Concentrating entanglement by local actions: Beyond mean values

Hoi-Kwong Lo
MagiQ Technologies, Inc., 275 Seventh Avenue, 26th Floor, New York, New York 10001-6708

Sandu Popescu
H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, United Kingdom and BRIMS, Hewlett-Packard Labs, Filton Road, Stoke Gifford, Bristol, BS34 8QZ, United Kingdom
(Received 2 September 1997; published 4 January 2001)

Suppose two distant observers Alice and Bob share a pure bipartite quantum state. By applying local operations and communicating with each other using a classical channel, Alice and Bob can manipulate it into some other states. Previous investigations of entanglement manipulations have been largely limited to a small number of strategies and their average outcomes. Here we consider a general entanglement manipulation strategy, and go beyond the average property. For a *pure* entangled state shared between two separated persons Alice and Bob, we show that the *mathematical* interchange symmetry of the Schmidt decomposition can be promoted into a *physical* symmetry between the actions of Alice and Bob. Consequently, the most general (multistep two-way-communications) strategy of entanglement manipulation of a pure state is, in fact, equivalent to a strategy involving only a single (generalized) measurement by Alice followed by one-way communications of its result to Bob. We also prove that strategies with one-way communications are generally more powerful than those without communications. In summary, one-way communications is necessary and sufficient for entanglement manipulations of a *pure* bipartite state. The supremum probability of obtaining a maximally entangled state (of any dimension) from an arbitrary state is determined, and a strategy for achieving this probability is constructed explicitly. One important question is whether collective manipulations in quantum mechanics can greatly enhance the probability of large deviations from the average behavior. We answer this question in the negative by showing that, given n pairs of identical partly entangled *pure* states ($|\Psi\rangle$) with entropy of entanglement $E(|\Psi\rangle)$, the probability of getting nK $[K>E(|\Psi\rangle)]$ singlets out of entanglement concentration tends to zero as n tends to infinity.

DOI: 10.1103/PhysRevA.63.022301 PACS number(s): 03.67.Dd, 03.65.Bz, 42.50.Dv, 89.70.+c

I. INTRODUCTION AND SUMMARY OF KEY RESULTS

Entanglement—the nonlocality of entangled state—as in the Einstein-Podolsky-Rosen paradox [1], discovered by Bell [2] in 1964, has long been regarded a hallmark of quantum mechanics. In the past, entanglement was often regarded as a qualitative property of a state. The last few years, however, witnessed a dramatic change in the approach to entanglement. Entanglement is now regarded as an important quantitative and useful resource in achieving tasks of quantum information processing such as dense coding [3], teleportation [4], and reduction of communication complexity [5].

The study of quantum information processing has been complicated by the fact that entanglement can appear in many nonstandard forms. Fortunately, we have learned that two distant parties sharing a bipartite pure state can apply local operations and classical communication to ''manipulate'' entanglement, thus converting one form to another [6]. To better understand quantum information processing, it is important to discover the fundmental laws of a general entanglement manipulation. Most previous investigations focused on some particular types of entanglement manipulations, namely, entanglement concentration and dilution of an ensemble of identical states, say $\Omega=(a|11\rangle+b|22\rangle)^N$, which are collectively processed. Moreover, the main interest was in the average properties, and little is known about the actual probability distribution of the outcomes of those manipulations.

This paper concerns mainly the fundamental laws of entanglement manipulations of *pure* bipartite states. Unlike previous investigations, here we allow the initial state to be a single copy of a general state Ψ. It is useful to note, however, that in fact all entanglement manipulation methods, both ''single-pair'' and ''collective'' ones can be reformulated as single-pair methods, by redefining the ''particles.'' Indeed, suppose Alice and Bob share n pairs of particles, and intend to process them by some collective method. We can now regard all n particles in each side as a single particle, living in a higher-dimensional Hilbert space (equal to the product of the Hilbert spaces of the original n particles). The n original pairs can thus be regarded a single pair of two (more complex) quantum particles, and the original ''collective'' manipulation can be regarded as a single-pair type manipulation of this new pair. Consequently, all questions concerning collective manipulations can be answered by studying single-pair manipulations of a generic state of two arbitrary particles. This is the path that we will follow in the present paper. Our results are the following.

(i) First of all, we show that, rather surprisingly, general entanglement manipulations of a pure bipartite state with only one-way communication are equally powerful as those with two-way communication, but are more powerful than those with no communication.

(ii) Then we specialize in a class of entanglement manipulations, namely, entanglement concentration, from a general pure bipartite initial state Ψ to a m-dimensionally maximally entangled state (which we shall call an m-ME state), $\Phi_m = (1/\sqrt{m})\Sigma_{i=1}^{m}|i\rangle_A|i\rangle_B$, where $|i\rangle_A$ and $|i\rangle_B$ are orthonormal vectors in Alice's and Bob's Hilbert spaces H_A and H_B, respectively, and m is fixed but arbitrary. (Note that Φ_1 is a direct product, Φ_2 is a singlet, and Φ_{2q} is equivalent to q singlets. Therefore, an m-ME state can be regarded as a generalization of singlets.) The main questions that we ask are the following: What is the optimal probability of getting a Φ_m from Ψ? What is the optimal strategy that will achieve such a probability? In this paper, we give complete answers to those questions.

(iii) After this, we specialize in the entanglement manipulation of a large number of identical pairs of Φ, and derive a bound to the probability of having a large deviation from the average property. More specifically, suppose that two remote observers, Alice and Bob, share n pairs of spin-1/2 particles, each pair in a nonmaximally entangled pure state $|\Psi\rangle = \alpha|1\rangle|1\rangle + \beta|2\rangle|2\rangle$. Then, by local actions (which may include local unitary transformations, measurements, and attachment of ancillary quantum systems) and classical communications, Alice and Bob can convert these pairs into a (smaller) number m of perfect singlets. It has been shown [6] that in the limit of large n, Alice and Bob can perform a *reversible* conversion of the n pairs Ψ into singlets, obtaining, *on average*, a number $\overline{m} = nE(\Psi)$ of singlets. (Here, $E(\Psi)$ is the "entropy of entanglement" of a pure bipartite state consisting of subsystems A and B, and is defined to be the von Neumman entropy of subsystem A (or B) [6].) Furthermore, as a consequence of this reversibility property, together with the fact that on average entanglement cannot increase via local actions and classical communications [6,7], this particular entanglement manipulation method yields the maximal possible average number of singlets. It has been shown [8] that $E(\Psi)$, the maximal average number of singlets which can be extracted per original pair Ψ, is up to a constant multiplicative factor, the *unique* measure of entanglement for Ψ that is nonincreasing under local operations and classical communication.

However, in all previous investigations of entanglement manipulations, the focus was on the *average values*, such as on the question "What is the average number of singlets which can be extracted from n pairs Ψ?" Indeed, the whole idea of reversibility refers only to the average properties. The point is that, as in all asymptotic results, there is always a very small but nonzero probability for an entanglement concentration procedure to give a substantially smaller amount of singlets than the expected number. Here we want to go beyond average values and ask about the actual distributions. For example, the same average number of singlets, $\overline{m} = nE(\Psi)$ might, in principle, be obtained from very different distributions: In the reversible procedure described in Ref. [6], out of n pairs Ψ a number m of singlets is obtained with some probability P_m, and the distribution is essentially Gaussian, peaked around $\overline{m} = nE(\Psi)$. In particular, via this procedure the probability to obtain a large number of singlets, $m \approx n$ is exponential small. However, one could envisage a distribution which yields the same average $\overline{m} = nE(\Psi)$ while having a non-negligible probability for obtaining a large number of singlets—for example, a distribution in which the probability of obtaining $m = n$ singlets is $E(\Psi)$, while in all other cases zero singlets are obtained. The question is "Does there exist any entanglement manipulation procedure which realizes the latter distribution?"

A main point of our investigation is to gain a better understanding of the *collective properties* involved in entanglement manipulation. Indeed, if Alice and Bob were to extract singlets by processing each of the n pairs Ψ separately, the law of large numbers tells that the probability distribution of the number of singlets will (asymptotically) be Gaussian. Deviations from this distribution can be obtained (if at all) only if Alice and Bob process all the n pairs together. But are such deviations possible? And if so, how large can they be? (To put things in the right perspective, we would like to mention that the reversible procedure [6] discussed above is *not* a procedure in which each pair is processed separately but a collective one; however, the distribution it yields is essentially Gaussian.)

In this paper, we show that, in the context of entanglement manipulations of a large ensemble of identical pure bipartite states, collective manipulations cannot substantially enhance the probability of large deviations from the average properties.

(iv) Afterward, we consider a subclass of entanglement manipulation strategies in which the final state is always one of the possible Φ_m's. (Say, the state Φ_1 appears with probability p_1, Φ_2 with probability p_2, Φ_n with probability p_n, etc.) We define a natural notion of a universal strategy for entanglement manipulations for all values of m, and show that such a universal strategy cannot possibly exist.

(v) Finally, we present open questions in the entanglement manipulations of pure states and briefly discuss problems in generalizing our results to mixed states.

In summary, on a conceptual level, the novelties of our investigation are the following.

(1) Some of our results—particularly the statement that manipulations with one-way communication are as powerful as those with two-way communication—apply to a general entanglement manipulation strategy. In contrast, all previous investigations focused on either entanglement concentration or dilution.

Incidentally, we clearly demonstrate the important role of symmetry in entanglement manipulations. Indeed, in our proof that manipulation strategies of a pure bipartite state with only one-way communication are as powerful as those with two-way communication, we are essentially promoting the mathematical interchange symmetry of the Schmidt decomposition into a physical symmetry between the local actions of Alice and Bob. (See Sec. II.)

(2) We allow the initial state to be a general Ψ. In contrast, almost all previous investigations restricted Ψ to be N identical copies of some state, say $u = a|11\rangle + b|22\rangle$, and were mainly concerned with properties in the large-N limit.

(3) We are concerned with the role of probability rather than the average property in entanglement manipulations.

II. ONE-WAY COMMUNICATIONS IS NECESSARY AND SUFFICIENT FOR ENTANGLEMENT MANIPULATIONS OF BIPARTITE PURE STATES

A. Reduction from two-way to one-way communications

The most general scheme of entanglement manipulations of a bipartite pure entangled state involves two-way communications between Alice and Bob. It goes as follows: Alice performs a measurement and tells Bob the outcome. Bob then performs a measurement (the type of measurement that Bob performs can depend on Alice's measurement outcome), and tells Alice the outcome, etc., etc. The goal of this subsection is to prove that any strategy of entanglement manipulation of a pure bipartite state is equivalent to a strategy involving only a *single* (generalized) measurement by Alice followed by the *one-way* communications of the result from Alice to Bob (and finally local unitary transformations by Alice and Bob).

Here we introduce some definition.

Definition 1. (ordered Schmidt coefficients): An arbitrary pure state Ψ can be written in Schmidt decomposition [9]

$$\Psi=\sum_{i}^{N}\sqrt{\lambda_i}|a_i\rangle|b_i\rangle, \tag{1}$$

where $\langle a_i|a_j\rangle=\langle b_i|b_j\rangle=\delta_{ij}$. We call $\sqrt{\lambda_i}$'s the *ordered Schmidt coefficients* if the λ_i's are ordered decreasingly, i.e., $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_N$. Note that all phases have been absorbed in the definition of the $|a_i\rangle$ states, so that the λ_i's are positive real numbers.

First of all, since it is more convenient to deal with projection operators than positive-operator-valued measures, we include any ancilla (measuring apparatus) in Alice and Bob's quantum systems. Therefore, without loss of generality, we regard Alice and Bob as sharing a pair of particles with an infinite (or an arbitrarily large) dimensional Hilbert space; however initially only N of the coefficients of the Schmidt decomposition [9] are nonzero, i.e., $|\Psi\rangle=\Sigma_{i=1}^{N}\sqrt{\lambda_i}|a_i\rangle|b_i\rangle$ where $\langle a_i|a_j\rangle=\delta_{ij}$ and $\langle b_i|b_j\rangle=\delta_{ij}$. We further assume that the above form of the Schmidt decomposition of $|\Psi\rangle$ is known to Alice and Bob.

Second, we consider only the most advantageous entanglement manipulation scheme in each step of which Alice keeps track of the results of all her measurements and tells Bob about them, and vice versa. Alice and Bob then update their information on the state they share in each step. Since it is a pure state $|\Psi\rangle$ that Alice and Bob start with, they always deal with a *pure state* in *each step*. Any scheme in which Alice and Bob choose to be sloppy or ignorant can be recast as a situation in which they fail to make full use of their information. Therefore, there is no loss in generality in our consideration [10].

We now argue that any two-way entanglement manipulation strategy for the state $|\Psi\rangle$ can be recast into an equivalent strategy which involves only one-way communications from Alice to Bob—that is to say, a strategy in which Alice performs all the measurements and informs Bob of the outcomes afterwards. This is so because (i) in entanglement manipulations we are mainly concerned with the coefficients of the Schmidt decomposition; and (ii) in *each* step of entanglement manipulation, the Schmidt decomposition of the *pure* state involved is always *symmetric* under the interchange of Alice and Bob. With such symmetry, there is no advantage in having Bob perform the measurement instead of Alice [11–16].

More concretely, consider a round of communications in a two-way scheme of entanglement manipulation. Suppose Alice has performed a measurement on $|\Psi'\rangle=\Sigma_k\sqrt{\lambda'_k}|a'_k\rangle|b'_k\rangle$, and obtained an outcome o_1. She can work out the Schmidt decomposition $P_{o_1}|\Psi'\rangle=\Sigma_k\sqrt{\lambda''_k}|a''_k\rangle|b''_k\rangle$ of the state that she now shares with Bob. Now Alice is supposed to tell Bob the outcome o_1 of her measurement, and Bob then will perform a measurement with a set of local projection operators, say $\{P_l^{Bob}\}$. However, it turns out that there exists a set of local projection operators $\{P_l^{Alice}\}$ by Alice which will do essentially the same trick, as far as entanglement manipulation is concerned. Mathematically, we claim the following proposition.

Proposition 1: Given any pure bipartite state $|\Psi\rangle_{AB}$ shared by Alice and Bob and any complete set of projection operators $\{P_l^{Bob}\}$'s by Bob, there exists a complete set of projection operators $\{P_l^{Alice}\}$'s by Alice and, for each outcome l, a direct product of local unitary transformations $U_l^A\otimes U_l^B$ such that, for each l,

$$(I\otimes P_l^{\mathrm{Bob}})|\Psi\rangle=(U_l^A\otimes U_l^B)(P_l^{\mathrm{Alice}}\otimes I)|\Psi\rangle. \tag{2}$$

Idea of the proof: Consider a pure bipartite state in its Schmidt decomposition

$$|\Psi\rangle=\sum_i\sqrt{\rho_i}|e_i\rangle_A|e_i\rangle_B \tag{3}$$

shared between Alice and Bob. Note that, by the definition of the Schmidt decomposition, it is symmetric under the interchange of $|e_i\rangle_A$ and $|e_i\rangle_B$. This is a mathematical symmetry. Now, in proposition 1, we promote this mathematical symmetry into a physical symmetry between the actions of Alice and Bob in the context of entanglement manipulations. More precisely, if Bob applies a set of projection operators $\{P_l^{Bob}\}$ on his system and obtains an outcome l, the state $|\Psi\rangle$ will be transformed into some state, say

$$|\Psi^B\rangle=\sum_i\sqrt{\mu_i}|a'_i\rangle_A|b'_i\rangle_B. \tag{4}$$

On the other hand, if Alice applies a corresponding set of projector operators $\{P_l^{Alice}\}$ on her system (instead of Bob), then we show that the corresponding outcome l will give her a state

$$|\Psi^A\rangle=\sum_i\sqrt{\mu_i}|a''_i\rangle_A|b''_i\rangle_B, \tag{5}$$

with exactly the same Schmidt coefficients as $|\Psi^B\rangle$. Consequently, there exists a bilocal unitary transformation that will rotate the state $|\Psi^A\rangle$ to $|\Psi^B\rangle$. In this sense, the states $|\Psi^A\rangle$

and $|\Psi^B\rangle$ are equivalent. The upshot is that there is no advantage for Bob to perform a measurement, in place of Alice. In summary, as far as entanglement manipulation of a pure bipartite state is concerned, there is a total symmetry between the actions of Alice and Bob [17].

Proof: See Appendix A.

One can repeat the above argument and prove that all the rounds of measurements can be performed by Alice alone, and Alice only needs to tell Bob her outcomes after the completion of all her measurements. What this means is that, for Alice and Bob manipulating a pure bipartite state, one can, without loss of generality, restrict oneself to schemes of entanglement manipulations using only one-way communications from Alice to Bob.

Finally, it is a well-known consequence of measurement theory that the entire sequence of Alice's measurements can be described as a *single* generalized measurement. [One may argue this well-known result as follows. Every measurement consists of two steps—the interaction of a measuring devise with a system, and the ''reading'' of the measuring device, i.e., a unitary transformation and a projection. Now, any arbitrary sequence of independent measurements can be replaced by an equivalent single measurement, by simply letting all the interactions to be performed first, and reading all the measuring devices simultaneously at the end. In this case one can view all the independent measuring devices as a (more complicated) *single* measuring device, performing a *single* interaction with the measured system (the unitary transformation describing this interaction being simply the product of the unitary transformations describing the individual measuring devices) and followed by a *single* reading stage. Furthermore, even if the measurements are not independent of each other, i.e., some measurements depend on the results of previous measurements, we can still replace the sequence by a single measurement: In this case too, the human observer can postpone ''reading'' the results obtained by the different measuring devices until the end. Indeed, there is no need for the observer to read the results of the measurements in order to tune the subsequent measurements accordingly. The entire process can be realized by the measuring devices interacting with each other as well as with the system under observation. Then, once again, we have a single measuring device, performing a single interaction (except that the interactions between the measuring device and the system contain also some internal interactions between the different parts of the measuring device—corresponding to one part reading the result of the other), and a single reading stage.]

In summary, the most general strategy of entanglement manipulation of a pure bipartite state is equivalent to a strategy involving only a single (generalized) measurement performed by Alice followed by the one-way communications of the result from Alice to Bob (and finally local unitary transformations by Alice and Bob).

B. One-way communications are provably better than no communications

We have shown above that two-way communications are not necessary for the entanglement manipulation of a pure bipartite state—the most general entanglement manipulation strategy can be realized with only one-way communication. A natural question to ask is whether communication is needed at all. We show that, indeed, communication is necessary. This is to say that entanglement manipulation strategies without communication cannot achieve all that could be achieved with communication. The proof of relegated to Appendix B.

In conclusion, more powerful strategies are generally obtained with one-way communications than without communications. On the other hand, we proved in the above paragraphs that one-way communications are sufficient for any strategy. Combining these two results, we conclude that one-way communications are necessary and sufficient for implementing a general strategy of entanglement manipulations of pure bipartite states.

III. OBTAINING A GIVEN MAXIMALLY ENTANGLED STATE Φ_m FROM AN ARBITRARY STATE Ψ

Definition 2 (*m-ME state:* Φ_m): We shall denote by Φ_m a standard *m-dimensional maximally entangled state*

$$|\Phi_m\rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} |i\rangle_A |i\rangle_B, \tag{6}$$

where $|i\rangle_A$'s ($|i\rangle_B$'s, respectively) form an orthonormal basis for a Hilbert space H_A (H_B, respectively). In particular, Φ_1 is a direct product, Φ_2 is (equivalent to) a singlet, and Φ_{2^q} is equivalent to q singlet pairs. In what follows, we shall call Φ_m an *m-ME state*.

We now come to one of the main results of our paper. We consider the following particular problem. Suppose Alice and Bob share a pair of particles in some arbitrary pure state Ψ. By different entanglement manipulations strategies we can transform Ψ into a given m-dimensional maximally entangled state Φ_m. In general such a process does not succeed with certainty but only with some probability p_m. Here we enquire as to the maximal probability with which such a transformation could occur.

Incidentally, one can even obtain a maximally entangled state whose degree of entanglement is *greater* than that of the initial state. Since the average degree of entanglement cannot increase, it is obvious that such a transformation occurs with a probability less than 1. To describe such a situation we sometime use the term ''gambling with entanglement''—indeed, Alice and Bob try to achieve a better than average outcome while taking the risk of losing entanglement if the result turns out to be unfavorable.

Definition 3 (p_m^{MAX}): For any positive integer *m*, we define p_m^{MAX} [18] to be the supremum over all manipulation strategies of the probability p_m of obtaining an *m* state Φ_m from a pair initially in the state Ψ.

We will prove the following theorem.

Main theorem: If we write the initial state Ψ in an ordered Schmidt decomposition as $\Psi = \Sigma_{i=1}^{N} \sqrt{\lambda_i} |a_i\rangle |b_i\rangle$, the supremum probability p_m^{MAX} of obtaining Φ_m over all possible entanglement manipulation strategies is given by the following.

(i) If $m>N$ (N being the number of terms in the Schmidt decomposition of Ψ), then $p_m^{MAX}=0$.

(ii) If $m\leqslant N$, then

$$p_m^{MAX}=\min_{1\leqslant r\leqslant m}\frac{m}{r}(\lambda_{m-r+1}+\lambda_{m-r+2}+\cdots+\lambda_N). \quad (7)$$

The proof of this main theorem is divided into two parts. In Sec. IV, we will derive an upper bound on p_m^{MAX} (see theorem 1). In Sec. V, we demonstrate an explicit strategy that saturates the bound and is, thus, optimal (see theorem 2).

IV. UPPER BOUND ON P_M^{MAX}: THEOREM 1

Theorem 1: If we write the initial state Ψ in an ordered Schmidt decomposition as $\Psi=\Sigma_{i=1}^{N}\sqrt{\lambda_i}|a_i\rangle|b_i\rangle$, the supremum probability p_m^{MAX} of obtaining Φ_m over all possible entanglement manipulation strategies satisfies the following.

i) If $m>N$ (N being the number of terms in the Schmidt decomposition of Ψ), then $p_m^{MAX}=0$.

ii) If $m\leqslant N$, then

$$p_m^{MAX}\leqslant\min_{1\leqslant r\leqslant m}\frac{m}{r}(\lambda_{m-r+1}+\lambda_{m-r+2}+\cdots+\lambda_N). \quad (8)$$

A. The number of Schmidt decomposition terms can never increase: Part (i) of theorem 1

The following lemma is useful.

Lemma 1: The number of terms in a Schmidt decomposition can *never* increase under local measurements and classical communications [19].

Proof: Let us suppose that the initial state $|\Phi\rangle=\Sigma_{i=1}^{N}\sqrt{\lambda_i}|a_i\rangle|b_i\rangle$ has only N nonvanishing terms in its Schmidt decomposition. For each measurement outcome l on $|\Phi\rangle$, the resulting state $P_l^{Alice}|\Phi\rangle=\Sigma_{i=1}^{N}\sqrt{\lambda_i}|a_i^l\rangle|b_i\rangle$ (where $|a_i^l\rangle$ is the projected state $P_l^{Alice}|a_i\rangle$) can be expressed as a sum of N terms. Consequently, its Schmidt decomposition must have at most N terms. QED.

Proof of part (i) of theorem 1: As a corollary of lemma 1, for an initial state $|\Phi\rangle=\Sigma_{i=1}^{N}\sqrt{\lambda_i}|a_i\rangle|b_i\rangle$ with only N nonvanishing terms in its Schmidt decomposition, $p_m^{MAX}=0$ if $m>N$. QED.

This leads to the following apparent paradox. Suppose Alice and Bob share s standard singlets. What is the probability that they can gamble successfully and obtain S ($>s$) singlets? Naively, one might expect the probability to be nonzero: One may use quantum data dilution [6] to dilute s standard singlets into say S pairs of $|\Phi\rangle$ each of entanglement $E(|\Phi\rangle)=s/S$, and then apply the Procrustean (i.e., local filtering) method [6] of entanglement gambling to each of S pairs of $|\Phi\rangle$. For each $|\Phi\rangle$, the Procrustean method gives a nonzero probability, say p', of obtaining a maximally entangled pair out of it. Thus, it looks as if there would be a nonzero probability $(p')^S$ of obtaining S singlets from s singlets. But, as we have seen above, this argument is erroneous—the probability of obtaining S singlets out of gambling with s singlets is strictly zero. The reason is that quantum data dilution is an *inexact* process which is valid only on average.

B. An upper bound on p_m^{MAX}: Part (ii) of theorem 1

It remains to prove part (ii) of theorem 1. It is convenient to introduce the following notation.

Notation (B_r^m): We denote the rth bound in theorem 1 by B_r^m. i.e.,

$$B_r^m\equiv\frac{m}{r}(\lambda_{m-r+1}+\lambda_{m-r+2}+\cdots+\lambda_N). \quad (9)$$

Restatement of part (ii) of theorem 1: Given a state $|\Psi\rangle$ with the ordered Schmidt decomposition $|\Psi\rangle=\Sigma_{i=1}^{N}\sqrt{\lambda_i}|a_i\rangle|b_i\rangle$, the supremum probability p_m^{MAX} of obtaining an m-ME state out of manipulating $|\Psi\rangle$ satisfies a set of constraints $p_m^{MAX}\leqslant B_r^m$ for $1\leqslant r\leqslant m$.

Motivation for the proof of part (ii) of theorem 1: For a fixed r, if the right-hand side of Eq. (9) is zero, then there are only $m-r$ terms in the Schmidt decomposition of $|\Psi\rangle$. From lemma 1, Alice will definitely fail to obtain an m-dimensional maximally entangled pair state because there will be at most $m-r$ terms in the Schmidt decomposition of the resulting state. In the proof, we would like to turn this argument around to show the following. If Alice does succeed, the remaining r [i.e., from ($m-r+1$th to mth) terms in the maximally entangled state must have come from the remaining [i.e., from ($m-r+1$th to Nth)] terms of the Schmidt decomposition of the original state $|\Phi\rangle$. (Surprisingly, classical reasoning is, in fact, valid here. This is because when one considers the reduced density matrix of Alice, Bob's system provides a ''record'' for its history. Therefore, no interference effect is possible. See below.) Let us multiply both sides of the inequality and consider the inequality $rp_m^{MAX}/m\leqslant rB_r^m/m$. Now the left-hand side of the new inequality is simply the probability that Alice's state is projected into the remaining r terms. (There is a supremum probability p_m^{MAX} of successfully obtaining an m-dimensional maximally entangled state. Now, such a state gives a totally random density matrix for Alice's subsystem, which has a support of m dimensions. Moreover, given a fixed r-dimensional subspace of the m-dimensional space in the support of Alice's system, consider the projection operators into that subspace and into its complement, respectively. There is a conditional probability r/m of the random state being projected into the r-dimensional space, rather than into its complement.) It must therefore be constrained by the probability of Bob's system being projected into the space spanned by the $(m-r+1)$th to Nth terms in $|\Phi\rangle$, which is given by the right-hand side.

Proof of part (ii) of theorem 1: Given an initial state $|\Phi\rangle$, for $1\leqslant r\leqslant m$, we decompose $|\Phi\rangle=|\Phi_1^r\rangle+|\Phi_2^r\rangle$ where $|\Phi_1^r\rangle=\Sigma_{i=1}^{m-r}\sqrt{\lambda_i}|a_i\rangle|b_i\rangle$. [Define $|\Phi_1^m\rangle=0$.] Alice and Bob now attempt to manipulate $|\Phi\rangle$ into an m-ME state. Alice can divide up the outcomes into two sets: $\{s_1,s_2,\ldots,s_p\}$ (success) and $\{f_1,f_2,\ldots,f_q\}$ (failure). Let us consider a *successful* outcome s_l. Then $P_{s_l}|\Phi\rangle=P_{s_l}|\Phi_1^r\rangle+P_{s_l}|\Phi_2^r\rangle$ is an

m-ME state. Denoting by $\rho_A^{s_l}$ (similarly $\rho_{A,i}^{r,s_l}$ where $i=1$ or 2) the *unnormalized* density matrix $\mathrm{Tr}_B P_{s_l}|\Phi\rangle\langle\Phi|P_{s_l}^\dagger$ (similarly $\mathrm{Tr}_B P_{s_l}|\Phi_i^r\rangle\langle\Phi_i^r|P_{s_l}^\dagger$ where $i=1$ or 2, respectively), we have $\rho_A^{s_l}=\rho_{A,1}^{r,s_l}+\rho_{A,2}^{r,s_l}$.

We emphasize that the interference term arising from $\mathrm{Tr}_B P_{s_l}|\Phi_1\rangle\langle\Phi_2|P_{s_l}^\dagger$ is identically zero. This is because, when one considers the reduced density matrix of Alice, Bob's system provides a ''record'' of its history. In taking the partial trace over Bob's system, all the interference terms disappear. It is very interesting that classical intuition is valid here. This greatly simplifies our discussion.

The supports satisfy $\mathrm{supp}(\rho_{A,1}^{r,s_l})\subset\mathrm{supp}(\rho_A^{s_l})$. Since $\mathrm{supp}(\rho_{A,1}^{r,s_l})$ has a dimension of most $m-r$ and yet $\mathrm{supp}(\rho_A^{s_l})$ has a dimension m ($P_{s_l}|\Phi\rangle$ is an m-ME state), we can pick r orthonormal vectors $|u_1^{s_l}\rangle,|u_2^{s_l}\rangle,\ldots,|u_r^{s_l}\rangle$ in $\mathrm{supp}(\rho_A^{s_l})$ such that $\langle u_i^{s_l}|v\rangle=0$ for all $|v\rangle\in\mathrm{supp}(\rho_{A,1}^{r,s_l})$. Let us define the projection operator $P_{u^{s_l}}^r=\Sigma_{i=1}^r|u_i^{s_l}\rangle\langle u_i^{s_l}|$. From its definition, it is clear that $P_{u^{s_l}}^r\rho_{A,1}^{r,s_l}P_{u^{s_l}}^{\dagger r}=0$. For a fixed but arbitrary strategy of entanglement concentration, let us denote by p_m^{arb} the probability of successfully obtaining an m-ME state. Therefore,

$$
\begin{aligned}
r p_m^{arb}/m &= \mathrm{Tr}_A\left(\sum_{s_l} P_{u^{s_l}}^r\rho_A^{s_l}P_{u^{s_l}}^{\dagger r}\right)\\
&= \mathrm{Tr}_A\left(\sum_{s_l} P_{u^{s_l}}^r\rho_{A,1}^{r,s_l}P_{u^{s_l}}^{\dagger r}\right)+\mathrm{Tr}_A\left(\sum_{s_l} P_{u^{s_l}}^r\rho_{A,2}^{r,s_l}P_{u^{s_l}}^{\dagger r}\right)\\
&= \mathrm{Tr}_A\left(\sum_{s_l} P_{u^{s_l}}^r\rho_{A,2}^{r,s_l}P_{u^{s_l}}^{\dagger r}\right)\\
&= \mathrm{Tr}_A\ \mathrm{Tr}_B\left(\sum_{s_l} P_{u^{s_l}}^r P_{s_l}|\Phi_2^r\rangle\langle\Phi_2^r|P_{s_l}^\dagger P_{u^{s_l}}^{\dagger r}\right)\\
&\leqslant \mathrm{Tr}_A\ \mathrm{Tr}_B|\Phi_2^r\rangle\langle\Phi_2^r| = \lambda_{m-r+1}+\lambda_{m-r+2}+\cdots+\lambda_N\\
&= rB_r^m/m
\end{aligned}
\tag{10}
$$

for $1\leqslant r\leqslant m$. The equality sign in the first line holds because $\rho_A^{s_l}$ is proportional to the identity matrix in an m-dimensional space, and its trace is proportional to its probability of occurring. Since the total probability of success is p_m^{arb} and $P_{u^{s_l}}^r$ projects an m-ME state into an r-dimensional subspace of the m-dimensional space, the probability of this occurring is clearly rp_m^{arb}/m. Now, one takes the supremum over all entanglement manipulation strategies in Eq. (10) to find that $p_m^{MAX}\leqslant\frac{m}{r}(\lambda_{m-r+1}+\lambda_{m-r+2}+\cdots+\lambda_N)=B_r^m$ for $1\leqslant r\leqslant m$. QED.

Example: Consider $N=3$ and $m=2$ for part (ii) of theorem 1. Theorem 1 now states that $p_2^{MAX}\leqslant\min\{2(\lambda_2+\lambda_3),1\}$.

V. OPTIMAL STRATEGY AND VALUE OF P_M^{MAX}: THEOREM 2

Theorem 1 gives an upper bound to the probability p_m^{MAX}. We now prove that an optimal strategy actually saturates this bound.

Theorem 2: Given a state $|\Psi\rangle=\Sigma_{i=1}^N\sqrt{\lambda_i}|a_i\rangle|b_i\rangle$ (where $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_N$) with only N nonvanishing terms in its Schmidt decomposition. There exists a way to convert Ψ into an m-dimensional maximally entangled state with probability $\min_{r\in\{1,2,\ldots,m\}}(m/r)(\lambda_{m-r+1}+\lambda_{m-r+2}+\cdots+\lambda_N)=\min_r B_r^m$.

Proof of theorem 2: Let us separate the proof into two cases: (a) $\min_r B_r^m=1$ and (b) $\min_r B_r^m<1$.

A. Case (a) of theorem 2

Case (a): Let $\min_r B_r^m=1$. We shall prove that for an optimal strategy, the probability of getting an m-ME state is 1.

It is convenient to start with a simple case, namely transforming maximally entangled states into maximally entangled states of lower dimension. We will prove the following.

Lemma 2: There is a way of transforming with probability 1 any maximally entangled state into a maximally entangled state of lower dimension. Consequently, $p_r^{MAX}\leqslant p_s^{MAX}$ if $r\geqslant s\geqslant 1$.

Proof: First consider the cases $r=3$ and $s=2$. (Here we omit the obvious normalization factors.) A maximally three-dimensionally entangled state has a Schmidt decomposition $|u\rangle_{AB}=|1\rangle_A|1\rangle_B+|2\rangle_A|2\rangle_B+|3\rangle_A|3\rangle_B$. We now show that it can be reduced with certainty to a standard singlet $|1\rangle_A|1\rangle_B+|2\rangle_A|2\rangle_B$. Suppose Alice prepares an ancilla in the state $|0\rangle_a$, and evolves the system in such a way that $|0\rangle_a|1\rangle_A\to(|2\rangle_a+|3\rangle_a)|1\rangle_A$, $|0\rangle_a|2\rangle_A\to(|1\rangle_a+|3\rangle_a)|2\rangle_A$, and $|0\rangle_a|3\rangle_A\to(|1\rangle_a+|2\rangle_a)|3\rangle_A$. The entire state will evolve as follows:

$$
\begin{aligned}
|0\rangle_a|u\rangle_{AB} &= |0\rangle_a(|1\rangle_A|1\rangle_B+|2\rangle_A|2\rangle_B+|3\rangle_A|3\rangle_B)\\
&\to |211\rangle_{aAB}+|311\rangle_{aAB}+|122\rangle_{aAB}+|322\rangle_{aAB}\\
&\quad +|133\rangle_{aAB}+|233\rangle_{aAB}\\
&= |1\rangle_a(|22\rangle_{AB}+|33\rangle_{AB})+|2\rangle_a\\
&\quad \times(|11\rangle_{AB}+|33\rangle_{AB})+|3\rangle_a(|11\rangle_{AB}+|22\rangle_{AB}).
\end{aligned}
\tag{11}
$$

Now Alice measures the state of her ancilla, and obtains a singlet shared with Bob. The exact singlet which is obtained depends on the result of Alice's measurement, but it can always be transformed into the standard one $(1/\sqrt{2})(|11\rangle_{AB}+|22\rangle_{AB})$. This can be realized by Alice communicating the result of her measurement to Bob, such that both of them know which singlet has been obtained, and then having both of them perform the appropriate unitary rotations.

A similar proof can be constructed to show that, starting with a k-ME state (a maximally entangled pair of k state particles), Alice and Bob can with probability 1 convert it to a $(k-1)$-ME state [a maximally entangled pair of $(k-1)$-state particles]. See Appendix C for details. QED.

We remark that, using lemma 2, one can convert with probability 1 a maximally entangled state of dimension i into r standard singlets provided that $i\geqslant 2^r$. Just note that, as mentioned above, r standard singlets are equivalent to a single 2^r-dimensional maximally entangled state, and use the above lemma. This simplifies a related discussion made in Ref. [6], and raises the probability of success from about $1-\epsilon$ to 1.

Now we turn to the general case. The first thing to note is that the condition $\min_r B_r^m = 1$ is completely equivalent with the constraint that the square of the largest normalized Schmidt coefficient is smaller or equal to $1/m$. This is because

$$\lambda_{m-r} \leqslant \cdots \leqslant \lambda_1 \leqslant 1/m \tag{12}$$

implies that

$$\lambda_1 + \lambda_2 + \cdots + \lambda_{m-r} \leqslant \frac{1}{m}(m-r). \tag{13}$$

Since $\lambda_1 + \lambda_2 + \cdots + \lambda_N = 1$, we find that

$$\lambda_{m-r+1} + \cdots + \lambda_N \geqslant 1 - \frac{1}{m}(m-r) = \frac{r}{m}, \tag{14}$$

which is equivalent to

$$B_r^m = \frac{m}{r}(\lambda_{m-r+1} + \cdots + \lambda_N) \geqslant 1. \tag{15}$$

Also recall that $B_m^m = 1$. Therefore, we conclude that, if $\lambda_1 \leqslant 1/m$, then $\min_r B_r^m = 1$. Conversely, if $\lambda_1 > 1/m$, $B_{m-1}^m < 1$. Combining these two results, we see that $\min_r B_r^m = 1$ iff $\lambda_1 \leqslant 1/m$.

Idea of the proof of case (a) of theorem 2: Naively, one might proceed by extracting an m-ME state from Ψ iteratively. At each step we could decompose the state Ψ' into $\Psi' = \Psi_1' + \Psi_2'$ such that Ψ_1' is an (unnormalized) m-ME state and Ψ_2' is residual state that, when properly normalized, still satisfies $\min_r B_r^m = 1$. One simple way to ensure that $\min_r B_r^m = 1$ (or $\lambda_1 \leqslant 1/m$) is always satisfied by Ψ_2' (if properly normalized) is to allow *only* the first m Schmidt terms to contribute to Ψ_1' and, therefore, the λ_1 term of Ψ_2' decreases fast enough.

However, this does not quite work as an iterative procedure. The reason is that, at some point of such a procedure, the mth Schmidt coefficient of the state Ψ will become *degenerate* with the $(m+1)$th and possibly other coefficients. In other words, $\lambda_m = \lambda_{m+1}$, etc. Dealing with this problem is one of the major technicalities in the proof. Let us start by making the following definition.

Definition 5 (precursor state): Consider a state of the form

$$|\Psi_{\text{pre}}^{m,p,q}\rangle = \frac{1}{\sqrt{m}}\left[\sum_{j=1}^{m-p} |j\rangle|j\rangle + \sum_{j=m-p+1}^{m+q}\left(\frac{p}{p+q}\right)^{1/2}|j\rangle|j\rangle\right], \tag{16}$$

where $p > 0$ and $q \geqslant 0$. Let us call it a precursor state of an m-ME state.

Remark: Note that the case $q=0$ corresponds to an m-ME state. For $q>0$, a precursor is a coherent sum of an $(m-p)$-ME state and an $(p+q)$-ME-state. The factor $[p/(p+q)]^{1/2}$ in the definition of $|\Psi_{\text{pre}}^{m,p,q}\rangle$ is needed for the following important result.

Lemma 3: A precursor state of an m-ME state can be converted with certainty an m-ME state.

Proof of lemma 3: The proof is essentially a generalization of the proof of lemma 2. See Appendix D.

In our proof, it is convenient to make use of the following definition.

Definition 6 (mth *Schmidt degeneracy number)*: For any pure bipartite state Ψ in an ordered Schmidt decomposition $|\Psi\rangle = \Sigma_{i=1}^N \sqrt{\lambda_i}|i\rangle_A|i\rangle_B$, let us define the mth ($m<N$) Schmidt degeneracy number (or simply the degeneracy number when there is no ambiguity) to be the number of Schmidt coefficients that are degenerate with λ_m.

Proof of case (a) of theorem 2: Consider the entanglement manipulation of a general state $|\Psi\rangle = \Sigma_{i=1}^N \sqrt{\lambda_i}|i\rangle|i\rangle$ satisfying $\min_r B_r^m = 1$. We construct a multistep procedure such that in each step Alice and Bob either (i) obtain a precursor state which, as shown in lemma 3, can readily be reduced with probability 1 to an m-dimensional maximally entangled state; or (ii) obtain a residual state whose (mth) Schmidt degeneracy number is increased by 1, while still obeying the relation $\min_r B_r^m = 1$ when properly normalized.

If Alice and Bob obtain an m-ME state, they have accomplished their task. If they obtain a residual state, they repeat the procedure. Since with each step the residual state increases its degeneracy number by 1, we are certain that in a finite number of steps ($\leqslant N$) either Alice and Bob obtain an m-ME state, or end up with a residual state Φ_N, which, by lemma 2, can subsequently be converted with certainty to Φ_m.

We now describe each step in more detail.

Suppose the initial state in ordered Schmidt decomposition is

$$|\Psi\rangle = \sum_{i=1}^N \sqrt{\lambda_i}|i\rangle_A|i\rangle_B. \tag{17}$$

Suppose further that λ_m is $(p+q)$-fold degenerate, such that

$$\lambda_{m-p+1} = \cdots = \lambda_m = \cdots = \lambda_{m+q}. \tag{18}$$

The decomposition of $|\Psi\rangle$ into a precursor and a residual state is done by the attachment of an ancilla prepared in the state $|0\rangle_a$ and a subsequent measurement by Alice. For $1 \leqslant i \leqslant m-p$, the evolution goes as

$$\sqrt{\lambda_i}|0\rangle_a|i\rangle_A \rightarrow \sqrt{\frac{a}{m}}|1\rangle_a|i\rangle_A + \sqrt{\lambda_i - \frac{a}{m}}|0\rangle_a|i\rangle_A, \tag{19}$$

where $|0\rangle_a$ and $|1\rangle_a$ are orthonormal, and a is the minimal value needed for a new degeneracy to occur in Schmidt coefficients of the residual state $|\Psi_{\text{res}}\rangle$. In fact,

$$a=\min\left(\frac{m(p+q)}{q}(\lambda_{m-p}-\lambda_{m-p+1}),\ \frac{m(p+q)}{p}(\lambda_{m+q}-\lambda_{m+q+1})\right).$$

For $m-p+1\leqslant i\leqslant m+q$, the evolution goes as:

$$\sqrt{\lambda_i}|0\rangle_a|i\rangle_A\rightarrow\sqrt{\left(\frac{a}{m}\right)\left(\frac{p}{p+q}\right)}|1\rangle_a|i\rangle_A+\sqrt{\lambda_i-\left(\frac{a}{m}\right)\left(\frac{p}{p+q}\right)}|0\rangle_a|i\rangle_A. \quad (20)$$

For $m+q+1\leqslant i\leqslant N$, the state is unchanged, i.e.,

$$|0\rangle_a|i\rangle_A\rightarrow|0\rangle_a|i\rangle_A. \quad (21)$$

Hence we find that

$$|0\rangle_a|\Psi\rangle\rightarrow\sqrt{a}|1\rangle_a|\Psi_{\text{pre}}^{m,p,q}\rangle+\sqrt{1-a}|0\rangle_a|\Psi_{\text{res}}\rangle, \quad (22)$$

where

$$|\Psi_{\text{pre}}^{m,p,q}\rangle=\frac{1}{\sqrt{m}}\left(\sum_{i=1}^{m-p}|i\rangle|i\rangle+\sum_{i=m-p+1}^{m+q}\left(\frac{p}{p+q}\right)^{1/2}|i\rangle|i\rangle\right) \quad (23)$$

is the precursor and

$$|\Psi_{\text{res}}\rangle=(1-a)^{-1/2}\left[\sum_{i=1}^{m-p}\sqrt{\lambda_i-\frac{a}{m}}|i\rangle|i\rangle+\sum_{i=m-p+1}^{m+q}\sqrt{\lambda_i-\left(\frac{a}{m}\right)\left(\frac{p}{p+q}\right)}|i\rangle|i\rangle+\sum_{i=m+q+1}^{N}\sqrt{\lambda_i}|i\rangle|i\rangle\right] \quad (24)$$

is the residual state. Now, since

$$a=\min\left(\frac{m(p+q)}{q}(\lambda_{m-p}-\lambda_{m-p+1}),\ \frac{m(p+q)}{p}(\lambda_{m+q}-\lambda_{m+q+1})\right),$$

we have either (1) $\lambda'_{m-p}=\lambda'_{m-p+1}$ or (2) $\lambda'_{m+q}=\lambda'_{m+q+1}$. In other words, a new degeneracy occurs in the Schmidt coefficients.

Now Alice measures the state of the ancilla. If the outcome is 1, she obtains a precursor state which can be converted with certainty to an m-ME-state. If the outcome is 0, she obtains a residual state with its degeneracy number increased by 1.

It is also easy to see that, just like the original state Ψ, the intermediate residual state $|\Psi_{\text{res}}\rangle$ also has the property that $\min_r B_r^m=1$. The final residual state will be totally degenerate and, hence, has the form Φ_N. This multistep method establishes our proof. QED.

Example: Consider $N=3$ and $m=2$ for case (a) of theorem 2. The initial state is $\sqrt{\lambda_1}|11\rangle+\sqrt{\lambda_2}|22\rangle+\sqrt{\lambda_3}|33\rangle$. The requirement that $\min_r B_r^m=1$ is that $\lambda_1\leqslant 1/2$. Case (a) of theorem 2 asserts that Alice and Bob can obtain a singlet with certainty. The extraction procedure in the proof now goes as follows. In the first step, a singlet state is extracted from the first two Schmidt terms until the second and third terms become degenerate. In other words, Alice attaches an ancilla in the state $|0\rangle_a$ to the original system, and applies the transformation

$$\begin{aligned}\sqrt{\lambda_1}|0\rangle_a|1\rangle&\rightarrow\sqrt{\lambda_2-\lambda_3}|1\rangle_a|1\rangle+\sqrt{\lambda_1-(\lambda_2-\lambda_3)}|0\rangle_a|1\rangle,\\ \sqrt{\lambda_2}|0\rangle_a|2\rangle&\rightarrow\sqrt{\lambda_2-\lambda_3}|1\rangle_a|2\rangle+\sqrt{\lambda_3}|0\rangle_a|2\rangle\\ \sqrt{\lambda_3}|0\rangle_a|3\rangle&\rightarrow\sqrt{\lambda_3}|0\rangle_a|3\rangle.\end{aligned} \quad (25)$$

As a consequence, the combined system evolves as

$$|0\rangle_a(\sqrt{\lambda_1}|11\rangle+\sqrt{\lambda_2}|22\rangle+\sqrt{\lambda_3}|33\rangle)\rightarrow\sqrt{\lambda_2-\lambda_3}|1\rangle_a(|11\rangle+|22\rangle)+|0\rangle_a[\sqrt{\lambda_1-(\lambda_2-\lambda_3)}|11\rangle+\sqrt{\lambda_3}|22\rangle+\sqrt{\lambda_3}|33\rangle]. \quad (26)$$

Alice measures her ancilla. An outcome of 1 gives a singlet, whereas an outcome of 0 will require Alice and Bob to execute the second step. In the second step, the precursor state $|11\rangle+(1/\sqrt{2})(|22\rangle+|33\rangle)$ is extracted. To this end Alice lets her ancilla interact again with her original system, and implements the transformation

$$\begin{aligned}\sqrt{\lambda_1-(\lambda_2-\lambda_3)}|0\rangle_a|1\rangle&\rightarrow\sqrt{2}\sqrt{\lambda_1-\lambda_2}|1\rangle_a|1\rangle+\sqrt{\lambda_3-(\lambda_1-\lambda_2)}|0\rangle_a|1\rangle\\ \sqrt{\lambda_3}|0\rangle_a|2\rangle&\rightarrow\sqrt{\lambda_1-\lambda_2}|1\rangle_a|2\rangle+\sqrt{\lambda_3-(\lambda_1-\lambda_2)}|0\rangle_a|2\rangle\\ \sqrt{\lambda_3}|0\rangle_a|3\rangle&\rightarrow\sqrt{\lambda_1-\lambda_2}|1\rangle_a|3\rangle+\sqrt{\lambda_3-(\lambda_1-\lambda_2)}|0\rangle_a|3\rangle\end{aligned} \quad (27)$$

so that the combined system evolves to

$$|0\rangle_a[\sqrt{\lambda_1-(\lambda_2-\lambda_3)}|11\rangle+\sqrt{\lambda_3}|22\rangle+\sqrt{\lambda_3}|33\rangle]\rightarrow\sqrt{\lambda_1-\lambda_2}|1\rangle_a[\sqrt{2}|11\rangle+|22\rangle+|33\rangle]+\sqrt{\lambda_3-(\lambda_1-\lambda_2)}|0\rangle_a[|11\rangle+|22\rangle+|33\rangle]. \quad (28)$$

Alice now measures the ancilla. An outcome of 1 gives a precursor state which can be converted to a singlet with certainty. On the other hand, an outcome of 0 will give a 3 state which, by lemma 2, can also be converted to a singlet with certainty. In summary, Alice and Bob can obtain a singlet with certainty in this example.

B. Properties of B_r^m

1. Lemma 4

Before moving to case (b), let us prove some lemmas. For any initial state $|\Psi\rangle$, the bounds in theorem 1, $B_r^m = (m/r)(\lambda_{m-r+1}+\lambda_{m-r+2}+\cdots+\lambda_N)$, obey the following.

Lemma 4: If $B_{r+1}^m > B_r^m$, then $B_{r+2}^m > B_{r+1}^m$.

Remark: In other words, for a fixed m, consider B_r^m as a function of r. Once it starts to increase, it will continue to do so.

Proof: See Appendix E.

2. Lemma 5

By adding the condition [which is valid for case (b) of theorem 2] that $\min_r B_r^m < 1$, the following lemma can be proven.

Lemma 5: Given $\min_r B_r^m < 1$, there exists a *unique* r_1 such that $B_1^m \geqslant B_2^m \geqslant \cdots \geqslant B_{r_1}^m < B_{r_1+1}^m < \cdots < B_m^m = 1$.

Proof: See Appendix F.

Remark: Since B_r^m is defined to be $(m/r)(\lambda_{m-r+1} + \lambda_{m-r+2}+\cdots+\lambda_N)$, in terms of λ_i's, the conditions that $B_1^m \geqslant B_2^m \geqslant \cdots \geqslant B_{r_1}^m < B_{r_1+1}^m < \cdots < B_m^m = 1$ can be written as the following set of equations:

$$\lambda_{m-1} \leqslant \lambda_m + \lambda_{m+1} + \cdots + \lambda_N,$$

$$\lambda_{m-2} \leqslant \frac{1}{2}(\lambda_{m-1} + \lambda_m + \cdots + \lambda_N)$$

$$\cdots \leqslant \cdots,$$

$$\lambda_{m-r_1+1} \leqslant \frac{1}{(r_1-1)}(\lambda_{m-r_1+2} + \lambda_{m-r_1+3} + \cdots + \lambda_N), \tag{29}$$

$$\lambda_{m-r_1} > \frac{1}{(r_1)}(\lambda_{m-r_1+1} + \lambda_{m-r_1+2} + \cdots + \lambda_N)$$

$$\cdots > \cdots,$$

$$\lambda_1 > \frac{1}{(m-1)}(\lambda_2 + \lambda_3 + \cdots + \lambda_N).$$

Inspired by the above discussion, let us consider the following set of equations.

$$\lambda_{m-1} \leqslant \lambda_m + \lambda_{m+1} + \cdots + \lambda_N,$$

$$\lambda_{m-2} \leqslant \frac{1}{2}(\lambda_{m-1} + \lambda_m + \cdots + \lambda_N),$$

$$\cdots \leqslant \cdots, \tag{30}$$

$$\lambda_{m-r} \leqslant \frac{1}{r}(\lambda_{m-r+1} + \lambda_{m-r+2} + \cdots + \lambda_N)$$

$$\cdots \leqslant \cdots,$$

$$\lambda_1 \leqslant \frac{1}{(m-1)}(\lambda_2 + \lambda_3 + \cdots + \lambda_N).$$

Consider putting $\lambda_{m-1}, \lambda_{m-2}, \ldots, \lambda_1$ into the left-hand side of Eqs. (30) one by one. We find from Eqs. (29) that $\lambda_{m-1}, \lambda_{m-2}, \ldots, \lambda_{m-r_1+1}$ satisfy Eqs. (30), whereas $\lambda_{m-r_1}, \lambda_{m-r_1-1}, \ldots, \lambda_1$ violate Eqs. (30). Let us focus on the point of first violation, namely, λ_{m-r_1}. We note that the maximal value of $\lambda_{m-r_1}^{\max}$ that will still satisfy Eq. (30) is

$$\lambda_{m-r_1}^{\max} \equiv \frac{1}{r_1}(\lambda_{m-r_1+1} + \cdots + \lambda_N) = \frac{B_{r_1}^m}{m} = \min_r B_r^m. \tag{31}$$

With lemmas 4 and 5 proven, we now return to the proof of case (b) of theorem 2.

C. Case (b) of theorem 2

Case (b): $\min_r B_r^m < 1$.

Idea of our proof: We construct an explicit strategy which saturates the bound $p_m = \min_r B_r^m$ as follows. By attaching an ancilla prepared in the state $|0\rangle_a$ to the system $|\Psi\rangle$, Alice divides up $|\Psi\rangle$ into two pieces—successful and failing pieces—by the evolution

$$|0\rangle_a|\Psi\rangle = |1\rangle_a|\Psi_s\rangle + |0\rangle_a|\Psi_f\rangle, \tag{32}$$

where $|0\rangle_a$ and $|1\rangle_a$ are orthonormal states of the ancilla, $|\Psi_s\rangle$ [when properly normalized belongs to case (a), i.e., $\min_r B_r^m = 1$ and hence] gives a probability 1 of success, and a state $|\Psi_f\rangle$ (has less than m terms in its Schmidt decomposition and hence) gives a probability 0 of success. Alice now reads off the state of the ancilla. A state $|1\rangle_a$ indicates a success and $|0\rangle_a$ a failure. One can then read off the probability of success of this explicit strategy from the norm of $|\Psi_s\rangle$. It turns out to be equal to $\min_r B_r^m$.

Proof of case (b) of theorem 2: Recall from Eq. (31) that the maximal acceptable value of the $(m-r_1)$th squared Schmidt coefficient for it to satisfy Eq. (30) is

$$\lambda_{m-r_1}^{\max} = \frac{B_{r_1}^m}{m} = \min_r B_r^m. \tag{33}$$

Now the successful piece $|\Psi_s\rangle$ in Eq. (32) is obtained by trimming the redundant contribution to $\lambda_1, \lambda_2, \ldots, \lambda_{m-r_1}$ from $|\Psi\rangle$. This is done by the attachment of an ancilla prepared in the state $|0\rangle_a$. The evolution goes as

$$\sqrt{\lambda_i}|0\rangle_a|i\rangle_A \rightarrow \sqrt{\lambda_{m-r_1}^{\max}}|1\rangle_a|i\rangle_A + \sqrt{\lambda_i - \lambda_{m-r_1}^{\max}}|0\rangle_a|i\rangle_A \tag{34}$$

for $1 \leqslant i \leqslant m-r_1$. For $m-r_1+1 \leqslant i \leqslant N$, the evolution is

$$\sqrt{\lambda_i}|0\rangle_a|i\rangle_A \rightarrow \sqrt{\lambda_i}|1\rangle_a|i\rangle_A. \tag{35}$$

Alice now reads off the state of her ancilla. We shall argue in the following paragraph that an outcome 0 means that Alice has failed in obtaining an m-ME state, whereas an outcome 1 means that she has succeeded in obtaining a state satisfying $\min_r B_r^m = 1$, which by Sec. V B [i.e., case (a) of theorum 2] can be reduced with certainty to an m-ME state.

If the outcome is 0, the resulting (failing) state $|\Psi_f\rangle$ has unnormalized squared Schmidt coefficients $\lambda_1 - \lambda_{m-r_1}^{\max}, \lambda_2 - \lambda_{m-r_1}^{\max}, \ldots, \lambda_{m-r_1} - \lambda_{m-r_1}^{\max}, 0, \ldots, 0$. Since it has at most $m-r_1$ terms in its Schmidt decomposition, it follows from lemma 1 that it gives a zero probability of obtaining an m-ME state. On the other hand, if the outcome is 1, the *un-normalized* squared Schmidt coefficients of the resulting (successful) state $|\Psi_s\rangle$ are given by $\lambda_{m-r_1}^{\max}, \ldots, \lambda_{m-r_1}^{\max}, \lambda_{m-r_1+1}, \lambda_{m-r_1+2}, \ldots, \lambda_N$. i.e., the first $(m-r_1)$th squared Schmidt coefficients are all replaced by $\lambda_{m-r_1}^{\max}$. By construction $|\Psi_s\rangle$ belongs to case (a) of theorem 2. Therefore, it always succeeds to give an m-ME state. Moreover, using Eq. (31), it has a norm

$$\begin{aligned}(m-r_1)\lambda_{m-r_1}^{\max} + \lambda_{m-r_1+1} + \cdots + \lambda_N \\ = \frac{m}{r_1}(\lambda_{m-r_1+1} + \lambda_{m-r_1+2} + \cdots + \lambda_N) \\ = B_{r_1}^m = \min_r B_r^m. \end{aligned} \tag{36}$$

This proves that our explicit strategy saturates the bound and completes our proof for the case (b) of theorem 2. QED.

Example: Consider $N=3$ and $m=2$ for case (b) of theorem 2. The initial state is $\sqrt{\lambda_1}|11\rangle + \sqrt{\lambda_2}|22\rangle + \sqrt{\lambda_3}|33\rangle$. The requirement $\min_r B_r^m < 1$ here corresponds to $\lambda_2 + \lambda_3 < 1/2$ (i.e., $\lambda_1 > 1/2$.) Now, according to theorem 2, the probability of obtaining a singlet successfully is $2(\lambda_2 + \lambda_3)$. The extraction is done by attaching an ancilla and applying the unitary transformation

$$\begin{aligned}&\sqrt{\lambda_1}|0\rangle_a|1\rangle \to \sqrt{\lambda_2+\lambda_3}|1\rangle_a|1\rangle + \sqrt{\lambda_1-\lambda_2-\lambda_3}|0\rangle_a|1\rangle, \\ &\sqrt{\lambda_2}|0\rangle_a|2\rangle \to \sqrt{\lambda_2}|1\rangle_a|2\rangle, \\ &\sqrt{\lambda_3}|0\rangle_a|3\rangle \to \sqrt{\lambda_3}|1\rangle_a|3\rangle.\end{aligned} \tag{37}$$

As a consequence, the evolution of the combined system is as follows:

$$\begin{aligned}&|0\rangle_a(\sqrt{\lambda_1}|11\rangle + \sqrt{\lambda_2}|22\rangle + \sqrt{\lambda_3}|33\rangle) \\ &\to |1\rangle_a(\sqrt{\lambda_2+\lambda_3}|11\rangle + \sqrt{\lambda_2}|22\rangle + \sqrt{\lambda_3}|33\rangle) \\ &\quad + |0\rangle_a\sqrt{\lambda_1-\lambda_2-\lambda_3}|11\rangle.\end{aligned} \tag{38}$$

Alice can now measure the ancilla. If the outcome is 1, which occurs with a probability $2(\lambda_2+\lambda_3)$, she has reduced the problem to case (a) of theorem 2, and the resulting state can be reduced to a singlet with certainty. On the other hand, if the answer is 0, she now has a product state and has, thus, failed in getting a singlet.

Recall that theorems 1 and 2 combined together are equivalent to our main theorem. Since we have by now proven both theorems 1 and 2, our main theorem has been established.

VI. LAW OF LARGE NUMBERS

In this section, we derive some constraint on the probabilities of having large deviations from the average properties. Consider the question raised in the Abstract and in Sec. I: Can collective measurements defeat the law of large numbers? We now show that the answer is no. That is, suppose Alice and Bob share n pairs of particles, each pair in a state $|\Psi\rangle$ with an entropy of entanglement $E(|\Psi\rangle)$. We shall show in theorem 3 below that the maximal probability of obtaining nK singlets, with $K > E(|\Psi\rangle)$, goes to zero as n goes to infinity.

Once again, we want to emphasize that this result *does not* follow automatically from the fact that *on average* we cannot obtain more than nE singlets. Indeed, an average of nE singlets could conceivably be obtained if with a *non-negligible* probability $p = E/K$ we obtain nK singlets, while with probability $1 - E/K$ we obtain no singlets at all.

Theorem 3: In the entanglement manipulation of n pairs Ψ, the optimal probability (over all possible strategies) of obtaining nK singlets, $p_{2^{nK}}^{MAX}$, tends to 1 (0, respectively) when $K < E(|\Psi\rangle)$ [$K > E(|\Psi\rangle)$, respectively] in the limit $n \to \infty$.

Remark: It can also be shown that, as a function of K, the jump from 0 to 1 in the value of $p_{2^{nK}}^{MAX}$ occurs in a region of width $O(n^{-1/2})$ around $E(|\Psi\rangle)$. We shall skip the proof here.

Proof of theorem 3: That $p_{2^{nK}}^{MAX}$ tends to 1 in the large-n limit when $K < E(|\Psi\rangle)$ follows trivially from Bennett *et al.*'s reversible strategy [6] and from lemma 2. Let us now consider the case $K > E(|\Psi\rangle)$. Here we view the n pairs Ψ as a single pair in state $\tilde{\Psi} = \Psi^n$, by considering all n Alice's (Bob's) particles to form a single (more complex) quantum system. Similarly, the final nK singlet pairs can be viewed as a single pair in a 2^{nK}-dimensionally maximally entangled state. Then the problem of extracting nK singlets from the n pairs Ψ can be rephrased as the problem of extracting an 2^{nK}-dimensionally maximally entangled state from $\tilde{\Psi}$. The maximal probability for success is $p_{2^{nK}}^{MAX}$, which can be bounded by using theorem 1.

Let $\tilde{\lambda}_i$'s represent the squared Schmidt coefficients of $\tilde{\Psi}$; they are also the eigenvalues of Alice's reduced density matrix. Since Alice's reduced density matrix has a product form, (originating from the n pairs $|\Psi\rangle$) its weight must be concentrated on a ''typical'' space of dimension roughly 2^{nE}. [Here we simplify our notation and use E to denote $E(|\Psi\rangle)$. This is essentially the law of large numbers in classical probability theory. Also see the quantum noiseless source coding theorem [20].] Let us pick a value of K_0 such that $K > K_0 > E$. Since $K_0 > E$, given any $\delta > 0$, for a sufficiently large n, we have that $\sum_{i=2^{nK_0}}^{t^n} \tilde{\lambda}_i < \delta$ where t is the number of terms in the Schmidt decomposition of $|\Psi\rangle$. (An ''atypical'' space has a small weight.) Let us apply theorem 1 to the cases $N=t^n$, $m=2^{nK}$ and $m-r+1=2^{nK_0}$. Note that $r/m>1/2$ for a sufficiently large n. Hence, $p_m^{MAX}/2 < r p_m^{MAX}/m \leqslant \Sigma_{i=m-r+1}^{t^n} \tilde{\lambda}_i < \delta$. Substituting $m=2^{nK}$, back, we obtain $p_{2^{nK}}^{MAX} \to 0$ as $n\to\infty$. QED.

In fact, any particular strategy which transforms n copies of the state Ψ into an average of nE singlets gives a singlet number probability distribution similar to that of reversible strategy [6]. This follows immediately from the result in Sec. VII.

VII. SPECIAL STRATEGIES

In the previous sections we were interested in the question of the maximal probability required to transform an arbitrary entangled state Ψ into a given maximally entangled state, say Φ_m (where m is some given fixed dimension). What happens to the original state Ψ in those cases in which the transformation into Φ_m is *not* successful was not important to us. We will now consider special manipulation strategies, such that for *every* outcome the initial state is transformed into some maximally entangled state. (Note that, by extension, we denote direct product states as ''maximally entangled states of dimension 1''). Such a strategy $\mathcal{S}$ can be characterized by the probabilities $p_1(\mathcal{S})$, $p_2(\mathcal{S})$, . . . , with which the initial state Ψ is transformed into Φ_1, Φ_2, . . . , respectively.

A convenient way to describe this probability distribution is to use, instead of the probabilities $p_m(\mathcal{S})$, the ''cumulative probability'' $p_m^{tot}(\mathcal{S})$:

$$p_m^{tot}(\mathcal{S}) = \sum_{k\geqslant m} p_k(\mathcal{S}). \tag{39}$$

In the present section we find an upper bound on the cumulative property for an arbitrary strategy $\mathcal{S}$,

$$p_m^{tot}(\mathcal{S}) \leqslant p_m^{MAX}, \tag{40}$$

where p_m^{MAX} is the supremum probability over all possible strategies to convert Ψ into an m-dimensional maximally entangled state (an m state). Since $p_m(\mathcal{S})$ represents the probability to convert Ψ into an m state by using the particular strategy $\mathcal{S}$, while p_m^{MAX} represents the supremum probability (over all possible strategies) to convert Ψ into an m state, it is obvious that $p_m(\mathcal{S}) \leqslant p_m^{MAX}$. But why should the sum $p_m(\mathcal{S}) + p_{m+1}(\mathcal{S}) + \cdots$ be smaller than p_m^{MAX}?

The reason for the is that, as we showed in lemma 2, a maximally entangled state of dimension k can always be converted, *with certainty*, into a maximally entangled state of smaller dimension m $(m<k)$. Then suppose that Alice and Bob, using strategy $\mathcal{S}$, convert Ψ into a maximally entangled state of dimension k larger than m. They can then, with certainty convert, this state into a maximally entangled state of dimension equal to m. Consequently, by appending this reduction strategy to strategy $\mathcal{S}$, we obtain a new strategy $\mathcal{S}'$ which converts Ψ into an m state with probability $p_m(\mathcal{S}') = \Sigma_{k\geqslant m} p_m(\mathcal{S}) = p_m^{tot}(\mathcal{S})$ (with a zero probability to convert Ψ into maximally entangled states of dimension larger than m). Now, as p_m^{MAX} is the supremum probability (over all possible strategies) of converting Ψ into an m state, we must have, in particular, $p_m^{MAX} \geqslant p_m(\mathcal{S}') = p_m^{tot}(\mathcal{S})$, which proves the bound in Eq. (40).

VIII. NONEXISTENCE OF UNIVERSAL STRATEGY

As shown in Sec. VII, for any strategy $\mathcal{S}$ which transforms an arbitrary state Ψ into different maximally entangled states Φ_m, the cumulative probability p_m^{tot} of obtaining some maximally entangled state of dimension m or larger is bounded by

$$p_m^{tot} \leqslant p_m^{MAX}. \tag{41}$$

We have also seen in Sec. V that for any particular m there exists a strategy which saturates this bound (the strategy which yields Φ_m with probability equal to p_m^{MAX} and Φ_k, $k>m$ with zero probability). The question is whether there exists a ''universal'' strategy $\mathcal{S}^{univ}$ whose cumulative distribution saturates this bound for *all* m's. The reason we call such a strategy universal is that such a strategy, followed by the reduction of some of the final maximally entangled states into maximally entangled states of lower dimension, it could generate any possible distribution consistent with bound (41). However, we shall show that such a universal strategy does not exist.

Proof: We show that a universal strategy generally cannot exist for the case $N=3$ and $m=2$ or 3. Consider

$$|\Psi\rangle = \sqrt{\lambda_1}|11\rangle + \sqrt{\lambda_2}|22\rangle + \sqrt{\lambda_3}|33\rangle, \tag{42}$$

with $p_2^{MAX}=1$ and $\lambda_2+\lambda_3-\lambda_1 \geqslant 0$. Assume, by means of contradiction, that a universal strategy does exist. We shall use projection operators rather than positive-operator-valued measures in our discussion. As noted in Sec. II, there is no loss of generality. Let $P_1, P_2, \ldots, P_r$ be the set of all projection operators by Alice that give some 3-ME state in a particular universal entanglement manipulation strategy. By definition, $(P_1+P_2+\cdots+P_r)|\Psi\rangle$ has a norm p_3^{MAX}. Note that it follows from theorem 2 that $p_3^{MAX}=3\lambda_3$. Since $p_2^{MAX}=1$, it is necessary for a universal strategy that the residual state $|\Psi_r\rangle = (1-P_1-P_2-\cdots-P_r)|\Psi\rangle$ when properly normalized has $p_2^{MAX}=1$. But this requires the squared eigenvalues of the reduced density matrix of $|\Psi_r\rangle$ to satisfy the constraint $\lambda_2'+\lambda_3'-\lambda_1' \geqslant 0$. We shall show that this is generally impossible. The point of our argument is that, as shown by lemma 6 below, the extraction of a 3-ME state will lead to an equal decrease in all three squared eigenvalues (of the reduced density matrix of $|\Psi_r\rangle$). i.e., $\lambda_i' = \lambda_i - p_3^{MAX}/3 = \lambda_i - \lambda_3$. Therefore, unless $\lambda_1=\lambda_2$, the residual state $|\Psi_r\rangle$ has $\lambda_2'+\lambda_3'-\lambda_1' = \lambda_2-\lambda_1 < 0$, thus contradicting the requirement that $p_2^{MAX}(|\Psi_r\rangle)=1$.

In the above proof, we have used the following lemma.

Lemma 6: Consider a state

$$|\Psi\rangle = \sqrt{\lambda_1}|11\rangle + \sqrt{\lambda_2}|22\rangle + \sqrt{\lambda_3}|33\rangle \tag{43}$$

in Schmidt decomposition. Any strategy that extracts a 3-ME state with a probability p from Ψ will lead to an equal decrease in all three eigenvalues of the reduced density matrix of the *unnormalized* residual state. i.e., $\lambda_i' = \lambda_i - p/3$ where the λ_i''s are eigenvalues of the reduced density matrix of the un-normalized residual state.

Proof of lemma 6: See Appendix G.

IX. MIXED STATES

Let us now consider the case when Alice and Bob share a mixed initial state ρ_{ini}. Since ρ_{ini} is impure, one generally cannot write it in terms of a Schmidt decomposition. More importantly, even if ρ_{ini} *happens* to be symmetric under the interchange of Alice and Bob, there is no guarantee that the intermediate states that they obtain during the entanglement manipulation process will respect such a symmetry [21]. Therefore, the symmetry argument emphasized in the earlier part of this paper will no longer be valid. Manipulations of a mixed state using two-way communications are generally more advantageous than a one-way strategy. Indeed, Bennett *et al.* showed that one-way and two-way capacities for purification are provably different [7].

We also proved in Sec. II that, for a pure bipartite state, entanglement manipulation strategies with one-way communications are provably better than no communications. Note that one-way communications are useful for an entanglement manipulation strategy that has a probability of success strictly between 0 and 1, but not for (deterministic) quantum error correction [7]. The role of communications in entanglement manipulations deserves future investigations.

For a mixed state, there are generally four distinct supremum probabilities to consider: p_m^2, $p_m^{A\rightarrow B}$, $p_m^{B\rightarrow A}$, and p_m^0, corresponding to entanglement manipulation schemes with two-way communications, one-way communications from Alice, to Bob, one-way communications from Bob to Alice and no communications, respectively. While simple bounds on the success probability for manipulating mixed states may be derived, many interesting questions remain unanswered. For example, we do not know the value of $p_{2^{nA}}$ in the asymptotic limit $n\rightarrow\infty$ in the region $D_0(\rho) \leqslant A \leqslant E(\rho)$, where $D_0(\rho)$ is the entanglement of distillation (without any classical communications between Alice and Bob). To conclude, we expect the subtle interplay of the concepts of probability, classical communications, collective manipulations and symmetry in the case of mixed states to be even more challenging than the pure state case considered in this paper.

X. OPEN QUESTIONS ON PURE STATES

Even for the case of a pure initial state, many interesting questions remain unsolved. For instance, what is the supremum probability p_m^0 of obtaining an m-ME state without any classical communications? Note that Bennett *et al.*'s reversible strategy [6] (but not the local filtering strategy [6]) is an example of a strategy which does not require any classical communications. It is an open question whether one can do better than Bennett *et al.*'s strategy without any classical communications.

We emphasize that the symmetry that we have found here applies not only to the entanglement concentration, but also to all types of entanglement manipulations including entanglement dilution [6]. For instance, the usual procedure of entanglement dilution via teleportation falls inside our general framework of using a single generalized measurement by Alice, followed by one-way communications of its result to Bob and a subsequent unitary transformation by Bob. A more systematic investigation of our formalism in applications other than the entanglement concentration may prove rewarding.

ACKNOWLEDGMENTS

H.-K.L. particularly thanks P. Shor for enlightening discussions that indirectly inspired this line of research. Our proof of Theorem 1 was simplified following a critical comment by R. Jozsa. S. P. thanks C. H. Bennett and J. Smolin for helpful communications on their independent proof of lemma 1. Useful discussions with R. Cleve, D. Gottesman, D. Leung, M. A. Nielson, and J. Preskill are greatly appreciated. Part of this paper was written during a visit of H.-K.L. to the Quantum Information and Computing (QUIC) Institute at Caltech, whose hospitality is gratefully acknowledged. This research was done while H.-K.L. was at Hewlett-Packard Labs, Bristol, UK and while S. Popescu was at the Isaac Newton Institute, Cambridge and BRIMS, Hewlett-Packard Labs, Bristol, UK.

APPENDIX A: PROOF OF PROPOSITION 1

Let us write Ψ in its Schmidt decomposition:

$$|\Psi\rangle = \sum_k \sqrt{\lambda_k}|a_k\rangle|b_k\rangle. \tag{A1}$$

Consider any of Bob's projection operator

$$P_l^{\mathrm{Bob}} = \sum_{i,j} m_{ij}^l |b_i\rangle\langle b_j|. \tag{A2}$$

After the projection, the state he shares with Alice becomes

$$|\Psi^B\rangle = (I\otimes P_l^{\mathrm{Bob}})|\Psi\rangle = \sum_{i,k} \sqrt{\lambda_k} m_{ik}^l |a_k\rangle|b_i\rangle. \tag{A3}$$

On the other hand, if, instead of Bob, Alice performs a measurement using the corresponding operator defined by

$$P_l^{\mathrm{Alice}} = \sum_{i,j} m_{ij}^l |a_i\rangle\langle a_j|, \tag{A4}$$

an outcome l will give the state

$$|\Psi^A\rangle = (P_l^{\mathrm{Alice}}\otimes I)|\Psi\rangle = \sum_{i,k} \sqrt{\lambda_k} m_{ik}^l |a_i\rangle|b_k\rangle. \tag{A5}$$

Let us consider unitary transformations U ($|a_i\rangle \rightarrow \Sigma_p u_{ip}|a_p\rangle$) and V ($|b_k\rangle \rightarrow \Sigma_q v_{kq}|b_q\rangle$) that will put Ψ^A in a Schmidt decomposition. i.e.,

$$(U\otimes V)|\Psi^A\rangle=\sum_p \sqrt{\mu_p}|a_p\rangle|b_p\rangle. \tag{A6}$$

From the definitions of U and V and Eqs. (A5) and (A6), we find that

$$\sum_{ik} \sqrt{\lambda_k}m^l_{ik}u_{ip}v_{kq}=\sqrt{\mu_p}\delta_{pq}. \tag{A7}$$

Now consider $(V\otimes U)|\Psi^B\rangle$,

$$(V\otimes U)|\Psi^B\rangle=\sum_{ik}\sum_{pq} \sqrt{\lambda_k}m^l_{ik}v_{kq}u_{ip}|a_q\rangle|b_p\rangle$$
$$=\sum_{pq} \sqrt{\mu_p}\delta_{pq}|a_q\rangle|b_p\rangle=\sum_p \sqrt{\mu_p}|a_p\rangle|b_p\rangle, \tag{A8}$$

where Eq. (A7) is used in the second equality. From Eqs. (A6) and (A8), we find that

$$(V\otimes U)|\Psi^B\rangle=(U\otimes V)|\Psi^A\rangle,$$
$$|\Psi^B\rangle=(V^{-1}U\otimes U^{-1}V)|\Psi^A\rangle, \tag{A9}$$
$$(I\otimes P^{\mathrm{Bob}}_l)|\Psi\rangle=(U^A_l\otimes U^B_l)(P^{\mathrm{Alice}}_l\otimes I)|\Psi\rangle,$$

where $U^A_l=V^{-1}U$ and $U^B_l=U^{-1}V$ QED.

APPENDIX B: PROOF OF THE NECESSITY OF ONE-WAY COMMUNICATION IN ENTANGLEMENT MANIPULATIONS OF BIPARTITE PURE STATES

Definitions (2) and (3) in the main text are needed for this proof. The basic reason for the necessity of classical communication is that, whenever p^{MAX}_m as defined in the text is strictly less than 1, Bob generally needs Alice's help to figure out whether the entanglement manipulation is successful or not.

Consider the example of $|\Psi\rangle=a|11\rangle+b|22\rangle$ where $a>b>0$. We shall first argue that the supremum probability of obtaining a singlet satisfies $0<p_2^{\mathrm{MAX}}<1$: Since the local filtering strategy in Ref. [6] gives a nonzero probability of getting a singlet, we have $p_2^{MAX}\geqslant p_2^{\text{local filtering}}>0$. Moreover, since the entanglement $E(\Psi)<1$ and the average entanglement cannot increase upon entanglement manipulations, the supremum probability p_2^{MAX} of getting a singlet out of entanglement manipulations is less than 1.

Now consider any strategy that gives $0<p_2<1$. Let us divide up its outcomes into two classes: $\{s_1,s_2,\ldots,s_p\}$ (success) and $\{f_1,f_2,\ldots,f_q\}$ (failure), and denote the *unnormalized* reduced density matrix of Bob for an outcome s_i (f_j) by $\rho^{\mathrm{Bob}}_{s_i}$ ($\rho^{\mathrm{Bob}}_{f_j}$). Since $0<p_2<1$, Bob needs to determine the outcome of the entanglement manipulation by distinguishing with certainty between the two density matrices $\rho^{\mathrm{Bob}}_{\mathrm{success}}=\Sigma_i\rho^{\mathrm{Bob}}_{s_i}$ and $\rho^{\mathrm{Bob}}_{\mathrm{failure}}=\Sigma_j\rho^{\mathrm{Bob}}_{f_j}$. Now the distinguishability of two density matrices can be described by the fidelity [22]

$$F\left(\frac{\rho^{\mathrm{Bob}}_{\mathrm{success}}}{\mathrm{Tr}\rho^{\mathrm{Bob}}_{\mathrm{success}}},\frac{\rho^{\mathrm{Bob}}_{\mathrm{failure}}}{\mathrm{Tr}\rho^{\mathrm{Bob}}_{\mathrm{failure}}}\right).$$

The detailed definition and properties of the fidelity are irrelevant for our discussion. It suffices to note the following fact: In order to show that it is impossible for Bob to distinguish with certainty between the two density matrices without communications from Alice, all we need to prove is that

$$F\left(\frac{\rho^{\mathrm{Bob}}_{\mathrm{success}}}{\mathrm{Tr}\rho^{\mathrm{Bob}}_{\mathrm{success}}},\frac{\rho^{\mathrm{Bob}}_{\mathrm{failure}}}{\mathrm{Tr}\rho^{\mathrm{Bob}}_{\mathrm{failure}}}\right)\neq 0,$$

or, equivalently, the supports of $\rho^{\mathrm{Bob}}_{\mathrm{success}}$ and $\rho^{\mathrm{Bob}}_{\mathrm{failure}}$ are not orthogonal to each other. The proof of this claim is simple: Owing to causality, the density matrix of Bob is conserved throughout Alice's measurement, i.e.,

$$\rho^{\mathrm{Bob}}_{\mathrm{success}}+\rho^{\mathrm{Bob}}_{\mathrm{failure}}=\rho^{\mathrm{Bob}}_{\mathrm{initial}}=a^2|1\rangle\langle 1|+b^2|2\rangle\langle 2|. \tag{B1}$$

Since $\rho^{\mathrm{Bob}}_{\mathrm{initial}}$ has a two-dimensional support, $\rho^{\mathrm{Bob}}_{\mathrm{success}}$ must have a support of at most two dimensions. On the other hand, as $\rho^{\mathrm{Bob}}_{s_i}$ is the reduced density matrix for a singlet, $\rho^{\mathrm{Bob}}_{\mathrm{success}}$, being the sum of $\rho^{\mathrm{Bob}}_{s_i}$'s, must have a support of at least two dimensions. Combining these two statements, $\rho^{\mathrm{Bob}}_{\mathrm{success}}$ has a support of exactly two dimensions. Now that both $\rho^{\mathrm{Bob}}_{\mathrm{initial}}$ and $\rho^{\mathrm{Bob}}_{\mathrm{success}}$ have two-dimensional supports, the support of $\rho^{\mathrm{Bob}}_{\mathrm{failure}}$ must be a subspace of the support of $\rho^{\mathrm{Bob}}_{\mathrm{success}}$. Therefore, we conclude that $\rho^{\mathrm{Bob}}_{\mathrm{success}}$ and $\rho^{\mathrm{Bob}}_{\mathrm{failure}}$ do *not* have orthogonal supports and hence the fidelity

$$F\left(\frac{\rho^{\mathrm{Bob}}_{\mathrm{success}}}{\mathrm{Tr}\rho^{\mathrm{Bob}}_{\mathrm{success}}},\frac{\rho^{\mathrm{Bob}}_{\mathrm{failure}}}{\mathrm{Tr}\rho^{\mathrm{Bob}}_{\mathrm{failure}}}\right)\neq 0.$$

QED

APPENDIX C: SOME DETAILS OF PROOF OF LEMMA 2

As before Alice attaches an ancilla to her system A and the evolution needed now is

$$|0\rangle_a|j\rangle_A\rightarrow\left(\frac{1}{\sqrt{k-1}}\sum_{i=1;i\neq j}^{k}|i\rangle_a\right)|j\rangle_A. \tag{C1}$$

That is, the state $|j\rangle_A$ of the particle remains unchanged, but the ancilla is brought to an equal superposition of all states $|1\rangle_a,\ldots,|k\rangle_a$, with the exception of $|j\rangle_a$. The evolution of the state of the ancilla and the pair can, therefore, be summarized as

$$|0\rangle_a|\Phi_k\rangle=|0\rangle_a\left(\frac{1}{\sqrt{k}}\sum_{j=1}^{k}|j\rangle_A|j\rangle_B\right)$$
$$\rightarrow\frac{1}{\sqrt{k}}\sum_{i=1}^{k}|i\rangle_a\left(\frac{1}{\sqrt{k-1}}\sum_{j=1;j\neq i}^{k}|j\rangle_A|j\rangle_B\right), \tag{C2}$$

i.e., each state $|i\rangle_a$ of the ancilla is correlated with a different $(k-1)$-dimensional maximally entangled state.

Next, Alice measures the state of her ancilla. No matter what result she obtains, the pair of particles is left in a $(k-1)$-dimensional maximally entangled state. Which particular state is obtained will depend on Alice's result. Suppose Alice finds the ancilla in the state $|i_0\rangle_a$. Then the pair is in the state $(1/\sqrt{k-1})\Sigma_{j=1;j\neq i_0}^{k}|j\rangle_A|j\rangle_B$. If they wish, Alice and Bob can now convert this state into the standard $(k-1)$-dimensional maximally entangled state $(1/\sqrt{k-1})\Sigma_{j=1}^{k-1}|j\rangle_A|j\rangle_B$. This can be realized by Alice communicating to Bob the result of her measurement, such that both of them know which $(k-1)$-dimensional maximally entangled state has been obtained, and then having both of them perform appropriate local unitary transformations of their particles.

Now, starting with a maximally entangled r-dimensional state, one can repeat our argument to reduce it to a maximally entangled $(r-1)$-dimensional state, a $(r-2)$-dimensional state, etc., until we obtain an s-dimensional state. This shows that any maximally entangled state can be reduced to one with a lower dimension.

APPENDIX D: PROOF OF LEMMA 3

Since $|\Psi_{\rm pre}^{m,p,0}\rangle$ is an m-ME state, all we need to show is the reduction with certainty from $|\Psi_{\rm pre}^{m,p,q}\rangle$ to $|\Psi_{\rm pre}^{m,p,q-1}\rangle$ whenever $q\geqslant 1$. The proof here is analogous to that of lemma 2.

Suppose Alice attaches an ancilla to her system and evolves them in the following manner:

$$|0\rangle_a|j\rangle_A\rightarrow\left(\frac{1}{\sqrt{p+q}}\sum_{i=1}^{p+q}|i\rangle_a\right)|j\rangle_A, \quad \text{for } 1\leqslant j\leqslant m-p,$$
$$|0\rangle_a|j\rangle_A\rightarrow\left(\frac{1}{\sqrt{p+q-1}}\sum_{i=1;i\neq j-(m-p)}^{p+q}|i\rangle_a\right)|j\rangle_A, \quad \text{for } m-p+1\leqslant j\leqslant m+q. \tag{D1}$$

In words, the ancilla is brought to an equal superposition of all states $|1\rangle_a,\ldots,|p+q\rangle_a$ if the state of Alice's system is $|j\rangle_A$ where $1\leqslant j\leqslant m-p$. However, when Alice's system is in $|j\rangle_A$ where $m-p+1\leqslant j\leqslant m+q$, the ancilla is brought to an equal superposition of all states $|1\rangle_a,\ldots,|p+q\rangle_a$ with the exception of $|j-(m-p)\rangle_a$. Upon measuring the state of the ancilla and applying local unitary transformations to their respective systems, Alice and Bob end up in a new precursor $|\Psi_{\rm pre}^{m,p,q-1}\rangle$. This proves the reduction from $|\Psi_{\rm pre}^{m,p,q}\rangle$ to $|\Psi_{\rm pre}^{m,p,q-1}\rangle$. By repeating this reduction process, one can, with certainty, reach $|\Psi_{\rm pre}^{m,p,0}\rangle$ which is an m-ME state.

APPENDIX E: PROOF OF LEMMA 4

It is convenient here to define $S_{m-r+1}=\Sigma_{i=m-r+1}^{N}\lambda_i$. Then,

$$\begin{aligned}
B_{r+1}^m &> B_r^m\\
\frac{m}{r+1}[S_{m-r+1}+\lambda_{m-r}] &> \frac{m}{r}S_{m-r+1}\\
rS_{m-r+1}+r\lambda_{m-r} &> (r+1)S_{m-r+1}\\
r\lambda_{m-r} &> S_{m-r+1}.
\end{aligned} \tag{E1}$$

Now,

$$\begin{aligned}
B_{r+2}^m &= \frac{m}{(r+2)}[S_{m-r+1}+\lambda_{m-r}+\lambda_{m-r-1}]\\
&\geqslant \frac{m}{(r+2)}[S_{m-r+1}+2\lambda_{m-r}]\\
&= \frac{m}{(r+2)(r+1)}[(r+1)S_{m-r+1}+2(r+1)\lambda_{m-r}]\\
&= \frac{m}{(r+2)(r+1)}[(r+1)S_{m-r+1}+r\lambda_{m-r}\\
&\quad +(r+2)\lambda_{m-r}]\\
&> \frac{m}{(r+2)(r+1)}[(r+1)S_{m-r+1}+S_{m-r+1}\\
&\quad +(r+2)\lambda_{m-r}]\\
&= \frac{m}{(r+2)(r+1)}[(r+2)S_{m-r+1}+(r+2)\lambda_{m-r}]\\
&= \frac{m}{(r+1)}[S_{m-r+1}+\lambda_{m-r}]=B_{r+1}^m,
\end{aligned} \tag{E2}$$

where Eq. (E1) is used in obtaining the fifth line. QED.

APPENDIX F: PROOF OF LEMMA 5

Let us consider the list of values of $B_1^m,B_2^m,\cdots,B_m^m$. Since $B_m^m=1>\min_r B_r^m$, as a function of r, B_r^m must start to increase at some point. i.e., there exists r_0 such that $B_{r_0+1}^m>B_{r_0}^m$. But then, by lemma 4, $B_{r_0+2}^m>B_{r_0+1}^m$, $B_{r_0+3}^m>B_{r_0+2}^m$, etc. In words, once B_r^m starts to increase, it will continue to do so. Let us focus on the *last* minimal point of the function B_r^m. i.e., the *largest* value r_1 such that $B_{r_1}^m=\min_r B_r^m$. By definition, $B_{r_1+1}^m>B_{r_1}^m$ which, from lemma 4, implies that $B_{r_1}^m<B_{r_1+1}^m<\cdots<B_m^m=1$. This completes the first part of the proof.

Moreover, we claim that $B_1^m\geqslant B_2^m\geqslant\cdots\geqslant B_{r_1}^m$. We prove this by contradiction. Assuming the contrary, there exists an $a\leqslant r_1$ such that $B_{a-1}^m<B_a^m$. Then lemma 4 implies that $B_{r_1-1}^m<B_{r_1}^m$, which is impossible because it contradicts the fact that $B_{r_1}^m=\min_r B_r^m$. Combining the results of the above two paragraphs, we conclude that $B_1^m\geqslant B_2^m\geqslant\cdots\geqslant B_{r_1}^m<B_{r_1+1}^m<\cdots<B_m^m=1$. QED.

APPENDIX G: PROOF OF LEMMA 6

The following proves the claim in lemma 6 that $\lambda_i' = \lambda_i - p/3$. For simplicity, we shall use projection operators rather than POVMs. As noted in Sec. II, there is no loss in generality. Let $P_1, P_2, \ldots, P_r$ be the set of projection operators for extracting some 3-ME state from Ψ.

Now suppose P gives a 3-ME state with a probability α,

$$|\Psi\rangle = P|\Psi\rangle + (1-P)|\Psi\rangle, \tag{G1}$$

with

$$P|\Psi\rangle = (\sqrt{\lambda_1}P|1\rangle)|1\rangle + (\sqrt{\lambda_2}P|2\rangle)|2\rangle + (\sqrt{\lambda_3}P|3\rangle)|3\rangle. \tag{G2}$$

Since $P|\Psi\rangle$ is 3-ME state with a norm α, its reduced density matrix for B is

$$\rho_B = \sum_{i=1}^{3} \frac{\alpha}{3} |i\rangle\langle i|. \tag{G3}$$

Equating this with the partial trace of $P|\Psi\rangle\langle\Psi|P$ over H_A, we find that the $(\sqrt{\lambda_i}/\sqrt{\alpha/3})P|i\rangle$'s form an orthonormal set. The residual state is

$$(1-P)|\Psi\rangle = \sqrt{\lambda_1 - \frac{\alpha}{3}}|1''1\rangle + \sqrt{\lambda_2 - \frac{\alpha}{3}}|2''2\rangle + \sqrt{\lambda_3 - \frac{\alpha}{3}}|3''3\rangle. \tag{G4}$$

Notice that the $|i''\rangle$'s are orthonormal because

$$\langle j|(1-P)(1-P)|i\rangle = \langle j|(1-2P+PP)|i\rangle = \langle j|(1-2PP + PP)|i\rangle = \langle j|(1-PP)|i\rangle = 0. \tag{G5}$$

Here the last equality follows from the fact that the $P|i\rangle$'s are orthogonal to one another. This shows that an extraction of a 3-ME state of probability α leads to a decrease of each λ's by $\alpha/3$. The same argument can be applied to each of $P = P_1, P_2, \ldots, P_r$. This shows that after the extraction with a probability p of a 3-ME state from Ψ, the eigenvalues of the reduced density matrix of the unnormalized residual state satisfy $\lambda_i' = \lambda_i - p/3$. QED.

[1] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).

[2] J. S. Bell, Physics (Long Island City, N.Y.) **1**, 195 (1964); J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969). These two papers can also be found in J. A. Wheeler and W. H. Zurek, *Quantum Theory and Measurement* (Princeton University Press, Princeton, 1983), pp. 403 and 409, respectively.

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

[4] C. H. Bennett *et al.*, Phys. Rev. Lett. **70**, 1895 (1993).

[5] H. Buhrman, R. Cleve, and A. Wigderson, in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC' 98)* (ACM Press, New York, 1999), p. 63.

[6] C. H. Bennett *et al.*, Phys. Rev. A **53**, 2046 (1996).

[7] C. H. Bennett *et al.*, Phys. Rev. A **54**, 3824 (1996).

[8] S. Popescu and D. Rohrlich, Phys. Rev. A **56**, R3319 (1997).

[9] See, for example, the Appendix of L. P. Hughston, R. Jozsa, and W. K. Wootters, Phys. Lett. A **183**, 14 (1993).

[10] A more precise discussion goes as follows. Suppose the goal is to transform an initial state Ψ into Υ_1. By keeping track of the results of all her measurement, Alice will generally obtain the state of a larger system $\Upsilon = \Upsilon_1 \otimes \Upsilon_2$. However, she can easily reduce its Schmidt coefficients to those of Υ_1 simply by performing a complete measurement along the second system.

[11] This symmetry is originally defined only on the subspace of the Hilbert space with $\lambda_k \neq 0$, but it can be trivially extended to the whole Hilbert space by pairing, in the subspace where $\lambda_k = 0$, *any* orthonormal basis $|a_k''\rangle$'s of H_A with *any* orthonormal basis $|b_k''\rangle$'s of H_B.

[12] This interchange symmetry is reminiscent of the symmetry in two-party cryptographic protocols discussed by, for example, J. Kilian, in *Proceedings of the 20 th Annual Symposium on the Theory of Computing* (ACM, New York, 1988), p. 20. The potential relevance of this interchange symmetry in quantum two-party protocols was speculated by Mayers [13] in the discussion of the impossibility of unconditionally secure quantum bit commitment [13–16].

[13] D. Mayers, Phys. Rev. Lett. **78**, 3414 (1997).

[14] H.-K. Lo and H. F. Chau, Phys. Rev. Lett. **78**, 3410 (1997).

[15] H.-K. Lo and H. F. Chau, Physica D **120**, 177 (1998).

[16] D. Mayers, e-print quant-ph/9603015.

[17] This equivalence (or invariance) between the outcomes of Alice and Bob's local experiments is easy to understand in the case where Alice and Bob share no initial entanglement. In this case, consider, for instance, Bob prepares a spin-1/2 object in his own laboratory along the *z* axis and then measures its spin along the *x* axis. The outcome of this simple experiment is, of course, equally probable to be up or down. Such an experiment by Bob can be mapped into an experiment by Alice in which she prepares a spin-1/2 object in her own laboratory along the *z* axis, and then measures its spin along the *x* axis. Just like Bob's experiment, Alice's experiment also gives equiprobable outcomes. In this sense, the two experiments are equivalent. On the contrary, suppose that Alice, but not Bob, shares some initial entanglement with Charles. Alice can then teleport states to and from with Charles, whereas Bob cannot. It is then clear that Alice's local experiments (plus classical communications) are not generally equivalent to those of Bob. In conclusion, entanglement with a third party generally destroys the equivalence of local experiments between two observers. In this paper, we show, however, that two persons, Alice and

Bob, sharing a pure entangled initial state still respect the equivalence in local experiments. This observation, which greatly simplifies our analysis, is not *a priori* obvious. Note that this equivalence is used here to prove that two-way communications can be reduced to one-way communications in the context of entanglement manipulations of a *pure* entangled state. Curiously, another equivalence (symmetry) argument was previously used to prove that two-way communications is provably better than one-way communications in entanglement purification of *mixed* states [7]. In our opinion, the power of symmetry arguments in entanglement manipulations remains to be fully explored.

[18] We use the superscript MAX because, as shown in Sec. V, the supremum probability is attainable by the optimal strategy.

[19] This lemma was also proven by other groups such as by C. H. Bennett and J. Smolin (private communications) and M. Nielson (private communications). We thank them for helpful discussions.

[20] B. Schumacher, Phys. Rev. A **51**, 2738 (1995); R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).

[21] It is an interesting open question whether there exists any mixed state that respects an interchange symmetry between Alice and Bob for all strategies of entanglement manipulations. We thank M. A. Nielson for raising this question.

[22] R. Jozsa, J. Mod. Opt. **41**, 2315 (1994).