line in Table II.

Twirling is advantageous because it transforms any mixed state into Bell-diagonal form, after which it can be more easily dealt with as an ensemble of Bell pairs. (Note: for our purposes, there is no added benefit coming from the equalization of the triplet eigenvalues in the Werner state.) The disadvantage is that some of the distillable entanglement could easily have been lost in the process, because twirling can never increase the fully entangled fraction $f(M)$ of a mixed state. To see this, note that after the twirl the fully entangled fraction of the state is given by

$$f(W_F) = \langle \Psi^- | W_F | \Psi^- \rangle = F \tag{109}$$

provided $F \geq \frac{1}{4}$. Therefore, the entanglement of formation is $E(W_F) = h(F)$. (If $F < \frac{1}{4}$, then $f = (1-F)/3 < \frac{1}{2}$, so $E(W_F) = 0$ by Eq. (98), and therefore $D(W_F) = 0$, so the twirl has certainly not increased the amount of distillable entanglement.) Before the twirl,

$$f(M) \geq \langle \Psi^- | M | \Psi^- \rangle = F, \tag{110}$$

where the inequality is due to the nature of f as the maximum over all fully entangled states (and $|\Psi^-\rangle$ is just a particular one), and the equality follows because twirling conserves the M_{00} matrix element. Because $h(f)$ is a monotonically increasing function of f, the entanglement of formation of the mixed state can only be greater than the entanglement of formation of the resulting Werner state,

$$E(M) \geq h(F) = E(W_F). \tag{111}$$

This implies that the twirl decreases the upper bound on the distillable entanglement. Granted, this is not sufficient to prove that the distillable entanglement $D(W_F)$ is no greater that $D(M)$, since it is only a statement about bounds. But clearly the the twirl only introduces uncertainty into the system. In other words, Alice and Bob know less information about the state W_F than they did about the original state M, due to their newfound lack

of knowledge about which SU(2) rotation was chosen. There is absolutely no reason why Alice and Bob should be able to distill *more* entanglement from a state that they know *less* about.

The twirl works because of the invariance of the singlet state under bilateral rotations, and therefore preserves the matrix element $M_{00} = \langle\Psi^-|M|\Psi^-\rangle$. Since our goal is to distill entanglement, we want to maximize the fully entangled fraction of the post-twirl state W_F. Based on the above discussion, however, it is clear that $f(W_F)$ will either equal M_{00} or be less than $\frac{1}{2}$. Hence, twirling will end any hope of distilling entanglement unless $M_{00} > \frac{1}{2}$. This seems rather wasteful, for there are many mixed states which are highly entangled which nevertheless satisfy $M_{00} \leq \frac{1}{2}$. We seek a more robust procedure, call it a modified twirl T', which preserves not the M_{00} element, but rather the largest of the set of diagonal elements $\{M_{00}, M_{11}, M_{22}, M_{33}\}$. For example, if a mixed state M contains a large fraction of the state $|\Phi^+\rangle$, then we wish to transform M into the state W_F with $F = \langle\Phi^+|M|\Phi^+\rangle$. Luckily, this can easily be done by first applying a unilateral σ_y, which swaps $|\Phi^+\rangle$ with $|\Psi^-\rangle$ (Table I), then applying the standard twirl T, followed by another unilateral σ_y which swaps the states back. With this procedure, the $|\Phi^+\rangle$ state is invariant under the modified twirl T'.

Having discussed some of the properties of the twirl operation and touted its utility in creating Bell-diagonal mixtures, we must return to the most unsettling feature of the operation: its reliance on a voluntary lack of information. For the twirl to work, Alice and Bob must willfully abstain from knowing which of the 12 possible SU(2) rotations has been applied to their pair of qubits. This is akin to suggesting that a guilty child who fears the wrath of his mother can better his situation by covering his eyes, as if his lack of information of his mother's whereabouts somehow reduces the likelihood of her finding him. Metaphors aside, we can address this issue in a very illuminating way. Suppose that the random SU(2) rotation is chosen by a computer, which then communicates the result to an apparatus at both Alice's and Bob's end which implements the appropriate rotation on the mixed state M. The computer is also attached to two printers, one at each end, and it prints out a number between 1 and 12 corresponding to the rotation that was chosen. Now, if Alice and Bob both remove their sheet from the printer, and, without looking at the number, lay the sheet face down in front of them, then they share the Werner state W_F between them. Now they can implement a distillation protocol designed to work on

Bell-diagonal mixtures, and by doing so they distill some number of maximally entangled states. Afterwards, suppose they turn over the sheet of paper and learn which rotation the twirl in fact applied. They learn that what they originally believed to be a Werner state shared between them was in fact just some rotated version of the state M. This does not change the fact that their protocol succeeded, and that they currently share some maximally entangled pairs. Furthermore, it would be absurd to suggest that the act of removing the paper from the printer and placing it face down was somehow instrumental in the success of the protocol. No, they could equally well look at the paper and learn immediately which rotation was performed; as long as they proceed by executing the same protocol that was used in the case where they did not peek at the paper, they will achieve the same successful result. So, the protocol works just as well whether they know which SU(2) rotation was applied or not. In particular, the rotation could have been the identity operation, meaning that the protocol distilled some entanglement directly from M. Therefore, the twirl was unnecessary in the first place!

We have demonstrated that any distillation procedure which relies on a twirl will work equally well without the twirl. Therefore, the twirl is best applied hypothetically rather than experimentally, and in this regard it serves as a book-keeping device: by keeping track of F, we know the upper bound on the available distillable entanglement at each step in the protocol. Since we have found it unnecessary to apply the twirl, one can ask why we bothered deriving any results for the twirl in the first place. For example, we claimed that the twirl can only reduce the amount of distillable entanglement present in a state M. This result is still true, but should be reinterpreted. Realize that we are still applying the twirl hypothetically, in the sense that our distillation procedures are limited to ones which work only when we can think of each pair as being in one of the four Bell states (and not any other state) with some probability. In that sense, the result should be restated as: 'by limiting our imagination to invent protocols which only work on Bell-diagonal *ensembles*, we could be neglecting other, more efficient protocols.' This argument should be re-read after understanding the dilution protocols, at which point it will be more clear.

(3) It will be useful to have a bitwise notation for the Bell states, given by

$$\begin{aligned} |\Phi^+\rangle &= 00 \\ |\Psi^+\rangle &= 01 \\ |\Phi^-\rangle &= 10 \\ |\Psi^-\rangle &= 11. \end{aligned} \tag{112}$$

The right, least significant bit is called the 'amplitude' bit and determines the Φ/Ψ nature of the state: $\Phi = 0$, $\Psi = 1$. The left, most significant bit is the 'phase' bit: $0 = +$, $1 = -$.

3.3.2. Recurrence method

The recurrence method is an iterative, 2-way protocol which takes as input a large number of pairs of particles, where each pair is characterized by the same density matrix M with $f(M) > \frac{1}{2}$. By making measurements on some of the pairs, it is possible to increase our information about the remaining unmeasured pairs. The output of each iteration is a reduced number of pairs, where each pair is characterized by a new density matrix M' such that $f(M') > f(M)$. In other words, the fully entangled fraction f is steadily increased by iterating the procedure. Of course, the measurement process necessarily wastes much of the initial entanglement; this is the price that must be paid in order to maneuver a mixed state into a maximally entangled pure state. As we will see, the method relies entirely on BXOR operations; for this reason, it is convenient to take the $|\Phi^+\rangle$ state as our 'standard' state, which has the desirable property of remaining invariant under the BXOR operation when used as both source and target pair. (This will make the algebra much simpler.) By 'standard' state, we mean that the fully entangled fraction should be given by $f(M) = \langle\Phi^+|M|\Phi^+\rangle$; given any Bell ensemble, this can always be arranged by applying appropriate unilateral Pauli matrices, as described for the modified twirl T'.

Suppose we begin with $2n$ pairs of particles, each in the bipartite mixed state M. We can apply the modified twirl T' (again, this need only be done hypothetically), resulting in the Werner state W_F, where $p_{00} = F$, and $p_{01} = p_{10} = p11 = (1 - F)/3$ (using the bit notation for the Bell states). Alice and Bob divide up the $2n$ pairs into n pairs of pairs. For each of these n sets, Alice and Bob apply the BXOR operation, using one of the pairs as the source and one as the target. Each of the 16 possible initial source/target combinations is drawn

with a particular probability dependent on W_F, and is deterministically mapped to another pair of states. The possibilities are listed in Table III, which is essentially a reproduction of the BXOR section of Table I, expressed in the bit notation. For example, the fourth line of

	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

TABLE III: Reproduced from [2]. Probabilities for each of the 16 initial source/target combinations, given that each pair is drawn from the same Bell ensemble, and the resulting states following a BXOR operation. The final column shows whether the target state passes (P) or fails (F) the test for being in a Φ state. This table is essentially a reproduction, in bit notation, of the BXOR section of Table I.

the table is arrived at by reasoning: the probability of choosing a $|\Phi^+\rangle$ state as the source pair and a $|\Psi^-\rangle$ state as the target pair is simply $p_{00}p_{11}$; by Table I, the BXOR maps the source pair to $|\Phi^-\rangle$ and the target pair to $|\Psi^-\rangle$, which is indicated by the 10 and 11 in the 'After BXOR' Source and Target columns, respectively.

Up to this point, nothing irreversible has been done to the system, so no entanglement has been distilled. Now, Alice and Bob perform a measurement on every target pair to distinguish whether it is in a Φ or Ψ Bell state. (This is easily done by locally measuring the z-component of spin of each qubit and comparing results: parallel spins $\Rightarrow \Phi$, antiparallel spins $\Rightarrow \Psi$.) Alice and Bob group together those source pairs whose corresponding target pairs were measured in a Φ state, and refer to this set as the 'Pass' set. Similarly, the source pairs whose corresponding target pairs were measured in a Ψ state are grouped into a 'Fail' subset. (Note: this grouping into two sets requires two-way communication, since Alice and Bob both need to know the result of each other's measurement.) The group membership of each source state is indicated in the rightmost column of Table III. To summarize: if the amplitude bit of the post-BXOR target state is 0, then the corresponding source state passes the test.

The probability that a given source pair passes the test is just the sum of all the starting probabilities which lead to a pass result,

$$p_{\text{pass}} = p_{00}^2 + p_{01}^2 + p_{10}^2 + p_{11}^2 + 2p_{00}p_{10} + 2p_{01}p_{11} \tag{113}$$

$$= \tfrac{1}{9}(8F^2 - 4F + 5). \tag{114}$$

Of the initial $2n$ pairs, n were measured and discarded, and of the remaining n pairs only the fraction p_{pass} make it into the 'Pass' subset. Within this subset, there is a new set of probabilities $\{p'_{00}, p'_{01}, p'_{10}, p'_{11}\}$ for each of the Bell states. For example, the only 00 states in the 'Pass' subset are descended from the initial pairings $\{00, 00\}$ and $\{10, 10\}$, and therefore p'_{00} (which is equal to the new fidelity F') is given by

$$p'_{00} = F' = (p_{00}^2 + p_{10}^2)/p_{\text{pass}} \tag{115}$$

$$= \frac{10F^2 - 2F + 1}{8F^2 - 4F + 5}. \tag{116}$$

To determine whether this is an improvement, the functions F' (blue) and F (red) are plotted in Fig. 7. Because the same plot holds for further iterations (provided a modified twirl T' is applied; explained at the end of this section), we can think of the recurrence protocol as an equilibrium-seeking process. There are three equilibrium points for this process–given by the intersection of the two plots–at $F = 0.25$, 0.5, and 1. For $0 \leq F \leq 0.25$, F' will quickly ascend to 0.25 and remain there. For $0.25 \leq F < 0.5$, the fidelity will slowly descend to 0.25 and remain there. Hence, $F = 0.25$ is a stable equilibrium. (The quick vs. slow convergence

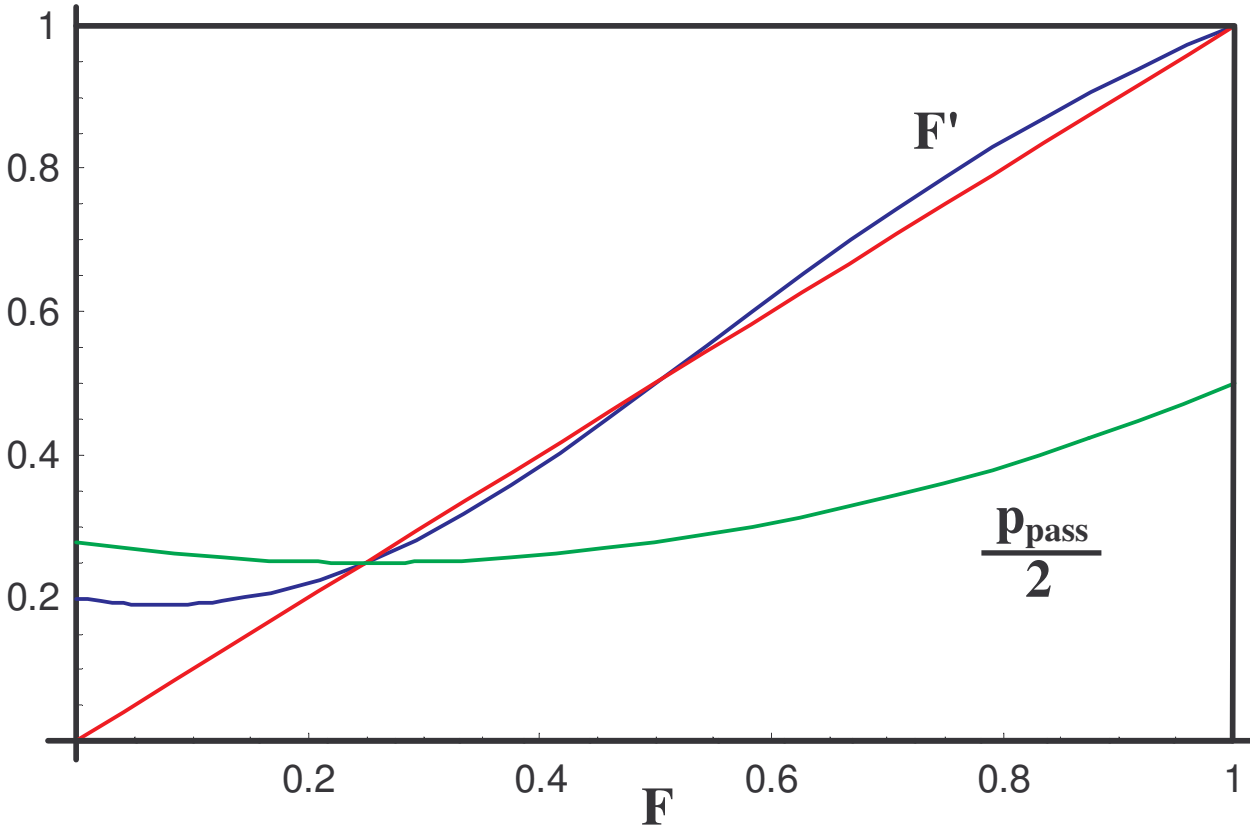


FIG. 7: Modified from [2]. Effect on the fidelity of Werner states by one iteration of the recurrence protocol. The protocol is able to distill entanglement from states with $F > 0.5$ because the final fidelity F' (blue) of the 'Pass' subset is greater than the initial fidelity F (red) in that region. The fraction of pairs $p_{\text{pass}}/2$ which remain after one iteration is also plotted (green).

is determined by the difference $|F - F'|$; larger differences imply faster convergence.) For $F = 0.5$ and $F = 1$, the fidelity will forever remain constant. Finally, the most important scenario is $F > 0.5$, in which case the recurrence protocol slowly raises the fidelity towards unity. (Recall: we had no right to expect the protocol to work for $F \leq 0.5$, since $E(W_F) = 0$ for those states. Therefore, the recurrence protocol is nice in the sense that it can distill at least some entanglement from any state which we can expect to distill entanglement from.) Also plotted in Fig. 7 is $p_{\text{pass}}/2$, the fraction of pairs remaining after each iteration (division by two takes care of the pairs which were measured). Since at least one half of the pairs are discarded at each iteration, the recurrence protocol is not efficient. In fact, since every small increase in fidelity requires a doubling of the initial number of pairs, the yield of the

protocol approaches zero for high output fidelity. The hashing method, to be introduced in Sec. 3.3.3, is much more efficient, but only works when the initial fidelity is greater than ≈ 0.8107. Hence, the recurrence method can be used to increase the fidelity past this point, after which the hashing method takes over. One can ask whether the efficiency can be increased by distilling some entanglement from the 'Fail' subset. It is easy to show that that the probabilities for each of the four Bell states in the 'Fail' subset is equal to $\frac{1}{4}$, so the answer is no, since this state is characterized by zero entanglement of formation.

The plot in Fig. 7 implicitly relies on a modified twirl T' applied after each iteration. The reason for this is that the probabilities of the other three states in the 'Pass' subset,

$$p'_{01} = (p_{01}^2 + p_{11}^2)/p_{\text{pass}} \tag{117}$$

$$p'_{10} = 2p_{00}p_{10}/p_{\text{pass}} \tag{118}$$

$$p'_{11} = 2p_{01}p_{11}/p_{\text{pass}}, \tag{119}$$

are not equal; hence, Eq. (116) is not valid for future iterations unless the twirl is used to equalize these probabilities. We stress again that the twirl, by introducing uncertainty, cannot increase the amount of distillable entanglement. In other words, the fidelity increase from F to F' given by Eq. (116) must represent the worst case scenario; any distribution other than $p_{01} = p_{10} = p_{11}$ will result in at least as high a fidelity increase. Therefore, a more efficient procedure is to ignore the twirl in between iterations, and always recalculate the expression for the subsequent fidelity based on the current values of p_{01}, p_{10}, and p_{11}. (Then, of course, the fidelity at each iteration would be a complicated function of the probabilities rather than rely only on the single parameter F, and could not be explained in a two dimensional plot such as Fig. 7.)

3.3.3. Universal hashing

The universal hashing protocol is a one-way communication protocol which begins with n pairs, each drawn from the same Werner mixed state W_F (after twirling), and distills $m \approx n(1 - S(W_F))$ pairs, each of arbitrarily high fidelity relative to the $|\Phi^+\rangle$ state, in the limit $n \to \infty$. For this reason, it works only when $S(W_F) < 1$, or equivalently $F > \approx 0.8107$.

A pure state given by the tensor product of n Bell pairs can be naturally represented using the bit notation as a $2n$-bit string x_0. For example, the $n = 4$ state $|\Phi^-\rangle \otimes |\Psi^-\rangle \otimes |\Phi^+\rangle \otimes |\Phi^-\rangle$

is written '10,11,00,10' (where the commas are only for clarity). Similarly, if each of the n pairs is drawn from a Bell-diagonal mixed state W_F, then the state can be represented as an ensemble of all possible $2n$-bit strings, where the probability of each string depends on its specific bit content and the probabilities associated with W_F. The parity of a bit string x is the modulo-2 sum of its bits, and can be calculated by taking the dot product of x with a string of the same length consisting of all 1s. For example, the parity of the $n = 4$ string mentioned above is $11111111 \cdot 10110010 = 0$. The parity of a given subset s of x is calculated using the dot product $s \cdot x$; for example, the parity of the subset consisting of the 3rd, 4th, 6th, and 7th bits of x is given by $00110110 \cdot 10110010 = 1$. Let us denote the true state of the n pairs shared between Alice and Bob by the string x_0, though of course Alice and Bob are unaware of its bit values. The hashing method relies on the ability of Alice and Bob to determine the parity of a random subset s_0 (known as the index string) of the string x_0, by only a single measurement on one of the n pairs. We will first demonstrate how this is done, and then explain why it is useful in distilling entanglement.

Again, let us examine the $n = 4$ case, and we assume that the randomly chosen subset is given by the string $s_0 = 00, 11, 01, 10$. (This is useful as an example because each of the four possible bit pairs are encountered.) Note that Alice can randomly choose s_0 and inform Bob of the result, so that both of them know the value of s_0 without any need for Bob to communicate back to Alice. Alice and Bob agree beforehand to choose as the 'destination pair' the pair of qubits which corresponds to the first pair of bits in s_0 that contains at least one non-zero bit. The 'destination pair' is the pair which will eventually be measured in order to determine the parity of the subset s_0 of x_0. In our example, the destination pair will be the second pair of x_0, because the first pair of s_0 is 00. (This stipulation is not absolutely necessary; Alice and Bob could agree to always use the first pair as the destination pair, regardless of the value of s_0. However, because the Bell pairs corresponding to 00 pairs in the index string have no effect on the desired subset parity, our convention conveniently avoids performing any operation–unitary or measurement–on these Bell pairs. This will reduce the 'back-action', discussed at the end of the section.)

Our goal is to make the amplitude bit of the destination pair equal to the desired subset parity. (Then, Alice and Bob each locally measure the spin of the destination qubit, and Alice sends her result to Bob. Bob will know that the parity of the subset is 0 if the spins are parallel and 1 if the spins are antiparallel.) This is accomplished in two steps. First,

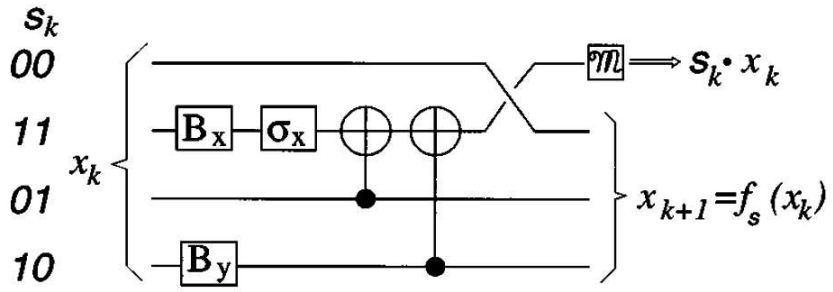


FIG. 8: Reprinted from [2]. The quantum circuit to find the parity of an unknown 8-bit string x_k, which represents the tensor product of 4 unknown Bell pairs, on the subset $s_k = 00, 11, 01, 10$. The desired quantity $s_k \cdot x_k$ is first accumulated in the amplitude but of the destination pair, after which a parallel spins measurement on that pair reveals the result. The remaining three Bell pairs are described by a 6-bit string $x_{k+1} = f_{s_k}(x_k)$.

Alice and Bob look at each pair of x_0 individually, and perform an operation which replaces the amplitude bit of that pair with the 'desired parity' of that pair. By 'desired parity,' we mean the parity of the Bell pair *on the subset specified by the corresponding pair in the index string.* (For example, if the fourth Bell pair happened to be in the state 11, then its desired parity would be 1 if the fourth pair of bits in s_0 was 10 or 01, or 0 if the fourth pair of index bits was 11 or 00.) Of course, Alice and Bob do not know the state of each of their Bell pairs, so the operation will have to be versatile enough to accomplish this mapping for any Bell pair. The operation for each pair will, however, depend on the corresponding index bits, so we treat the four cases separately:

00: As mentioned before, we do not have to perform any operations on Bell pairs corresponding to a 00 in the index string, since these pairs do not affect the parity.

01: Here, the desired parity of the Bell pair is already given by its amplitude bit, i.e. the desired parity of 00 is 0, of 01 is 1, of 10 is 0, and of 11 is 1. Hence, no operation is necessary.

10: In this case, the desired parity of the Bell pair is given by its phase bit, so we seek an operation which will replace the amplitude bit of a state with its phase bit. There is conceivably more than one operation which accomplishes this, since the value

to which the phase bit is mapped is irrelevant. One viable operation is the swap operation, which interchanges the two bits of the state. In our previous notation, the operation should map $\Phi^+(00) \rightarrow \Phi^+(00)$, $\Psi^+(01) \rightarrow \Phi^-(10)$, $\Phi^-(10) \rightarrow \Psi^+(01)$, and $\Psi^-(11) \rightarrow \Psi^-(11)$. This mapping is achieved by the B_y operation, as seen in Table I. Thus, Alice and Bob apply the B_y operation to each Bell pair whose corresponding index pair is 10.

11: The desired parity of the Bell pair is now the same as the full parity of the pair, which can be found by XORing the bits together. We seek an operation which XORs the two bits together and places the result in the amplitude bit. By following along in Table I, it can be seen that the operation $\sigma_x B_x$ achieves the mapping $\Phi^+(00) \rightarrow \Phi^+(00)$, $\Psi^+(01) \rightarrow \Psi^-(01)$, $\Phi^-(10) \rightarrow \Psi - +(11)$, and $\Psi^-(11) \rightarrow \Phi^-(10)$. Thus, Alice and Bob apply the $\sigma_x B_x$ operation to each Bell pair whose corresponding index pair is 11. Note: they should agree beforehand which of them will perform the unilateral σ_x. Also, $\sigma_x B_x = B_x \sigma_x$ (since rotations about the same axis commute), so Alice and Bob do not need to orchestrate when to perform each rotation; they simply perform the necessary rotation at their leisure.

Now that Alice and Bob have maneuvered the desired parity of each pair into the amplitude bit, the second step is to sum all of the amplitude bits modulo 2, and place the result in the amplitude bit of the destination pair. From Table I, note that the BXOR operation sums the amplitude bits of the initial source and target states and places the result in the amplitude bit of the resulting target state. (If the source is Φ, a target Φ is mapped to Φ, and a target Ψ is mapped to Ψ. If the source is Ψ, a target Φ is mapped to Ψ, and a target Ψ is mapped to Φ.) Thus, Alice and Bob perform the BXOR operation many times, using each Bell pair (other than the destination pair and those pairs whose corresponding index pair is 00) as the source once, and always using the destination pair as the target. They then perform the 'parallel-spins' measurement on the destination pair to learn the value of $s_0 \cdot x_0$, discard the measured pair, and are left with a string $2(n-1)$-but string x_1. The quantum circuit for the protocol is illustrated in Fig. 8 for the $n = 4$ case discussed.

Of course, the same procedure can now be applied to x_1 to obtain the parity on a randomly chosen subset s_1. In general, the hashing protocol consists of $n - m$ iterations of this procedure. At the start of the $(k+1)$st round, $k = 0, 1, \ldots, n-m-1$, Alice and Bob share

$n-k$ pairs drawn from the state W_F, the full state of which is described by an unknown $2(n-k)$-bit string x_k. Alice chooses a random index string s_k and shares it with Bob. The two of them perform the appropriate operations described above for each Bell pair, and finally measure the destination pair to obtain $s_k \cdot x_k$. The remaining $n-k-1$ pairs are described by a $2(n-k-1)$-bit string $x_{k+1} = f_{s_k}(x_k)$, which is deterministically computable from s_k and x_k.

To understand how the hashing protocol enables Alice and Bob to distill entanglement, we require two results from probability theory.

(1) The initial state x_0 can be described by any one of 2^{2n} distinct strings. However, by the Typical Sequence Theorem, x_0 will be one of the ϵ-typical sequences with probability greater than $1-\delta$, where δ can be made arbitrarily close to zero by taking $n \to \infty$. The number of possible typical sequences is at most $2^{n(S(W_F)+\epsilon)}$, where ϵ can be chosen arbitrarily close to zero as well. Hence, Alice and Bob can assume that x_0 is one of $\approx 2^{nS(W_F)}$ rather than 2^{2n} strings, which is a wrong assumption only a small fraction δ of the time.

(2) *Claim:* Given any two distinct strings $x \neq y$ and a randomly chosen index string s, the probability that the parities of x and y agree on s, i.e. $s \cdot x = s \cdot y$, is equal to $\frac{1}{2}$. (All addition is done modulus 2.) *Proof*: The expression $(s \cdot x) \oplus (s \cdot y) = s \oplus (x+y)$ is equal to 0 when the parities of x and y agree on s, and 1 when they disagree. Because x and y are distinct, $z \equiv x+y$ is a bit string which has a nonzero bit in the mth location $z(m)$, for at least one m. Without loss of generality, fix m to be one such location. The randomly chosen index string s will include the mth location, meaning $s(m) = 1$, exactly one half of the time. To calculate the parity $s \cdot z$, we first perform the bitwise dot product over all locations excluding m, the result of which, $s \cdot z - z(m)s(m) = s \cdot z - s(m)$, can be 0 or 1, and not necessarily with equal probability. Upon inclusion of the mth bit into the dot product, however, half the time the result is flipped (when $s(m) = 1$), and half the time the result stays the same (when $s(m) = 0$). In the end, then, the value of $s \cdot z$ depends purely on a coin toss. Hence, the parities of x and y agree on s exactly half the time. QED.

Before Alice and Bob begin, the full state is described by the string x_0, but all they know is that the state is one of $\approx 2^{nS(W_F)}$ distinct candidate strings, which we shall refer to collectively as y_0 strings. The goal of the hashing method is to determine the residual string x_{n-m} after $n-m$ rounds of hashing. (Note that this is less ambitious than determining the original string x_0–there is no need to discern the original state of the pairs that have already

been measured and discarded.) Each round of hashing informs Bob of the parity of x_k on a random subset s_k; he can then eliminate from the set of candidate strings any string y_k whose parity on s_k disagrees with $s_k \cdot x_k$. For each of the remaining candidates, Bob will compute the hypothetical trajectory $y_{k+1} = f_{s_k}(y_k)$, eventually honing in on exactly one possible y_{n-m} with high probability, which must then equal the true state x_{n-m}. To see this, let us follow the trajectory of one of the original 'decoy' candidate strings y_0 (not equal to the true state x_0) through $n-m$ rounds of the hashing procedure. The scenario which worries us is the following: the trajectory y_k is such that for every k up to $n-m-1$, the parity $s_k \cdot y_k$ agrees with the measured parity $s_k \cdot x_k$, yet the residual string y_{n-m} is not equal to the true residual string x_{n-m}. Mathematically, $\forall_{k=0}^{n-m-1}(s_k \cdot y_k = s_k \cdot x_k)$ AND $y_{n-m} \neq x_{n-m}$. In this case, the decoy y_0 managed to pass all the same parity tests as the true string x_0 (without ever converging to the same residual string), thereby preventing Bob from determining the true identity of the string x_{n-m}. We can place an upper bound on the probability of this event occurring for any single y_0. Since $y_{n-m} \neq x_{n-m}$, it must be true that the intermediate strings along the trajectory, y_k and x_k, were distinct for every k (otherwise, they would have converged to the same string.) Therefore, the probability of yielding the same parity result for all $n-m$ rounds is $2^{-(n-m)}$ (by item (2), above). Additionally, y_k remains distinct from x_k after each round with probability ≤ 1, and therefore the probability of a particular candidate string fooling Bob into misidentifying x_{n-m} is bounded by

$$p[\forall_{k=0}^{n-m-1}(s_k \cdot y_k = s_k \cdot x_k) \text{ AND } y_{n-m} \neq x_{n-m}] \leq 2^{-(n-m)}. \tag{120}$$

This must be multiplied by the total number of candidate strings, $2^{n(S(W_F)+\epsilon)}$, to determine the total probability that more than one candidate string remains at the end of the procedure. Additionally, we should include the probability of zero candidate strings remaining at the end of the procedure, which is simply given by the probability that x_0 was not in fact one of the typical strings ($\leq \delta$, by item (1), above). Thus, the probability that the hashing method fails to uniquely identify the string x_{n-m} is bounded above by

$$p_{\text{failure}} \leq 2^{n[S(W_F)+\epsilon]-(n-m)} + \delta. \tag{121}$$

We already know $\delta \to 0$ as $n \to \infty$, and the first term can be made arbitrarily small by performing $n - m = n[S(W_F) + 2\epsilon]$ rounds of hashing. At this point, assuming Alice has been sending Bob the results of her measurements after every round and Bob has been

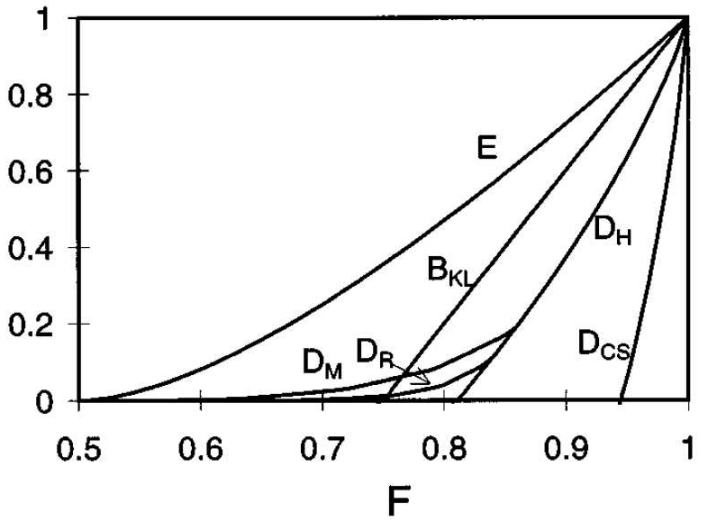


FIG. 9: Reprinted from [2]. Distillation yield of the mixed state distillation protocols. D_H gives the yield of the hashing protocol, and D_R gives the yield of the recurrence protocol when followed by the hashing protocol (once hashing is more efficient). E is the entanglement of formation of the Werner state W_F, and serves to illustrate the upper bound on distillable entanglement. (The lines D_M, B_{KL}, and D_{CS} are not discussed in this paper.)

concurrently keeping track of the hypothetical trajectories of each of the initial candidate strings y_0, Bob will with probability $1 - p_{\text{failure}}$ know the exact values of the bit string x_{n-m}. By applying unilateral Pauli rotations, he can transform each pair into the standard $|\Phi^+\rangle$ state, giving an asymptotic distillation yield of $m/n = 1 - S(W_F)$ (Fig. 9).

We have made a big fuss about this idea that Bob must keep track of the hypothetical trajectories of each of the $2^{nS(W_F)}$ candidate strings y_0 as he and Alice perform each round of hashing. Since the number of candidate strings increases exponentially with n, it is uncertain whether there is even an efficiently computable way to do this. One might wonder, could Bob wait until the measurement results from all $n - m$ rounds were in, and then use some algorithm to more efficiently determine x_{n-m}? This would certainly be possible if the unitary operations that were applied along the way did not affect the state of the unmeasured pairs. Then, at the end of the procedure Bob would know the parity of $n-m$ different subsets of x_0, and he could quickly and uniquely determine the states of the remaining m pairs. As it is, though, Bob only knows the parities of a random subset of each of the x_k's. Granted, these strings are all deterministically related to each other through the various functions f_{s_k}; the

problem is that the cumulative effect of the f_{s_k}'s at every round can be quite complicated, since the operations $\sigma_x B_x$, B_y, and BXOR are acting on unknown states. This complication is called 'back-action'–which refers to the difficult procedure of working backwards to find x_{n-m}, once all the measurement results are known. The method we have presented (where Bob keeps track of all hypothetical candidate trajectories) is a nice conceptual way to see how the problem of back-action can be dealt with, though not necessarily the most efficient.

3.3.4. Direct purification of non-Bell-diagonal mixtures

The recurrence and hashing methods are only able to distill entanglement from a mixed state M when we can think of M as arising from a Bell diagonal ensemble. Again, since we can always transform M into W_F by a hypothetical twirl, this is equivalent to saying that the recurrence and hashing methods can only distill entanglement from states M whose fully entangled fraction $f(M) > 0.5$. (Recall that a suitable modified twirl can always be used to ensure that the fidelity F of the resulting Werner state is equal to $f(M)$, and entanglement can only be distilled from W_F when $F > 0.5$.) Therefore, our two methods are unable to distill entanglement from, say, the state

$$M = \tfrac{1}{2}|00\rangle\langle 00| + \tfrac{1}{2}|\Psi^+\rangle\langle\Psi^+|, \tag{122}$$

for which $f(M) = \frac{1}{2}$. To find a protocol which can distill entanglement from this state requires that we not confine ourselves to the notion that all the intermediate states created during the protocol must be Bell states. (Then, we won't be limited by the entanglement of formation of Bell-diagonal mixtures, which is 0 when $f(M) \leq \frac{1}{2}$.) In fact, this protocol is the simplest of the three we have encountered.

We are free to think in terms of any ensemble which realizes M, and in this case we choose the obvious one: $\{50\%\,|00\rangle, 50\%\,|\Psi^+\rangle\}$. Alice and Bob begin with n pairs, each in the state M, and group them into $n/2$ pairs of pairs. These $n/2$ sets are each subjected to a BXOR operation, with one pair as the source and the other as the target. Then, both Alice and Bob measure the spin of their target qubit and compare results. Because each of the two pairs can start in one of two states, there are four equally weighted possibilities to consider.

(1) BXOR($|\Psi^+\rangle \otimes |\Psi^+\rangle$) = $|\Psi^+\rangle \otimes |\Phi^+\rangle$. Measurement yields up-up or down-down, each with probability one half. The source pair is left in the maximally entangled state $|\Psi^+\rangle$.

(2) BXOR($|\Psi^+\rangle \otimes |00\rangle$) = $\frac{1}{\sqrt{2}}(|01\rangle \otimes |01\rangle + |10\rangle \otimes |10\rangle)$. Measurement yields up-down or down-up, each with probability one half. The source pair is left in an unentangled state.

(3) BXOR($|00\rangle \otimes |\Psi^+\rangle$) = $|00\rangle \otimes |\Psi^+\rangle$. Measurement yields up-down or down-up, each with probability one half. The source pair is left in an unentangled state.

(4) BXOR($|00\rangle \otimes |00\rangle$) = $|00\rangle \otimes |00\rangle$. Measurement yields up-up with probability one. The source pair is left in an unentangled state.

Given all the possibilities, we see that the measurement will yield down-down with probability 1/8, and the remaining source state is guaranteed to be in the Bell state $|\Psi^+\rangle$. Thus, the yield from this procedure is $D_2 = 1/16$, since half of the pairs are always measured and discarded. (Two-way communication is needed because Alice and Bob must both know the result of the other's measurement in order to know which pairs to keep, hence the use of D_2.) The straightforward generalization of this analysis to the state

$$M = (1-p)|00\rangle\langle 00| + p|\Psi^+\rangle\langle\Psi^+| \tag{123}$$

shows that pure states $|\Psi^+\rangle$ can be distilled with yield $D_2 = p^2/4$.

It was left unmentioned in [2] that this yield can be improved by iterating the procedure in the following manner. First of all, whenever the measurement result is up-down or down-up, we discard the remaining source state because it is definitely unentangled. However, the measurement result up-up will occur 3/8 of the time; 1/8 of the time, this result comes from case (1), so the source state is $|\Psi^+\rangle$; 1/4 of the time, this result comes from case (4), so the source state is $|00\rangle$. Thus, the measurement result up-up places the source pair in the mixed state of Eq. (123), with the parameter $p' = 1/3$. There will be $n' = (3/16)n$ pairs left in this state, so we can distill another $(\frac{p'^2}{4})n' = n/192$ pairs, thereby increasing the yield to $D_2 = \frac{1}{16} + \frac{1}{192}$. Again, we can operate on the newly created mixed state consisting of the

source pairs whose targets were measured in the up-up state to further increase the yield, although the gains rapidly decrease. More generally, given n pairs in the state M of Eq. (123), $(p^2/4)n$ maximally entangled states can be distilled right away, and a quantity of

$$n' = \left(\frac{p^2}{4} + \frac{(1-p)^2}{2} \right) n \tag{124}$$

pairs in the state of Eq. (123), with parameter

$$p' = \frac{p^2}{\frac{p^2}{4} + \frac{(1-p)^2}{2}} \tag{125}$$

can be used to distill further entanglement by iterating the procedure. An expression for $D_2(p)$ in closed form was not found.

4. RELATIONSHIP OF MIXED STATE PURIFICATION PROTOCOLS TO QUANTUM ERROR CORRECTION

As suggested earlier, entanglement purification protocols (EPPs) offer a form of quantum error correction, by virtue of their ability to create maximally entangled pairs for use in noiseless quantum teleportation. In this section we will make this connection explicit.

Suppose Alice wants to send Bob a particular quantum state. The word 'send' is very general, since it does not necessarily involve physical relocation of a qubit–it may use teleportation. It is possible that noise affects the transmission of the qubit, so that Bob does not receive quite the same state that Alice intended him to receive. In general, a quantum channel χ can be viewed as a completely positive, trace-preserving map from pure states to mixed states. Now, a channel is only useful if there is some way of reliably transmitting information through it, and for this reason we use quantum error correcting codes (QECCs). Alice uses a QECC to introduce redundancy into her state $|\psi\rangle$, so that errors introduced by the channel do not prevent Bob from reconstructing (decoding) $|\psi\rangle$ faithfully at his end. For instance, if Alice's goal is to send m qubits to Bob, then she may introduce $n - m$ ancillas at her end and perform a collective encoding operarion U_e on the n qubits. She then requires n channel uses to send the full encoded state to Bob. If Bob can decode the original state of the intended m qubits successfully, then the QECC has succeeded. We define the rate R of the code as $R = m/n$, the number of intended qubits divided by the number of channel uses. Thus, $R \leq 1$ and serves as a measure of the redundancy required

of the code to overcome the noise in the channel. Of course, some codes may be better than others, but there is still a maximum rate of information that can be sent through a given channel χ. Therefore, we define the quantum capacity $Q(\chi)$ of a channel as the maximum R over all possible QECCs. As with EPPs, it is conceivable that the rates of QECCs can be improved by allowing a one-way or two-way classical communication side channel between Alice and Bob, so we must define the corresponding channel capacities $Q_1(\chi)$ and $Q_2(\chi)$. Since the possibility of classical communication cannot decrease a channel's capacity, clearly $Q(\chi) \leq Q_1(\chi) \leq Q_2(\chi)$. However, it can be shown [2], though it will not be proven here, that in fact

$$Q(\chi) = Q_1(\chi) < Q_2(\chi). \tag{126}$$

The first equality is surprising, because it means that any QECC that can be implemented using one-way classical communication can be implemented equally well without the classical communication. Because most of the literature on QECCs ignores the possibility of classical side channels, this result is important because it will allow us to make explicit connections between our 1-EPPs (one-way purification protocols) and the well-known QECCs already in existence. The strict inequality $Q(\chi) < Q_2(\chi)$ means that there is a class of QECCs relying on two-way classical communication which is more efficient than currently well-known QECCs. It turns out that the 2-EPPs we have discussed can be used to implement such codes.

In this paper, we will prove explicit relations between the capacity $Q(\chi)$ of the channel χ and the one-way entanglement distillation yield $D_1(M)$ of the mixed state M, by describing specific protocols for which a 1-EPP can be used as a QECC and vice versa. In order to compare the two, we clearly need a relationship between channels χ and mixed states M. For an important class of channels, including depolarizing channels, there exists a one-to-one mapping between χ and M, which we now describe. (Note: a p-depolarizing channel leaves the input state unaltered which probability $1-p$ and replaces it with the totally mixed state with probability $1-p$. Alternatively, we can say that the channel applies the identity operation with probability $1-\frac{3p}{4}$, and applies each of the Pauli operators σ_x, σ_y, and σ_z with probability $\frac{p}{4}$.)

The mapping from mixed states to channels $\hat{\chi}(M)$ is defined as the quantum channel that results when using the quantum teleportation protocol with the state M as the entangled resource. Clearly, $\hat{\chi}(M)$ will be a noisy channel unless M is a maximally entangled state.

For example, consider what happens when Alice tries to transmit the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ to Bob using teleportation when the shared resource is in the Werner state W_F. With probability F, the resource is in the state $|\Phi^+\rangle$ (for which Alice and Bob have designed the particular teleportation protocol), and faithful tranmission will result. If one of the other Bell states is used as the shared resource, then Bob will not apply the correct fix-up operation after receiving Alice's measurement result. The net effect is that Bob's final state is given by $\sigma_x|\psi\rangle$, $\sigma_y|\psi\rangle$, or $\sigma_z|\psi\rangle$, each with probability $\frac{1-F}{3}$. Thus, the teleportation procedure is equivalent to a $\frac{4(1-F)}{3}$-depolarizing quantum channel.

The inverse mapping $\hat{M}(\chi)$, from channels to mixed states, is defined as the mixed state M that results from the action of the channel χ on one half (say, Bob's half) of a maximally entangled input pair. We will see that this is a natural definition to use, because the most obvious way for Alice and Bob to share entanglement is for Alice to prepare a Bell pair, and send one of the qubits to Bob through a noisy channel. For instance, if Alice prepares the Bell state $|\Phi^+\rangle$ and sends Bob his qubit through a p-depolarizing channel, the final state will remain as $|\Phi^+\rangle$ with probability $1 - \frac{3p}{4}$, or will have become the state $|\Psi^+\rangle$, $i|\Psi^-\rangle$, or $|\Phi^-\rangle$, each with probability $\frac{p}{4}$. Hence, the final state is in fact a Werner state with fidelity $F = 1 - \frac{3p}{4}$.

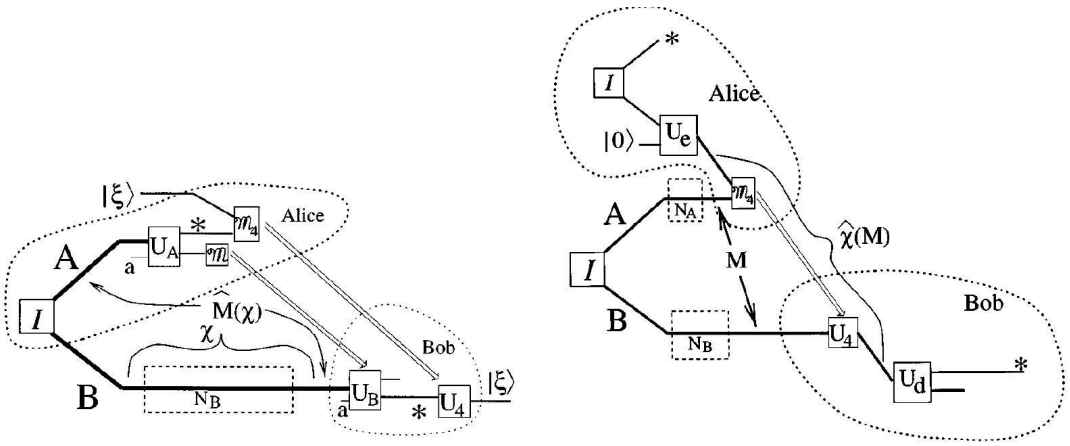
Note that for the 'Werner state-depolarizing channel' example we have done, the two mappings are inverses of each other:

$$\begin{aligned} \hat{M}(\hat{\chi}(W_F)) &= \hat{M}\left(\frac{4(1-F)}{3}\text{-depolarizing channel}\right) \\ &= 1 - \frac{3}{4}\left(\frac{4(1-F)}{3}\right) \\ &= W_F. \end{aligned} \tag{127}$$

In fact, this result is true not only for Werner states, but for all Bell-diagonal mixed states M and corresponding channels $\hat{\chi}(M)$. Consider a general Bell-diagonal state

$$M = p_0|\Phi^+\rangle\langle\Phi^+| + p_1|\Phi^-\rangle\langle\Phi^-| + p_2|\Psi^+\rangle\langle\Psi^+| + p_3|\Psi^-\rangle\langle\Psi^-|. \tag{128}$$

If a teleportation protocol, which assumes the shared resource is $|\Phi^+\rangle$, is used to teleport $|\psi\rangle$ using the Bell-diagonal state M, then faithful transmission will result with probability p_0. Otherwise, the the final state will be either $\sigma_x|\psi\rangle$, $\sigma_y|\psi\rangle$, or $\sigma_z|\psi\rangle$ with probability p_1, p_2, or p_3 (not necessarily respectively; the particular probability associated with each state



(a) (b)

FIG. 10: Reprinted from [2]. (a) A QECC is derived using a 1-EPP. Alice locally prepares Bell pairs and sends Bob the appropriate half using the channel χ, resulting in the shared state $\hat{M}(\chi)$. Using the 1-EPP, they distill maximally entangled pairs (*) and use them to faithfully teleport the state $|\xi\rangle$. (b) A 1-EPP is derived from a QECC. Alice and Bob share the mixed state M, which can be used as a channel $\hat{\chi}(M)$ via teleportation. Using a QECC $\{U_e, U_d\}$, Alice and Bob can share maximally entangled pairs (*) at the rate given by the capacity of the channel, thereby establishing a 1-EPP whose yield is equal to $Q(\hat{\chi}(M))$.

depends on the teleportation protocol). Therefore, $\hat{\chi}(M)$ is the channel which applies one of the four Pauli matrices (including the identity) with some probability. Clearly, the action of this channel on a fully entangled input state will result in the original Bell-diagonal mixed state that we started with (Eq. (128)). Hence, $\hat{M}(\hat{\chi}(M)) = M$ for Bell-diagonal states M, as claimed.

We are now ready to prove the following two inequalities concerning the distillable entanglement of a general mixed state and the capacity of a quantum channel,

$$Q(\chi) \geq D_1(\hat{M}(\chi)) \tag{129}$$

$$D_1(M) \geq Q(\hat{\chi}(M)). \tag{130}$$

Equation (129) says that the capacity of any channel χ must be at least as great as the one-way distillable entanglement of the corresponding mixed state $\hat{M}(\chi)$. To prove this, we consider a situation where Alice can send Bob quantum information using a channel χ, and

demonstrate an explicit QECC which relies on a 1-EPP (Fig. 10(a)). In fact, the code is very simple. Alice prepares n pairs in a maximally entangled state, and sends each one to Bob using the channel χ. The resulting shared state is, by definition, $\hat{M}(\chi)^{\otimes n}$. Then, using a 1-EPP, Alice and Bob can distill m pure maximally entangled states, where $m/n = D_1(\hat{M}(\chi))$. Once they share these m Bell pairs, Alice uses teleportation to faithfully transmit m qubits to Bob. In effect, they have used the quantum channel n times to faithfully transmit m qubits, so we have described a QECC whose rate is $R = m/n = D_1(\hat{M}(\chi))$. The inequality (129) follows because the capacity $Q(\chi)$ is the maximum rate R over all possible QECCs, and we have only described one.

Equation (130) says that the one-way distillable entanglement of a state M must be at least as great as the capacity of the corresponding quantum channel $\hat{\chi}(M)$. To prove this, we consider a situation where Alice and Bob share n pairs, each in the mixed state M, and demonstrate an explicit 1-EPP which relies on a QECC (Fig. 10(b)). As discussed before, teleportation using a mixed state M is equivalent to sending a qubit through a noisy quantum channel $\hat{\chi}(M)$. There exists a QECC whose rate R is equal to the channel capacity $Q(\hat{\chi}(M))$. Alice's and Bob's goal is to share m maximally entangled pairs. To accomplish this, Alice prepares a state of m maximally entangled pairs locally, then uses the QECC to encode the state of Bob's m qubits into n qubits, where m/n is the channel capacity $Q(\hat{\chi}(M))$. (She does this by appending $n-m$ ancillas and performing the encoding operation U_e.) Then, she can teleport these qubits to Bob using the n impure pairs that she and Bob already share. Using the decoding operation of the QECC, Bob is able to faithfully reconstruct the m halves of maximally entangled pairs that Alice prepared for him. In effect, Alice and Bob have managed to distill m maximally entangled states from n copies of the mixed state M, so we have described a 1-EPP whose distillation yield is $m/n = Q(\hat{\chi}(M))$. The inequality (130) follows because the distillable entanglement $D_1(M)$ is the maximum yield over all such possible 1-EPPs, and we have only described one.

In the case of Bell-diagonal mixed states, we can define $M = \hat{M}(\chi)$ and $\chi = \hat{\chi}(M)$ because the two mappings are inverses of each other. Hence, the inequalities (129) and (130) together imply that

$$D_1(M) = Q(\chi), \tag{131}$$

the one-way distillable entanglement of a mixed state M is equal to the capacity of the corresponding channel χ. The above proofs work equally well when considering two-way

classical communication, except that we can only compare the quantities D_2 and Q_2:

$$Q_2(\chi) \geq D_2(\hat{M}(\chi)) \tag{132}$$

$$D_2(M) \geq Q_2(\hat{\chi}(M)), \tag{133}$$

and if the mapping between χ and M is reversible then

$$D_2(M) = Q_2(\chi). \tag{134}$$

---

[1] C.H. Bennett, H.J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, **53**, 2046 (1996).

[2] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, and W.K. Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, **54**, 3824 (1996).

[3] B. Huttner, A. Muller, J.D. Gautier, H. Zbinden, and N. Gisin. Unambiguous quantum measurement of non-orthogonal states. *Phys. Rev. A*, **54**, 3783 (1996).

[4] P.M. Hayden, M. Horodecki, and B.M. Terhal. The asymptotic entanglement cost of preparing a quantum state. *J. Phys. A*, **34**, 6891 (2001).