Concentrating Entanglement by Local Actions – Beyond Mean Values

Hoi-Kwong Lo, Sandu Popescu*
Network Systems Department
HP Laboratories Bristol

HPL-97-122
October, 1997

entanglement, quantum theory, quantum information, quantum computation, foundations of quantum mechanics

Previous investigations of entanglement manipulations have focused on the average properties of the outcomes and little is known about the actual probability distribution. Here we go beyond the average properties. We show that, for a *pure* entangled state shared between two separated persons Alice and Bob, 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 (multi-step 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. One important question is whether coherent manipulations in quantum mechanics can enhance the probability of large deviations from the average behaviour. 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.

PACS Numbers: 03.65.Bz, 42.50.Dv, 89.70.+c

Internal Accession Date Only

*Isaac Newton Institute, Cambridge, United Kingdom

© Copyright Hewlett-Packard Company 1997

I. INTRODUCTION

A. Background

Non-locality of entangled states as in the Einstein-Podolsky-Rosen paradox [1], discovered by J. Bell [2] in 1964, is a hallmark of quantum mechanics. The last two years however witnessed a dramatic change in the approach to entanglement, with the advent of the idea of manipulating entanglement. This paper concerns mainly *pure* states. It has previously been realized that entangled pure states can be transformed from one into another via local actions and classical communication, procedures which do not affect the genuinely nonlocal properties of the different states. In effect entanglement is now viewed as a resource which can be transfered from a system to another, and cast into different forms while obeying certain conservation laws—very much like energy or entropy.

For example, suppose that two remote observers, Alice and Bob, share n pairs of spin 1/2 particles, each pair in a non-maximally 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 [3] 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 $\bar{m} = nE(\Psi)$ of singlets, with $E(\Psi)$ the "entropy of entanglement" [4]. Furthermore, as a consequence of this reversibility property, together with the fact that on average entanglement cannot increase via local actions and classical communications [3,5], it has been shown [6] that this particular entanglement manipulation method yields the maximal possible average number of singlets, and that $E(\Psi)$, the maximal average number of singlets which can be extracted per original pair Ψ, is the *unique* measure of entanglement for Ψ.

However, until now the study of entanglement manipulation was focused only on *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 is valid only on average. Here we want to go beyond average values and ask about the actual distributions. For example, the same average number of singlets, $\bar{m} = nE(\Psi)$ might, in principle, be obtained from very different distributions: In the reversible procedure described in Ref. [3], 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 $\bar{m} = nE(\Psi)$. 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 $\bar{m} = nE(\Psi)$ 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 later 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 would 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 big can they be?

[To put things in the right perspective, we would like to mention that the reversible procedure [3] discussed above, is *not* a procedure in which each pair is processed separately but a collective one—yet, the distribution it yields is essentially Gaussian.]

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 the questions raised in this introduction 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.

Another focus of our investigation is the role of *symmetry* in entanglement manipulations. The symmetry of the Schmidt decomposition is exploited in a crucial manner in deriving the main result 1) below. Our work also underscores the importance of classical communications in entanglement manipulations. In our opinion, in the context of entanglement manipulations (including quantum error correction) the subtle interplay of the concepts of probability, coherent manipulations, symmetry and classical communications deserves further investigations. Our work is a step in this direction.

B. Main Results

Our key results on entanglement manipulations can be summarized as follows:

1) Naively, the most general strategy of entanglement manipulation involves Alice and Bob taking turns in performing all sorts of local actions (local unitary transformations, measurements and attachment of ancillary quantum systems), and exchanging back and forth classical messages. However, we show that in the case of *pure* states, *any* strategy of entanglement manipulation is equivalent to one 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).

The key reason is that the Schmidt decomposition is symmetric under the interchange of Alice and Bob. We prove that this symmetry can be promoted to a symmetry between the actions of Alice and Bob on manipulating a pure state. Unfortunately, such a symmetry does not exist for density matrices.

We will also show in Sec. VI that one-way communications generally give strategies that are more powerful that those without communications. Combining this result with the above reduction from two-way to one-way communications, we conclude that one-way communications is necessary and sufficient for implementing the most general strategy of entanglement manipulations.

Notations and Definitions. To state our next results, we need to introduce some notations and definitions. An arbitrary pure state Ψ can be written in Schmidt decomposition [7]

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

where $\langle a_i|a_j\rangle = \langle b_i|b_j\rangle = \delta_{ij}$. Here we order the Schmidt coefficients λ_i decreasingly, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$.

We shall denote by Φ_m a standard *m-dimensional maximally entangled state*

$$|\Phi_m\rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^m |i\rangle|i\rangle. \tag{2}$$

In particular, Φ_1 is a direct-product, Φ_2 is (equivalent to) a singlet and Φ_{2^r} is equivalent to r singlet pairs. In what follows, we shall call Φ_m an *m-state.* All our following results center around entanglement manipulation schemes which aim to convert an arbitrary pure initial state Ψ, known to Alice and Bob, into an m-state:

2)For any positive integer m, we define p_m^{MAX} [8] to be the supremum over all manipulation strategies of the probability p_m of getting an m-state Φ_m from a pair initially in the state Ψ.

We determine p_m^{MAX} and formulate an explicit strategy which can realize it. In general, for a given initial state Ψ, each m requires a different optimal strategy.

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

This follows from the fact that the number of terms in the Schmidt decomposition never increases under local actions and classical communication [9].

ii) If $m \leq N$, then

$$p_m^{MAX} = min_{1 \leq r \leq m} \frac{m}{r} (\lambda_{m-r+1} + \lambda_{m-r+2} + \cdots + \lambda_N). \tag{3}$$

3) Consider a fixed strategy which transforms Ψ into different maximally entangled states Φ_m with corresponding probabilities p_m. A way to describe the probability distribution is by the *cumulative probability* distribution, i.e., by the total probability p_m^{tot} to obtain some maximally entangled state Φ_k with $k \geq m$

$$p_m^{tot} = \sum_{k \geq m} p_k. \tag{4}$$

We prove that for an arbitrary initial state Ψ,

$$p_m^{tot} \leq p_m^{MAX}. \tag{5}$$

That is, we prove that for any given strategy, p_m^{tot}, the total probability to obtain maximally entangled states of dimension m and larger, is upper bounded by p_m^{MAX}, the maximum over all strategies of the probability of getting an m-state.

4) We define a natural notion of a "universal" strategy for entanglement manipulation for all m's and prove that quantum mechanics forbids the existence of such a strategy.

5) We show that collective manipulations cannot yield large deviations from the law of large numbers. More concretely, suppose Alice and Bob share n pairs of particles with each pair in a state $|\Psi\rangle$. We show that the probability of getting nK singlets with $K > E(|\Psi\rangle)$ tends to zero as $n \to \infty$. In particular this means that *any* strategy which can transform n pairs Ψ into an *average* of nE singlets (the maximal allowed average) yields a singlet number probability distribution very similar to that of the reversible strategy by Bennett *et al.* [3]: Any such strategy yields a cumulative probability distribution roughly equal to 1 (0 respectively) when $K < E(|\Psi\rangle)$ ($K > E(|\Psi\rangle)$ respectively). It can be shown that the jump from 0 to 1 occurs in a region of width $O(n^{-1/2})$ around $E(|\Psi\rangle)$ in both cases.

C. Outline of the Paper

Except for Section 9, we will focus on the case where initially Alice and Bob share a known entangled state that is *pure*. In Section 2, we prove that only a single (generalized) measurement and one-way communications are needed for entanglement manipulations of a pure state. A function p_m^{MAX} is introduced in Section 3. We show, in Section 4, that the number of terms in a Schmidt decomposition never increases under entanglement manipulations. In Section 5, we derive an upper bound on p_m^{MAX}. For any given m, we devise in Section 6 an explicit strategy which saturates the bound. One might wonder about the existence of a "universal" strategy which (in some sense) saturates the bounds for all m's. We show in Section 7 that such a strategy generally does not exist. In Section 8, we show that collective manipulations cannot defeat the law of large numbers. In Section 9, the difficulty in attempts to generalize our results to the case where Alice and Bob initially share a *mixed* state is noted. Open questions on the case of pure states are discussed in Section 10.

II. REDUCTION FROM TWO-WAY TO ONE-WAY COMMUNICATIONS

The most general scheme of entanglement manipulations 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. In this Section, however, we prove that any strategy of entanglement manipulation of a pure 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).

First of all, since it is more convenient to deal with projection operators than positive operator valued measures (POVMs), we include any ancilla (measuring apparatus) in Alice and Bob's quantum machines as well. 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 but initially only N of the coefficients of the Schmidt decomposition [7] are non-zero, i.e., $|\Psi\rangle = \sum_{i=1}^{N} \sqrt{\lambda_i}|a_i\rangle|b_i\rangle$ 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 gambling [10] 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*. Notice that any scheme in which Alice and Bob choose to be sloppy or ignorant can be re-casted as a situation in which they fail to make full use of their information. Therefore, there is no loss in generality in our consideration.

We now argue that any two-way entanglement manipulation strategy for the state $|\Psi\rangle$ can be re-casted 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 (1) in entanglement manipulations we are mainly concerned with the coefficients of the Schmidt decomposition and (2) 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].

More concretely, consider a round of communications in a two-way scheme of entanglement manipulation. Suppose Alice has performed a measurement on $|\Psi'\rangle = \sum_k \sqrt{\lambda'_k}|a'_k\rangle|b'_k\rangle$ and obtained an outcome o_1. She can work out the Schmidt decomposition $P_{o_1}|\Psi'\rangle = \sum_k \sqrt{\lambda''_k}|a''_k\rangle|b''_k\rangle$ 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 $\{P_l^{Bob}\}$. Using the isomorphism $|b''_k\rangle \rightarrow |a''_k\rangle$ in the Schmidt decomposition [16], Alice can map the set $\{P_l^{Bob}\}$ into $\{P_l^{Alice}\}$ which is a set of *local* projection operators by Alice instead. Therefore, Alice can perform the measurement herself and obtain mathematically equivalent outcomes. The two experimental procedures (Bob measures with $\{P_l^{Bob}\}$ vs Alice measures with $\{P_l^{Alice}\}$) are *isomorphic*: They give the *same* set of probabilities for the corresponding outcomes. Moreover, for each outcome l, the resulting states in both cases have the *same* coefficients in Schmidt decomposition. If they like, Alice and Bob can apply a direct product of local unitary transformations to change one state into the other.

If the above is still unclear, let us discuss our reasoning in more detail. For simplicity, let us abuse our notation by writing $P_{o_1}|\Psi'\rangle$ simply as $|\Psi\rangle$ and by dropping the "primes" in λ_k'''s, $|a_k''\rangle$ and $|b_k''\rangle$. Suppose Alice obtains an outcome o_1 for her measurement, her state becomes

$$|\Psi\rangle = \sum_k \sqrt{\lambda_k}|a_k\rangle|b_k\rangle \tag{6}$$

in Schmidt decomposition. Consider any of Bob's projection operator

$$P_l^{\text{Bob}} = \sum_{i,j} m_{ij}^l |b_i\rangle\langle b_j|. \tag{7}$$

After the projection, the state he shared with Alice becomes

$$\begin{aligned} |\Psi^B\rangle &= \left(I \otimes P_l^{\text{Bob}}\right)|\Psi\rangle \\ &= \sum_{i,k} \sqrt{\lambda_k} m_{ik}^l |a_k\rangle|b_i\rangle. \end{aligned} \tag{8}$$

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

$$P_l^{\text{Alice}} = \sum_{i,j} m_{ij}^l |a_i\rangle\langle a_j|, \tag{9}$$

an outcome l will give the state

$$\begin{aligned} |\Psi^A\rangle &= \left(P_l^{\text{Alice}} \otimes I\right)|\Psi\rangle \\ &= \sum_{i,k} \sqrt{\lambda_k} m_{ik}^l |a_i\rangle|b_k\rangle \end{aligned} \tag{10}$$

Notice that the two resulting states corresponding to the outcomes in the two experimental situations (i.e., Alice measures and gets the result l vs Bob measures and gets the result l) are related to each other by the mapping $|a_k\rangle|b_i\rangle$ to $|a_i\rangle|b_k\rangle$. We now argue that this mapping is an isomorphism which preserves Schmidt coefficients. Our point is: this exchange operation can be physically (and also mathematically) realized by interchanging systems H_A and H_B and relabeling the state $|a_k\rangle$'s by $|b_k\rangle$'s and vice versa. Being a simple exchange and relabeling, this operation must, therefore, correspond to a symmetry operation which leaves physics invariant [17]. What we mean by physics here includes the probability amplitude of a state and the ordered set of coefficients of the Schmidt decomposition λ_i's. [If one were able to change the probability amplitude of a state or the coefficients of its Schmidt decomposition by interchanging the two systems and labeling their states, then the probability amplitude and Schmidt coefficients could not carry much physical meaning.] This fact can be verified mathematically: If $|\Psi^B\rangle$ can be put into a Schmidt basis by applying a direct product of local unitary transformations $U^1 \otimes U^2$, then $|\Psi^A\rangle$ can be put into a Schmidt basis with the same set of coefficients by applying $U^2 \otimes U^1$.

Mathematically, we are claiming that, given any pure state $|\Psi\rangle$, for each outcome l, there exists a direct product of local unitary transformations $U_l^A \otimes U_l^B$ such that

$$\left(I \otimes P_l^{\text{Bob}}\right)|\Psi\rangle = \left(U_l^A \otimes U_l^B\right)\left(P_l^{\text{Alice}} \otimes I\right)|\Psi\rangle. \tag{11}$$

In conclusion, the mathematical symmetry of the Schmidt decomposition can be promoted into a *physical* symmetry between the *actions* of Alice and Bob. Consequently, Alice can perform the measurement in each step herself and inform Bob of the result afterwards.

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. Mathematically, we can understand this result as follows: Suppose Alice and Bob go through $2r$ rounds of communications. Up to uninteresting local unitary transformations [18], a branch of history is described by

$$\begin{aligned}&\left(I \otimes P_{i_{2r}}^{\mathrm{B},2r}\right)\left(P_{i_{2r-1}}^{\mathrm{A},2r-1} \otimes I\right)\cdots\\ &\cdots\left(P_{i_3}^{\mathrm{A},3} \otimes I\right)\left(I \otimes P_{i_2}^{\mathrm{B},2}\right)\left(P_{i_1}^{\mathrm{A},1} \otimes I\right)|\Psi\rangle,\end{aligned} \tag{12}$$

where i_j denotes the particular outcome of the measurement in the j-th step of the entanglement manipulation. Applying Eq. (11) to each round of communication from Bob to Alice, we obtain

$$\left(U_{i_{2r}}^{\mathrm{A},2r} \otimes U_{i_{2r}}^{\mathrm{B},2r}\right)\left(P_{i_{2r}}^{\mathrm{A},2r} \otimes I\right)\left(P_{i_{2r-1}}^{\mathrm{A},2r-1} \otimes I\right)\cdots$$

$$\begin{aligned}&\left(P_{i_3}^{\mathrm{A},3}\otimes I\right)\left(U_{i_2}^{\mathrm{A},2}\otimes U_{i_2}^{\mathrm{B},2}\right)\left(P_{i_2}^{\mathrm{A},2}\otimes I\right)\left(P_{i_1}^{\mathrm{A},1}\otimes I\right)|\Psi\rangle\\&=\left(U_{i_{2r}}^{\mathrm{A},2r}P_{i_{2r}}^{\mathrm{A},2r}P_{i_{2r-1}}^{\mathrm{A},2r-1}U_{i_{2(r-1)}}^{\mathrm{A},2(r-1)}\cdots P_{i_3}^{\mathrm{A},3}U_{i_2}^{\mathrm{A},2}P_{i_2}^{\mathrm{A},2}P_{i_1}^{\mathrm{A},1}\right)\\&\otimes\left(U_{i_{2r}}^{\mathrm{B},2r}U_{i_{2(r-1)}}^{\mathrm{B},2(r-1)}\cdots U_{i_2}^{\mathrm{B},2}\right)|\Psi\rangle.\end{aligned}\tag{13}$$

Therefore, we conclude that, for Alice and Bob manipulating with entanglement and starting with a pure 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 from 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, (only 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.

Mathematically, this means that Eq. (13) can be further simplified as

$$\begin{aligned}
& \left(U_{i_{2r}}^{\mathrm{A},2r} P_{i_{2r}}^{\mathrm{A},2r} P_{i_{2r-1}}^{\mathrm{A},2r-1} U_{i_{2(r-1)}}^{\mathrm{A},2(r-1)} \cdots \right. \\
& \left. P_{i_3}^{\mathrm{A},3} U_{i_2}^{\mathrm{A},2} P_{i_2}^{\mathrm{A},2} P_{i_1}^{\mathrm{A},1}\right) \\
& \otimes \left(U_{i_{2r}}^{\mathrm{B},2r} U_{i_{2(r-1)}}^{\mathrm{B},2(r-1)} \cdots U_{i_2}^{\mathrm{B},2}\right) |\Psi\rangle \\
= & \left(U_I^A P_I \otimes U_I^B\right) |\Psi\rangle \\
= & \left(U_I^A \otimes U_I^B\right) \left(P_I \otimes I\right) |\Psi\rangle, \qquad (14)
\end{aligned}$$

where P_I is a projection operator and the index I is a shorthand for the multi-index $i_{2r}, i_{2r-1}, \cdots, i_2, i_1$. The first equality in the above equation holds because (1) for any projection operator P and unitary transformation U, $P' = UPU^\dagger$ is also a projection operator and consequently (2) unitary transformations can be permuted to the left of projection operators.

In summary, the most general strategy of entanglement manipulation of a pure 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). The upshot of the whole analysis is the following: In the case of a known initial pure state, an arbitrary but fixed entanglement manipulation strategy is equivalent to a set of local projection operators $\{P_l^{Alice}\}$ of Alice. This is so because as can be seen from Eq. (14) all we have ignored is just a direct product of local unitary transformations, which in no way affect the interesting physics—the coefficients of the Schmidt decomposition. This projection operator formulation greatly simplifies our following analysis.

III. CUMULATIVE PROBABILITIES

Suppose Alice and Bob share a pair of particles in some arbitrary state Ψ, and that by using some particular strategy $\mathcal{S}$ they convert it into different maximally entangled states of dimension m (m=1,2,...) with corresponding probabilities $p_m(\mathcal{S})$. As mentioned in Sec. I.B, a convenient way to describe this probability distribution is to use instead of the probabilities $p_m(\mathcal{S})$ the "cumulative probability" $p_m^{tot}(\mathcal{S})$,

$$p_m^{tot}(\mathcal{S}) = \sum_{k\geq m} p_k(\mathcal{S}). \tag{15}$$

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

$$p_m^{tot}(\mathcal{S}) \leq p_m^{MAX} \tag{16}$$

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 $p_m(\mathcal{S}) \leq p_m^{MAX}$. But why should the sum $p_m(\mathcal{S}) + p_{m+1}(\mathcal{S}) + ...$ be smaller than p_m^{MAX}?

The reason is that, as we show bellow, 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, by using the strategy $\mathcal{S}$ convert Ψ into a maximally entangled state of dimension k larger than m. They can then convert, with certainty, this state into a maximally entangled state of dimension equal to m. Consequently, by appending this reduction strategy to the strategy $\mathcal{S}$, we obtain a new strategy $\mathcal{S}'$ which converts Ψ into an m-state with probability $p_m(\mathcal{S}') = \sum_{k\geq m} p_m(\mathcal{S}) = p_m^{tot}(\mathcal{S})$, (while having 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 $p_m^{MAX} \geq p_m(\mathcal{S}') = p_m^{tot}(\mathcal{S})$ which proves the bound in Eq. (16). All that remains to prove is the following.

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

Proof: First, consider the case $r = 3$ and $s = 2$. (Here we omit the obvious normalization factors.) A maximally three-dimensionally entangled state has the 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 $|0\rangle_a|1\rangle_A \rightarrow (|2\rangle_a + |3\rangle_a)|1\rangle_A$, $|0\rangle_a|2\rangle_A \rightarrow (|1\rangle_a + |3\rangle_a)|2\rangle_A$, and $|0\rangle_a|3\rangle_A \rightarrow (|1\rangle_a + |2\rangle_a)|3\rangle_A$. The entire state will evolve as follows:

$$\begin{aligned} & |0\rangle_a|u\rangle_{AB} \\ = \ & |0\rangle_a(|1\rangle_A|1\rangle_B + |2\rangle_A|2\rangle_B + |3\rangle_A|3\rangle_B) \\ \rightarrow \ & |211\rangle_{aAB} + |311\rangle_{aAB} + |122\rangle_{aAB} + |322\rangle_{aAB} \\ & +|133\rangle_{aAB} + |233\rangle_{aAB} \\ = \ & |1\rangle_a(|22\rangle_{AB} + |33\rangle_{AB}) + |2\rangle_a(|11\rangle_{AB} + |33\rangle_{AB}) \\ & +|3\rangle_a(|11\rangle_{AB} + |22\rangle_{AB}). \end{aligned} \tag{17}$$

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 $\frac{1}{\sqrt{2}}(|11\rangle_{AB} + |22\rangle_{AB})$. This can be realized by Alice communicating to Bob the result of her measurement, 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-state (a maximally entangled pair of k-state particles), Alice and Bob can with probability 1 convert it to a $(k-1)$-state (maximally entangled pair of $(k-1)$-state particles). As before Alice attaches an ancilla to her system A and the evolution needed now is

$$|0\rangle_a|j\rangle_A \rightarrow \left(\frac{1}{\sqrt{k-1}} \sum_{i=1;i\neq j}^{k} |i\rangle_a\right)|j\rangle_A. \tag{18}$$

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

$$\begin{aligned} |0\rangle_a|\Phi_k\rangle &= |0\rangle_a(\frac{1}{\sqrt{k}}\sum_{j=1}^{k}|j\rangle_A|j\rangle_B) \\ &\rightarrow \frac{1}{\sqrt{k}}\sum_{i=1}^{k}|i\rangle_a(\frac{1}{\sqrt{k-1}}\sum_{j=1;j\neq i}^{k}|j\rangle_A|j\rangle_B). \end{aligned} \tag{19}$$

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 $\frac{1}{\sqrt{k-1}}\sum_{j=1;j\neq i_0}^{k}|j\rangle_A|j\rangle_B$. If they wish, Alice and Bob can now convert this state into the standard (k-1)-dimensional maximally entangled state $\frac{1}{\sqrt{k-1}}\sum_{j=1}^{k-1}|j\rangle_A|j\rangle_B$. 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, $(r-2)$-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. QED.

We remark that using Lemma 1 one can convert with probability 1 a maximally entangled state of dimension i into r standard singlets provided that $i \geq 2^r$. Just note that, as mentioned before, 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. [3] and raises the probability of success from about $1-\epsilon$ to 1.

IV. NON-INCREASING PROPERTIES

Consider the following question. Suppose Alice and Bob share s standard singlets. What is the probability that they can gamble successfully and get S $(> s)$ singlets? Naively, one might expect the probability to be non-zero: One may use quantum data dilution [3] 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 [3] of entanglement gambling to each of S pairs of $|\Phi\rangle$. For each $|\Phi\rangle$, the Procrustean method gives a non-zero probability say, p', of getting a maximally entangled pair out of it. So, it looks as if there would be a non-zero probability $(p')^S$ of getting S singlets from s singlets. As we will see below, this argument is erroneous because quantum data dilution is an *inexact* process which is valid only on average. In contrast, in gambling with entanglement, we are interested in the deviation from average. We will prove that the probability of getting S singlets out of gambling with s singlets is strictly zero. In fact, we can prove a stronger result:

Lemma 2 : The number of terms in a Schmidt decomposition can *never* increase under local measurements and classical communications [9].

Proof: Let us suppose that the initial state $|\Phi\rangle = \sum_{i=1}^{N} \sqrt{\lambda_i} |a_i\rangle |b_i\rangle$ has only N non-vanishing terms in its Schmidt decomposition. For each measurement outcome l on $|\Phi\rangle$, the resulting state $P_l^{Alice}|\Phi\rangle = \sum_{i=1}^{N} \sqrt{\lambda_i} |a_i^l\rangle |b_i\rangle$ [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.

As a corollary, for an initial state $|\Phi\rangle = \sum_{i=1}^{N} \sqrt{\lambda_i} |a_i\rangle |b_i\rangle$ with only N non-vanishing terms in its Schmidt decomposition, $p_m^{MAX} = 0$, if $m > N$. Consequently, the probability that Alice and Bob get S singlets out of gambling with $s(< S)$ singlets (via local operations and classical communications) is exactly zero.

V. CONSTRAINTS ON P_M^{MAX}

Theorem 1: Given a state $|\Psi\rangle = \sum_{i=1}^{N} \sqrt{\lambda_i}|a_i\rangle|b_i\rangle$ (where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$) with only N non-vanishing terms in its Schmidt decomposition. The supremum probability p_m^{MAX} of obtaining an m-dimensional maximally entangled state satisfies a set of constraints $rp_m^{MAX}/m \leq \lambda_{m-r+1} + \lambda_{m-r+2} + \cdots + \lambda_N$ for $1 \leq r \leq m$.

Idea of the proof: For a fixed r, if the right hand side, $\lambda_{m-r+1} + \lambda_{m-r+2} + \cdots + \lambda_N$, is zero, then there are only $m - r$ terms in the Schmidt decomposition of $|\Psi\rangle$. From Lemma 2, Alice will definitely fail to get an m-dimensional maximally entangled pair state because there will be at most $m - r$ terms in the Schmidt decomposition of the resulting state. Turning this argument around, if Alice does succeed, the remaining r (i.e., from $m-r+1$-th to m-th) terms in the maximally entangled state must have come from the remaining (i.e., from $m - r + 1$-th to N-th) terms of the Schmidt decomposition of the original state $|\Phi\rangle$. Now the left hand side of the inequality is simply the probability that Alice's state gets projected into the remaining r terms. [There is a supremum probability p_m^{MAX} of gambling successfully (i.e., getting an m-dimensional maximally entangled state) and a conditional probability r/m of getting projected in an r-dimensional subspace of the m-dimensional space in the support of Alice's system.] It must therefore be constrained by the probability of Bob's system getting projected into the space spanned by the $m-r+1$-th to N-th terms in $|\Phi\rangle$, which is given by the right hand side.

Proof of Theorem 1: Given an initial state $|\Phi\rangle$, for $1 \leq r \leq m$, we decompose $|\Phi\rangle = |\Phi_1^r\rangle + |\Phi_2^r\rangle$ where $|\Phi_1^r\rangle = \sum_{i=1}^{m-r} \sqrt{\lambda_i}|a_i\rangle|b_i\rangle$ [Define $|\Phi_1^m\rangle = 0$.] and $|\Phi_2^r\rangle = \sum_{i=m-r+1}^{N} \sqrt{\lambda_i}|a_i\rangle|b_i\rangle$. [Define $|\Phi_2^r\rangle = 0$ whenever $N < m - r + 1$.] Alice and Bob now gamble with $|\Phi\rangle$ to get an m-state. Alice can divide up the outcomes into two sets: $\{s_1, s_2, \cdots, s_p\}$ (success) and $\{f_1, f_2, \cdots, f_q\}$ (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-state. Denoting by $\rho_A^{s_l}$ (similarly $\rho_{A,i}^{r,s_l}$ where $i = 1$ or 2) the *un-normalized* 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 $\rho_A^{s_l} = \rho_{A,1}^{r,s_l} + \rho_{A,2}^{r,s_l}$ and their supports satisfy $supp(\rho_{A,1}^{r,s_l}) \subset supp(\rho_A^{s_l})$. Since $supp(\rho_{A,1}^{r,s_l})$

has dimension at most $m-r$ and yet $supp(\rho_A^{s_l})$ has dimension m ($P_{s_l}|\Phi\rangle$ is an m-state.), we can pick r orthonormal vectors $|u_1^{s_l}\rangle, |u_2^{s_l}\rangle, \cdots, |u_r^{s_l}\rangle$ in $supp(\rho_A^{s_l})$ such that $\langle u_i^{s_l}|v\rangle = 0$ for all $|v\rangle \in supp(\rho_{A,1}^{r,s_l})$. Let us define the projection operator $P_{u^{s_l}}^r = \sum_{i=1}^r |u_i^{s_l}\rangle\langle u_i^{s_l}|$. 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 (gambling), let us denote by p_m^{arb} the probability of successfully getting an m-state. Therefore,

$$\begin{aligned}
& rp_m^{arb}/m \\
&= \mathrm{Tr}_A \left(\sum_{s_l} P_{u^{s_l}}^r \rho_A^{s_l} P_{u^{s_l}}^{\dagger r} \right) \\
&= \mathrm{Tr}_A \left(\sum_{s_l} P_{u^{s_l}}^r \rho_{A,1}^{r,s_l} P_{u^{s_l}}^{\dagger r} \right) + \mathrm{Tr}_A \left(\sum_{s_l} P_{u^{s_l}}^r \rho_{A,2}^{r,s_l} P_{u^{s_l}}^{\dagger r} \right) \\
&= \mathrm{Tr}_A \left(\sum_{s_l} P_{u^{s_l}}^r \rho_{A,2}^{r,s_l} P_{u^{s_l}}^{\dagger r} \right) \\
&= \mathrm{Tr}_A \ \mathrm{Tr}_B \left(\sum_{s_l} P_{u^{s_l}}^r P_{s_l} |\Phi_2^r\rangle\langle\Phi_2^r| P_{s_l}^\dagger P_{u^{s_l}}^{\dagger r} \right) \\
&\leq \mathrm{Tr}_A \ \mathrm{Tr}_B |\Phi_2^r\rangle\langle\Phi_2^r| \\
&= \lambda_{m-r+1} + \lambda_{m-r+2} + \cdots + \lambda_N,
\end{aligned} \tag{20}$$

for $1 \leq r \leq m$. The equality sign in the second line holds because $\rho_A^{s_l}$ is proportional to the identity matrix in a 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-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 gambling strategies in Eq. (20) to find that $rp_m^{MAX}/m \leq \lambda_{m-r+1} + \lambda_{m-r+2} + \cdots + \lambda_N$ for $1 \leq r \leq m$. QED.

VI. OPTIMAL STRATEGY

A. Theorem 2

Theorem 1 gives an upper bound to the probability p_m^{MAX}. We now prove that an optimal strategy saturates the bound. In other words, we have:

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

Proof of Theorem 2: We simplify our notation by denoting the r-th bound in Theorem 1, $\frac{m}{r}(\lambda_{m-r+1} + \lambda_{m-r+2} + \cdots + \lambda_N)$, by B_r^m. Let us separate the proof into two cases: (a) $\min_r B_r^m = 1$ and (b) $\min_r B_r^m < 1$.

B. Proof of 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-state is 1.

Obviously, if all Schmidt coefficients of Ψ are equal to each other, then Ψ is an N-dimensional maximally entangled state, and by Lemma 1, one can convert it with certainty into an m-dimensional maximally entangled state ($m < N$). As a generalization of lemma 1, we now consider a state of the form

$$|\Psi_{\rm pre}^{m,p,q}\rangle = \frac{1}{\sqrt{m}}\left(\sum_{j=1}^{m-p}|j\rangle|j\rangle + \sum_{j=m-p+1}^{m+q}(\frac{p}{p+q})^{1/2}|j\rangle|j\rangle\right) \tag{21}$$

where $p > 0$ and $q \geq 0$. Let us call it a "precursor" of an m-state. Note that the case $q = 0$ corresponds to an m-state. For $q > 0$, a precursor is a coherent sum of an $m - p$-state and an $(p + q)$-state. The factor $(\frac{p}{p+q})^{1/2}$ in the definition of $|\Psi_{\rm pre}^{m,p,q}\rangle$ is needed for the following important result: A precursor can be converted with certainty into an m-state. Since $|\Psi_{\rm pre}^{m,p,0}\rangle$ is an m-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 $q \geq 1$. Our proof here is analogous to the proof of Lemma 1. Suppose Alice attaches an ancilla to her system and evolves them in the following manner:

$$\begin{aligned} |0\rangle_a|j\rangle_A &\rightarrow (\frac{1}{\sqrt{p+q}}\sum_{i=1}^{p+q}|i\rangle_a)|j\rangle_A, \\ &\quad \text{for } 1 \leq j \leq m-p \\ |0\rangle_a|j\rangle_A &\rightarrow (\frac{1}{\sqrt{p+q-1}}\sum_{i=1;i\neq j-(m-p)}^{p+q}|i\rangle_a)|j\rangle_A, \\ &\quad \text{for } m-p+1 \leq j \leq m+q. \end{aligned} \tag{22}$$

In words, the ancilla is brought to an equal superposition of all states $|1\rangle_a, \cdots, |p+q\rangle_a$ if the state of Alice's system is $|j\rangle_A$ where $1 \leq j \leq m-p$. However, when Alice's system is in $|j\rangle_A$ where $m-p+1 \leq j \leq m+q$, the ancilla is brought to an equal superpositon of all states $|1\rangle_a, \cdots, |p+q\rangle_a$ 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-state.

Let us return to the entanglement manipulation of a general state $|\Psi\rangle = \sum_{i=1}^{N}\sqrt{\lambda_i}|i\rangle|i\rangle$ satisfying $\min_r B_r^m = 1$. The number of coefficients in the Schmidt decomposition that are degenerate with the m-th largest one (i.e., the number of λ_i's such that $\lambda_i = \lambda_m$) will play a pivotal role in the following discussion. Let us call this number the (m-th) "degeneracy number". The idea of our proof of case (a) of theorem 2 is to construct a multi-step procedure such that in each step Alice and Bob either:

i) obtain a precursor which can readily be reduced with probability 1 to an m-dimensional maximally entangled state; or

ii) obtain a residual state whose (m-th) 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-state, they have accomplished their task. If they get 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 ($\leq N$) either Alice and Bob obtain an m-state, or end up with a residual state which is maximally entangled (with dimension larger than or equal to m), which can subsequently be converted with certainty to an m-state.

We now describe each step in more detail. Suppose the initial state is

$$|\Psi\rangle = \sum_{i=1}^{N} \sqrt{\lambda_i} |i\rangle_A |i\rangle_B \tag{23}$$

with the Schmidt coefficients ordered decreasingly. Suppose further that λ_m is $(p+q)$-fold degenerate such that

$$\lambda_{m-p+1} = ... = \lambda_m = ... = \lambda_{m+q}. \tag{24}$$

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 $1 \leq i \leq m-p$, the evolution goes as:

$$\begin{aligned} & \sqrt{\lambda_i}|0\rangle_a|i\rangle_A \\ \to & \sqrt{\frac{a}{m}}|1\rangle_a|i\rangle_A + \sqrt{\lambda_i - \frac{a}{m}}|0\rangle_a|i\rangle_A, \end{aligned} \tag{25}$$

where $|0\rangle_a$ and $|1\rangle_a$ are orthonormal. For $m-p+1 \leq i \leq m+q$, it goes as:

$$\begin{aligned} & \sqrt{\lambda_i}|0\rangle_a|i\rangle_A \\ \to & \sqrt{(\frac{a}{m})(\frac{p}{p+q})}|1\rangle_a|i\rangle_A + \sqrt{\lambda_i - (\frac{a}{m})(\frac{p}{p+q})}|0\rangle_a|i\rangle_A. \end{aligned} \tag{26}$$

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

$$|0\rangle_a|i\rangle_A \to |0\rangle_a|i\rangle_A. \tag{27}$$

Hence, we find that

$$\begin{aligned} & |0\rangle_a|\Psi\rangle \\ \to & \sqrt{a}|1\rangle_a|\Psi_{\text{pre}}^{m,p,q}\rangle + \sqrt{1-a}|0\rangle_a|\Psi_{\text{res}}\rangle \end{aligned} \tag{28}$$

where

$$|\Psi_{\rm pre}^{m,p,q}\rangle = \frac{1}{\sqrt{m}} \left(\sum_{i=1}^{m-p} |i\rangle|i\rangle + \sum_{i=m-p+1}^{m+q} (\frac{p}{p+q})^{1/2} |i\rangle|i\rangle \right) \tag{29}$$

is the precursor and

$$\begin{aligned} |\Psi_{\rm res}\rangle = (1-a)^{-1/2} \Bigg[& \sum_{i=1}^{m-p} \sqrt{\lambda_i - \frac{a}{m}} |i\rangle|i\rangle \\ & + \sum_{i=m-p+1}^{m+q} \sqrt{\lambda_i - (\frac{a}{m})(\frac{p}{p+q})} |i\rangle|i\rangle \\ & + \sum_{i=m+q+1}^{N} \sqrt{\lambda_i} |i\rangle|i\rangle \Bigg] \end{aligned} \tag{30}$$

is the residual state and a is the minimal value needed for a new degeneracy to occur in Schmidt coefficients of the residual state $|\Psi_{\rm res}\rangle$. i.e., $a = \min\left(\frac{m(p+q)}{q}(\lambda_{m-p} - \lambda_{m-p+1}), \frac{m(p+q)}{p}(\lambda_{m+q} - \lambda_{m+q+1})\right)$, thus achieving either (1) $\lambda'_{m-p} = \lambda'_{m-p+1}$ or (2) $\lambda'_{m+q} = \lambda'_{m+q+1}$.

Now Alice measures the state of the ancilla. If the outcome is "1", she gets a precursor state which can be converted with certainty to an m-state. If the outcome is "0", she gets a residual state with its degeneracy number increased by 1. Also it is easy to see that, just like the original state Ψ, the residual state $|\Psi_{\rm res}\rangle$ also has the property that $\min_r B_r^m = 1$. Indeed, $\min_r B_r^m = 1$ is completely equivalent with the constraint that the largest normalized Schmidt coefficient is smaller or equal to $1/m$ which is satisfied by the residual state. This multi-step method establishes our proof. QED.

C. Lemma 3

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

Lemma 3: If $B_{r+1}^m > B_r^m$, then $B_{r+2}^m > B_{r+1}^m$.

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 of Lemma 3: Let $s' = \sum_{i=m-r+1}^{N} \lambda_i$.

$$
\begin{aligned}
B_{r+1}^m &> B_r^m \\
\frac{m}{r+1}[s' + \lambda_{m-r}] &> \frac{m}{r}s' \\
rs' + r\lambda_{m-r} &> (r+1)s' \\
r\lambda_{m-r} &> s'. \qquad (31)
\end{aligned}
$$

Now,

$$
\begin{aligned}
B_{r+2}^m &= \frac{m}{(r+2)}[s' + \lambda_{m-r} + \lambda_{m-r-1}] \\
&\geq \frac{m}{(r+2)}[s' + 2\lambda_{m-r}] \\
&= \frac{m}{(r+2)(r+1)}[(r+1)s' + 2(r+1)\lambda_{m-r}] \\
&= \frac{m}{(r+2)(r+1)}[(r+1)s' + r\lambda_{m-r} + (r+2)\lambda_{m-r}] \\
&> \frac{m}{(r+2)(r+1)}[(r+1)s' + s' + (r+2)\lambda_{m-r}] \\
&= \frac{m}{(r+2)(r+1)}[(r+2)s' + (r+2)\lambda_{m-r}] \\
&= \frac{m}{(r+1)}[s' + \lambda_{m-r}] \\
&= B_{r+1}^m, \qquad (32)
\end{aligned}
$$

where Eq. (31) is used in obtaining the fifth line. QED.

With lemma 3 proven, we now return to the proof of case (b) of theorem 2.

D. Proof of 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 following evolution:

$$|0\rangle_a|\Psi\rangle = |1\rangle_a|\Psi_s\rangle + |0\rangle_a|\Psi_f\rangle \qquad (33)$$

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 $|\Psi_f\rangle$

(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: It can be shown that the opposite statement $B_r^m \geq 1$ is equivalent to the following (redundant) recursive constraints on the individual Schmidt coefficients.

$$\begin{aligned}
\lambda_{m-1} &\leq \lambda_m + \lambda_{m+1} + \cdots + \lambda_N \\
\lambda_{m-2} &\leq \frac{1}{2}(\lambda_{m-1} + \lambda_m + \cdots + \lambda_N) \\
\lambda_{m-3} &\leq \frac{1}{3}(\lambda_{m-2} + \lambda_{m-1} + \cdots + \lambda_N) \\
&\cdots \\
\lambda_1 &\leq \frac{1}{(m-1)}(\lambda_2 + \lambda_3 + \cdots + \lambda_N)
\end{aligned} \tag{34}$$

and the normalization condition $\sum_i \lambda_i = 1$. Notice that this representation decouples the relations between the Schmidt coefficients and their overall normalization.

Consider the 'last minimal point' of the function B_r^m. i.e., r_0 such that

$$B_{r_0}^m = \min_r B_r^m < B_{r_0+1}^m. \tag{35}$$

Its existence is guaranteed by the fact that $\min_r B_r^m < 1 = B_m^m$. Lemma 3 shows that r_0 is unique. Moreover, it is straightforward to see that r_0 is the smallest number such that $\lambda_{m-r_0} > \frac{1}{r_0}(\lambda_{m-r_0+1} + \lambda_{m-r_0+2} + \cdots + \lambda_N)$, which violates the r_0-th equation in Eqs. (34). This implies that when we look at λ's in reversed order. i.e., $\lambda_N, \cdots, \lambda_m, \lambda_{m-1}, \lambda_1$, we find that $\lambda_N, \cdots, \lambda_{m-r_0+1}$ are fine (in the sense that they do not violate Eqs. (34) yet), but λ_{m-r_0} (when placed in the left hand side, violates Eqs. (34) and) is too big to be useful. Moreover, it follows from Lemma 3 that $\lambda_{m-r_0-1}, \lambda_{m-r_0-2}, \cdots, \lambda_1$ all violate Eqs. (34) when they are placed in the left hand side of the equations.

Let us define

$$\lambda_{m-r_0}^{\max} \equiv \frac{1}{r_0}(\lambda_{m-r_0+1} + \lambda_{m-r_0+2} + \cdots + \lambda_N) = \frac{B_{r_0}^m}{m}, \tag{36}$$

which is the maximal acceptable value of the $(m - r_0)$-th Schmidt coefficient. Now the successful piece $|\Psi_s\rangle$ in Eq. (33) is obtained by trimming the redundant contribution to $\lambda_1, \lambda_2, \cdots, \lambda_{m-r_0}$ from $|\Psi\rangle$. As discussed earlier, this is done by the attachment of an ancilla prepared in the state $|0\rangle_a$. The evolution goes as follows:

$$\begin{aligned} \sqrt{\lambda_i}|0\rangle_a|i\rangle_A \rightarrow & \sqrt{\lambda_{m-r_0}^{\max}}|1\rangle_a|i\rangle_A \\ & +\sqrt{\lambda_i - \lambda_{m-r_0}^{\max}}|0\rangle_a|i\rangle_A \end{aligned} \tag{37}$$

for $1 \le i \le m - r_0$. For $m - r_0 + 1 \le i \le N$, the evolution is

$$\sqrt{\lambda_i}|0\rangle_a|i\rangle_A \rightarrow \sqrt{\lambda_i}|1\rangle_a|i\rangle_A. \tag{38}$$

Alice now reads off the state of her ancilla. We now argue that an outcome "0" means that Alice has failed in getting an m-state whereas an outcome "1" means that she has succeeded in obtaining a state satisfying $\min_r B_r^m = 1$, which by Sec. VI B can be reduced with certainty to an m-state.

If the outcome is "0", the resulting (failing) state $|\Psi_f\rangle$ has unnormalized Schmidt coefficients $\lambda_1 - \lambda_{m-r_0}^{\max}, \lambda_2 - \lambda_{m-r_0}^{\max}, \cdots, \lambda_{m-r_0} - \lambda_{m-r_0}^{\max}, 0, \cdots, 0$. Since it has at most $m - r_0$ terms in its Schmidt decomposition, it follows from Lemma 2 that it gives a zero probability of getting a m-state. On the other hand, if the outcome is "1", the *un-normalized* Schmidt coefficients of the resulting (successful) state $|\Psi_s\rangle$ are given by $\lambda_{m-r_0}^{\max}, \cdots, \lambda_{m-r_0}^{\max}, \lambda_{m-r_0+1}, \lambda_{m-r_0+2}, \cdots, \lambda_N$. i.e., the first $m - r_0$-th Schmidt coefficients are all replaced by $\lambda_{m-r_0}^{\max}$. By construction $|\Psi_s\rangle$ belongs to Case (a) of Theorem 2. Therefore, it always succeeds to give an m-state. Moreover, it has a norm

$$\begin{aligned} & (m - r_0)\lambda_{m-r_0}^{\max} + \lambda_{m-r_0+1} + \cdots + \lambda_N \\ & = \frac{m}{r_0}(\lambda_{m-r_0+1} + \lambda_{m-r_0+2} + \cdots + \lambda_N) \\ & = B_{r_0}^m \\ & = \min_r B_r^m \end{aligned} \tag{39}$$

where the second and third lines come from Eq. (36) and the last from Eq. (35). This proves that our explicit strategy saturates the bound and completes our proof for the case (b) of Theorem 2. QED.

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

Unlike the Schmidt projection method (as used in the reversible strategy) in [3], the above optimal gambling strategy does require one-way communications. It is tempting to conclude that one-way communications give strategies that are intrinsically more powerful than those without. However, even for fixed m and Ψ, optimal strategies have not been proven to be unique. Hence, one could still imagine devising an optimal strategy that does not require one-way communications. But is this really possible?

We now show that one-way communications *do* generally give more powerful strategies that those without communications: When p_m^{max} is strictly less than 1, Bob generally needs Alice's help to figure out whether the gambling is successful is not. Consider the example of $|\Psi\rangle = a|11\rangle + b|22\rangle$ where $a > b > 0$ and $m = 2$. Consider *any* optimal strategy, which gives $p_2 = 2b^2$. Let us divide up its outcome into two classes: $\{s_1, s_2, \cdots, s_p\}$ (success) and $\{f_1, f_2, \cdots, f_q\}$ (failure) and denote the *un-normalized* reduced density matrix of Bob for an outcome s_i (f_j) by $\rho_{s_i}^{Bob}$ ($\rho_{f_j}^{Bob}$). Clearly, Bob needs to determine the outcome of the gambling by distinguishing with certainty between the two density matrices $\rho_{success}^{Bob} = \sum_i \rho_{s_i}^{Bob}$ and $\rho_{failure}^{Bob} = \sum_j \rho_{f_j}^{Bob}$. Now the distinguishability of two density matrices can be described by the fidelity [19] $F(\rho_{success}^{Bob}, \rho_{failure}^{Bob})$. 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 $F(\rho_{success}^{Bob}, \rho_{failure}^{Bob}) \neq 0$ or equivalently the supports of $\rho_{success}^{Bob}$ and $\rho_{failure}^{Bob}$ 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.,

$$\begin{aligned}\rho^{Bob}_{success} + \rho^{Bob}_{failure} &= \rho^{Bob}_{initial} \\ &= a^2|1\rangle\langle 1| + b^2|2\rangle\langle 2|. \end{aligned} \tag{40}$$

Since $\rho^{Bob}_{initial}$ has a two-dimensional support, $\rho^{Bob}_{success}$ must have a support of at most two dimensions. On the other hand, as $\rho^{Bob}_{s_i}$ is a singlet, $\rho^{Bob}_{success}$, being the sum of $\rho^{Bob}_{s_i}$'s, must have a support of at least two dimensions. Combining these two statements, $\rho^{Bob}_{success}$ has a support of exactly two dimensions. Now that both $\rho^{Bob}_{initial}$ and $\rho^{Bob}_{success}$ have two-dimensional supports, the support of $\rho^{Bob}_{failure}$ must be a subspace of the support of $\rho^{Bob}_{success}$. Therefore, we conclude that $\rho^{Bob}_{success}$ and $\rho^{Bob}_{failure}$ do *not* have orthogonal supports and hence the fidelity $F(\rho^{Bob}_{success}, \rho^{Bob}_{failure}) \neq 0$. QED

In conclusion, one-way communications generally give more powerful strategies than those without communications. On the other hand, we proved in Sec. II that one-way communications is sufficient for any strategy. Combining these two results, we conclude that one-way communications is necessary and sufficient for implementing any strategy of entanglement manipulations of pure states.

VII. NON-EXISTENCE OF UNIVERSAL STRATEGY

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

$$p^{tot}_m \leq p^{MAX}_m. \tag{41}$$

We have also seen in the previous section that for any particular m there exists a strategy which saturates this bound (the strategy which yields Φ_m with probability equal to p^{MAX}_m 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 could generate any possible distribution consistent with the bound (41). We shall show however 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 $\lambda_2 + \lambda_3 - \lambda_1 \geq 0$. Assume, by means of contradiction, that a universal strategy does exist. We shall use projection operators rather than positive operator valued measures (POVMs) in our discussion. As noted in Sec. 2, there is no loss of generality. Let $P_1, P_2, \cdots, P_r$ be the set of all projection operators by Alice that give some 3-state in a particular universal gambling 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 $|\Psi_r\rangle = (1 - P_1 - P_2 - \cdots - P_r)|\Psi\rangle$ has $p_2^{MAX} = 1$. But this requires the squared eigenvalues of the reduced density matrix of $|\Psi_r\rangle$ to satisfy the constraint $\lambda'_2 + \lambda'_3 - \lambda'_1 \geq 0$. We shall show that this is generally impossible. The point of our argument, as to be discussed in the next paragraph, is that the extraction of a 3-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 $p_2^{MAX}(|\Psi_r\rangle) = 1$.

The following proves our claim that $\lambda'_i = \lambda_i - p_3^{MAX}/3$. Suppose P gives a three-state with a probability α.

$$|\Psi\rangle = P|\Psi\rangle + (1 - P)|\Psi\rangle \tag{43}$$

with

$$\begin{aligned} P|\Psi\rangle = &\left(\sqrt{\lambda_1}P|1\rangle\right)|1\rangle \\ &+ \left(\sqrt{\lambda_2}P|2\rangle\right)|2\rangle \\ &+ \left(\sqrt{\lambda_3}P|3\rangle\right)|3\rangle. \end{aligned} \tag{44}$$

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

$$\rho_B = \sum_{i=1}^{3} \frac{\alpha}{3} |i\rangle\langle i|. \tag{45}$$

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

$$\begin{aligned} (1-P)|\Psi\rangle = & \sqrt{\lambda_1 - \frac{\alpha}{3}} |1''1\rangle \\ & + \sqrt{\lambda_2 - \frac{\alpha}{3}} |2''2\rangle \\ & + \sqrt{\lambda_3 - \frac{\alpha}{3}} |3''3\rangle. \end{aligned} \tag{46}$$

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

$$\begin{aligned} & \langle j|(1-P)(1-P)|i\rangle \\ = & \langle j|(1-2P+PP)|i\rangle \\ = & \langle j|(1-2PP+PP)|i\rangle \\ = & \langle j|(1-PP)|i\rangle \\ = & 0. \end{aligned} \tag{47}$$

Here the last equality follows from the fact that $P|i\rangle$'s are orthogonal to one another. This shows that an extraction of a 3-state of probability α leads to a decrease of each λ's by $\alpha/3$. The same argument can be applied to each of $P = P_1, P_2, \cdots, P_r$. This shows that $\lambda_i' = \lambda_i - p_3^{MAX}/3$ and completes our proof of the non-existence of a universal strategy. QED.

VIII. LAW OF LARGE NUMBERS

Consider the question raised in the abstract and the introduction: Can coherent 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 get nK singlets while with probability $1 - E/K$ we get no singlets at all.

In particular our result shows that *any* strategy that transforms n pairs Ψ into an *average* of nE singlets (the maximal allowed average) yields a singlet number probability distribution very similar to that of the reversible strategy of Bennett *et al.* [3]: Any such strategy yields a cumulative probability distribution roughly equal to 1 (0 respectively) when $K < E(\Psi)$ ($K > E(\Psi)$ respectively). Besides, the jump from 0 to 1 occurs in a region of width $O(n^{-1/2})$ around $E(\Psi)$.

Theorem 3. In the entanglement manipulation of n pairs Ψ, the optimal probability (over all possible strategies) of getting nK singlets, $p^{MAX}_{2^{nK}}$, 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 $p^{MAX}_{2^{nK}}$ 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 $p^{MAX}_{2^{nK}}$ tends 1 in the large n limit when $K < E(|\Psi\rangle)$ follows trivially from Bennett *et al.*'s reversible strategy [3] and Lemma 1. Let us now consider the case $K > E(|\Psi\rangle)$. As explained in the Introduction, we could view the n pairs Ψ as a single pair in state $\tilde{\Psi}$, 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 $p^{MAX}_{2^{nK}}$ which can be bounded by using Theorem 1.

Let $\tilde{\lambda}_i$'s represent the 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 simply our notation and use E to denote $E(|\Psi\rangle)$. This is essentially the law of large numbers in classical probability theory. See also quantum noiseless source coding theorem [20].] Let us pick a 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 case $N = t^n$, $m = 2^{nK}$ and $m - r + 1 = 2^{nK_0}$. Notice that $r/m > 1/2$ for a sufficiently large n. Hence, $p_m^{MAX}/2 < rp_m^{MAX}/m \leq \sum_{i=m-r+1}^{t^n} \tilde{\lambda}_i < \delta$. Substituting $m = 2^{nK}$ back, we get $p_{2^{nK}}^{MAX} \rightarrow 0$ as $n \rightarrow \infty$. QED.

The fact that any particular strategy which transforms n states Ψ into an average of nE singlets gives a singlet number probability distribution similar to that of the reversible strategy [3] follows immediately from Theorem 3 and Eq. (16).

IX. MIXED STATES

Let us now consider the case when Alice and Bob share a mixed initial state $\rho_{\rm ini}$. Since $\rho_{\rm ini}$ is impure, one generally cannot write it in terms of Schmidt decomposition. More importantly, even if $\rho_{\rm ini}$ *happens* to be symmetric under the interchange of Alice and Bob, there is no guarantee that the intermediate states that they get during the gambling process will respect such a symmetry [21]. Therefore, the symmetry argument much emphasized in the earlier part of this paper will no longer be valid. Gambling with a mixed state using two-way communications is generally more advantageous than a one-way strategy. Indeed, Bennett *et al.* have shown that one-way capacity and two-way capacity for purification are provably different [5].

We also proved in Sec. VI that in gambling with entanglement one-way communications are provably better than no communications. Notice that one-way communications is useful for gambling but not for (deterministic) quantum error correction [5]. The role of communications in a general entanglement manipulation (i.e., gambling plus quantum error correction) deserves future investigations.

For a mixed state, there are generally four distinct supremum probabilities to consider: p^2_m, $p^{A \to B}_m$, $p^{B \to A}_m$ and p^0_m corresponding to gambling 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 gambling with 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 $n \to \infty$ in the region $D_0(\rho) \leq A \leq \; E(\rho)$ 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, coherent 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^0_m of getting an m-state without any classical communications? Notice that Bennett *et al.*'s reversible strategy [3] (but not the local filtering strategy [3]) 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.

Another important open question is whether a central limit theorem holds for entanglement manipulations [22].

It cannot be over-emphasized that the symmetry that we have found here applies not only to entanglement concentration, but also to all types of entanglement manipulations including entanglement dilution [3] and quantum data compression [20]. 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 beside entanglement concentration may prove rewarding.

XI. ACKNOWLEDGMENTS

H.-K. Lo particularly thank P. Shor for enlightening discussions which indirectly inspired this line of research. Our proof of Theorem 1 has been simplified following a critical comment by R. Jozsa. S. Popescu thanks C. H. Bennett and J. Smolin for helpful communications on their independent proof of Lemma 2. Useful discussions with R. Cleve, D. Gottesman, D. Leung, M. A. Nielson and J. Preskill are greatly appreciated. Part of this work was written up during a visit of H.-K. Lo to Quantum Information and Computing (QUIC) Institute at Caltech, whose hospitality is gratefully acknowledged.

REFERENCES

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

[2] J. S. Bell, Physics **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), p. 403 and p. 409 respectively.

[3] C. H. Bennett *et al.*, Phys. Rev. **A53**, 2046 (1996).

[4] The entropy of entanglement $E(|\Phi\rangle)$ of a pure state $|\Phi\rangle$ is defined to be the von Neumann entropy of the reduced density matrix of either observer. i.e., $E(|\Phi\rangle) = S(\rho_A) = S(\rho_B)$ where $S(\rho) = -\mathrm{Tr}\rho\log_2\rho$, $\rho_A = \mathrm{Tr}_B|\Phi\rangle\langle\Phi|$ and $\rho_B = \mathrm{Tr}_A|\Phi\rangle\langle\Phi|$ are the reduced density matrices of the two subsystems.

[5] C. H. Bennett *et al.*, Phys. Rev. **A54**, 3824 (1996).

[6] S. Popescu and D. Rohrlich, Los Alamos preprint archive quant-ph/9610044.

[7] See, for example, the Appendix of L. P. Hughston, R. Jozsa and W. K. Wootters, Phys. Lett. **A183**, 14 (1993).

[8] We use the superscript MAX because, as will be shown in Section 6, the supremum probability is attainable by the optimal strategy.

[9] This result has also been found by other groups such as by C. H. Bennett and J. Smolin (private communications) and by M. Nielsen (private communications). We thank them for helpful discussions.

[10] We will use the words "gambling" and "entanglement manipulations" interchangeably.

[11] 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 has been speculated by D. Mayers [12] in the discussion of the impossibility of unconditionally secure quantum bit commitment [12–15].

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

[13] D. Mayers, Report No. quant-ph/9603015.

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

[15] H.-K. Lo and H. F. Chau, Report No. quant-ph/9605026; in *Proceedings of the fourth workshop on Physics and Computation*, (New England Complex Systems Institute, Boston, 1996).

[16] 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.

[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 equi-probable outcomes. In this sense, the two experiments are equivalent. On the contrary, suppose Alice, but not Bob, share 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 that

of Bob. In conclusion, entanglement with third party generally destroys 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 has previously been used to prove that two-way communications is provably better than one-way communications in entanglement purification of *mixed* states [5]. In our opinion, the power of symmetry arguments in entanglement manipulations remains to be fully explored.

[18] More generally, in each step of entanglement manipulations Alice and Bob may apply a direct product of local unitary transformations to their state after a measurement. It is easy to check that this in no way changes our basic arguments below. For simplicity, we shall ignore such local unitary transformations.

[19] R. Jozsa, J. Modern Optics **41**, 2315 (1994).

[20] B. Schumacher, Phys. Rev. **A51**, 2738 (1995); R. Jozsa and B. Schumacher, J. Modern Optics **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. Nielsen for raising this question.

[22] We thank M. A. Nielsen for raising this question.