Towards the Classical Communication Complexity of Entanglement Distillation Protocols with Incomplete Information

Andris Ambainis* Ke Yang †

Abstract

Entanglement is an essential resource for quantum communication and quantum computation, similar to shared random bits in the classical world. Entanglement distillation extracts nearly-perfect entanglement from imperfect entangled state. The classical communication complexity of these protocols is the minimal amount of classical information that needs to be exchanged for the conversion. In this paper, we focus on the communication complexity of protocols that operate with incomplete information*, i.e., where the inputs are mixed states and/or prepared adversarially.*

We consider three models of imperfect entanglement, namely, the bounded measurement model, the depolarization model, and the fidelity model. We describe there models as well as the motivations for studying them. For the bounded measurement model and the depolarization model, we prove tight and almost-tight bounds on the output quality of non-interactive protocols. For the fidelity model we prove a lower bound that matches the upper bound given by Ambainis et al., and thus completely characterizes communication complexity of entanglement distillation protocols for this model. Our result also suggests the optimality of the BB84 protocol in terms of communication complexity.

We emphasize that although some of the results appear intuitively straightforward, their proofs are not. In fact, two novel techniques are developed for proving these results. We believe that these techniques are of independent interests, too.

1 Introduction

Communication complexity studies the amount of communication needed to solve a certain computational problem [60, 31]. Communicating quantum bits instead of classical bits can decrease the amount of communication needed [16, 51, 52]. Besides new solutions to classical problems, quantum world also brings new open problems to communication complexity.

Entanglement distillation is a widely studied problem in quantum information theory. Entanglement Distillation Protocols (EDPs) are two-party protocols between Alice and Bob that take as input imperfectly entangled quantum states, and output near-perfect EPR pairs. In such protocols, Alice and Bob are allowed to perform local quantum operations and classical communications. However, they are not allowed to communicate in a quantum channel. Protocols of this type are called "LOCC protocols," for "Local Operation Classical Communication." For LOCC protocols, it is natural to ask what the communication complexity of these tasks is, i.e., how much information Alice and Bob need to exchange in order to produce near-perfect EPR pairs. Also, it is interesting to consider the trade-off between the amount of communication and the quality of the output.

Entanglement distillation protocols are closely related to a number of areas. We discuss some of these related areas, as well as how the communication complexity of EDPs are related in these areas.

Understanding Entanglement Entanglement, and particularly in the form of Einstein-Podolsky-Rosen pairs [18] (EPR pairs), is probably the most important phenomenon in quantum information theory, with exciting applications such as teleportation [6] and superdense coding [11]. Researchers have long trying to understand entanglement, and in particular, the *quantification* of entanglement. Given an entangled state ρ, *how much* entanglement does it have? Among the various proposes is the concept of *distillable entanglement* [10, 44], which is defined to be the asymptotic ratio of number of EPR pairs "distillable" from n copies of ρ using the optimal entanglement distillation protocol to n. A good understanding of EDPs, therefore, is essential for understanding entanglement.

Fighting Decoherence Quantum states are notoriously unstable and are easy to "decohere," that is, that they interact with the environment and become "corrupted." This can be a problem for, for example, teleportation, where Alice and Bob need to maintain a large collection of shared

*Institute for Advanced Studies, ambainis@ias.edu.
†Carnegie Mellon University, yangke@cs.cmu.edu.

EPR pairs before the teleportation starts, and imperfect EPR pairs will result in unfaithful teleportation. Naturally, Alice and Bob need to use EDPs to "extract" almost perfect EPR pairs.

Understanding Quantum Error Correcting Codes Quantum Error Correcting Codes (QECCs) are mechanisms for systematically encoding quantum information into "code-words", so that if parts of a code-words are corrupted, the original information can still be recovered by decoding. It is desirable to design QECCs with low overhead (the amount of redundancy added) that can tolerate a high rate of noise. Initiated by Shor [56] and Steane [57], the study of QECCs has become a very active area. Many constructions are proposed [20, 32, 49], and many bounds on the overhead of QECCs are known [17, 19, 54, 46, 47, 48]. Most of these bounds are proven using techniques from classical error correcting codes and are only for *non-degenerate* codes [20, 43]. On the other hand, much less is known for *degenerate* codes, since they don't have counterparts in classical error correction and novel techniques are needed to prove bounds for them.

Quantum entanglement distillation protocols can be viewed as an alternative to QECCs. Thanks to teleportation, a collection of shared EPR pairs is equivalent to a quantum channel. If Alice produces a number of EPR pairs and send over Bob's share through a noisy channel, they will share imperfect EPR pairs. Next, Alice and Bob can use an entanglement distillation protocol to distill perfect EPR pairs, and then use the distiled EPR pairs to transmit quantum information by teleportation. In this way, EDPs can be used to transmit quantum information reliably through a noisy channel. This connection was first pointed out by Bennett et al. [9], and later made more precise in [10]. Furthermore, Bennett et al. [10] showed a correspondence between one-way EDPs (where the communication is only from Alice to Bob) and QECCs. They proved that every one-way EDP implies a QECC that can tolerate the same noise rate, and vice versa.

Because of the correspondence between QECCs and EDPs, it is interesting to compare their efficiencies, and in particular, if their overheads are preserved in the conversion. We may measure the overhead of a QECC by the difference between the code length and the message length, and the overhead of EDPs is naturally measured by the amount of communication. Unfortunately, the conversion by Bennett et al. [10] does not preserve the overhead. However, the overhead is preserved for a large class QECCs, known as stabilizer codes [20]. See Nielsen and Chuang [43, pp. 597]. Such an equivalence suggests that the study of communication complexity of EDPs may provide more insights to the study of the bounds on QECCs, for both degenerating codes and non-degenerating codes. As a case in point, Leung et al. [32] considered a generalization of QECCs, which they call "approximate quantum error correcting codes", and showed that by relaxing the error correction condition, more efficient codes can be designed. This result, viewed from the perspective of the EDPs, simply corresponds to the trade-off between the amount of communication and the output quality of these protocols, which appears to be quite natural.

Understanding Quantum Key Distribution Consider the Quantum Key Distribution protocol (QKD) by Bennett and Brassard [5]. It is one of the very few results from quantum information theory that currently enjoy practical applications. See [4, 29, 39, 12, 13] for some experimental results. There also have been a sequence of proofs of security for such a protocol, with latter ones simplifying and/or strengthening the former ones; see [38, 14, 35, 33, 15, 55]. Lo and Chau [35] were the first one that made a connection from the key distribution protocols to EDPs, and the proof was further simplified by Shor and Preskill [55]. While all these studies focus on the security of such a protocol, they seem not to be concerned with the communication complexity, i.e., how efficient the BB84 protocol is in term of the classical bits exchanged.

Interestingly, quantum key distribution protocols are closely related to entanglement distillation protocols working in the so-called "fidelity noise model" (discussed later in our paper). There exists a significant amount of similarity between the definition of secure QKD protocols and the definition of conditional EDPs for the fidelity model. In particular, Lo and Chau and Shor and Preskill showed that the BB84 protocol is in some sense "equivalent" to a specific EDP, such that the security of the BB84 protocol corresponds to the "quality" of the EDP, and the communication complexity of the BB84 protocol directly corresponds to the communication complexity of the EDP. Therefore, an optimality result for this entanglement distillation protocol (as we shall show in this paper) implies that the BB84 protocol is optimal in terms of communication complexity for protocols.

1.1 Our Contributions

In this paper, we study the classical communication complexity of EDPs with *incomplete information*. In this setting, Alice and Bob don't have the complete knowledge about the input state they share. Rather, the input state is a mixed state, or is adversarially prepared.

We also focus on the *precise* communication complexity of EDPs, rather than their *asymptotic* behavior. In fact, we try to answer questions of the following fashion: "On this particular input state class, how many bits of classical communication are needed in order to just output a *single* EPR pair with certain quality?" We believe that it is important

to understand the communication complexity in this case, where the requirement appears to be *minimal*. Interestingly, as we shall see later, answers to this minimal question already yield a lot of insights into the more general problem, where Alice and Bob wish to generate EPR pairs of not only high quality, but also of large quantity.

We consider various formulations of "imperfect EPR pairs", which we call "noise models". We study the behavior of EDPs with different noise models and inputs. We summarize our results here.

A tight bound for the bounded measurement mode. In the bounded measurement model, Alice and Bob originally share n perfect EPR pairs, and then r out of these n pairs are measured in the computational basis, resulting in a mixed state $\frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)$. Alice and Bob have no information about which pairs are measured and which are not, but they know r. In other words, the measured qubit pairs are adversarially chosen. This is a simplified version of the noise model typically used in quantum error correction, where r pairs are arbitrarily corrupted. We choose to study the bounded measurement model since it is simpler for analysis yet rich enough to yield interesting results.

We prove a tight upper bound on the output fidelity (which measures the "quality" of a protocol) of non-interactive protocols, i.e., ones where Alice and Bob don't communicate. More precisely, we prove that maximal fidelity of a non-interactive protocol is at most $1 - r/2n$. This is tight since there exists a very simple protocol that achieves a fidelity of $1 - r/2n$. We view this result as the first step towards understanding EDPs for this model.

An almost tight bound for the depolarization model. In the depolarization model, Alice generates n EPR pairs by herself, and then sends to Bob his share over through a depolarization channel of parameter p, which independently leave each qubit unchanged with probability $(1-p)$ and replace it with a completely mixed state with probability p. It is a typical model for "noisy channels", and in particular was studied by Bennett et al. [9, 10].

We prove an almost tight upper bound for non-interactive protocols over this model. More precisely, we show that any non-interactive protocol has maximal fidelity $1 - p/2$ in its output. This bound if almost tight in that there exists a very simple protocol of output fidelity $1 - 3p/4$.

A complete characterization for the fidelity model. The fidelity model is an adversarial noise model, where the only information Alice and Bob have is that the fidelity of their input state and the perfect EPR pairs is $1 - \varepsilon$. Ambainis et al. [1], studied this model in the name of "general error" model. This model was also independently studied by Lo and Chau [35] and Shor and Preskill [55] in proving security of the BB84 quantum key distribution protocol, who showed that the BB84 protocol is, in fact, an entanglement distillation protocol for the fidelity model.

We present a complete characterization of EDPs over the fidelity model. We prove an almost tight lower bound (up to an additive constant) on the communication complexity of EDPs over the fidelity model. More precisely, we prove that the maximal conditional fidelity of an EDP with t bits of communication is at most $1 - \varepsilon \cdot p/2^{t+1}$, even if the EDP is only required to output 1 qubit pair. Here, ε is the fidelity of the input state, p is the probability that the EDP succeeds with perfect EPR pairs, and the conditional fidelity is the fidelity of the EDP conditioned on it succeeding (we allow an EDP to fail in this case). Therefore, to achieve a fidelity or $1 - \delta$ on the output, $\log(1/\delta) + \log(\varepsilon \cdot p) - 1$ bits of classical communication is needed. Comparing the result from [1], which contains a protocol that (with simple modification) uses $\log(1/\delta) + \log(1-\varepsilon)$ bits, our lower bound is tight up to an additive constant (under the reasonable assumption that both ε and p are constant). Our result implies that the "random hashing" protocol by Ambainis et al. [1] optimal. Since essentially the same protocol is used in the BB84 key distribution protocols, as pointed out by Lo and Chau [35], the BB84 protocol is also optimal in terms of communication complexity.

We stress that some of these results may seem intuitively straightforward, their proofs do not appear so. In fact, in order to prove these results, we need to develope two novel techniques that might be interesting by themselves.

Alternative definition of fidelity We give an alternative definition of the fidelity of a pure state and an EPR pair. We first notice that an EPR pair (denoted by Φ^+) is the unique state that remains unchanged under a group of operators. Then we show that for an arbitrary pure state $|\phi\rangle$, its "deviation" from this group of operators is exactly the fidelity of $|\phi\rangle$ and Φ^+. See Lemma 3.

This technique is used to prove the two results for the bounded measurement model and the depolarization model. It is interesting to compare this technique to the stabilizer formalism [21], where a state is defined as the unique elements that is "stabilized" by a group of operations, i.e., that is remains unchanged under these operations. Our alternative definition suggests that it may be interesting to consider states that are "partially" stabilized as well.

Analysis of protocols with mixed state input We introduce a technique to analyze general LOCC protocols with mixed states as input. Prior to our work, most of the work on LOCC protocols only deal with pure states as input. Having a pure state as input greatly simplifies the analysis, since the Schmidt decomposition can be used. Many researchers have used Schmidt decomposition in their analysis, including Lo and Popescu [36], Nielsen [41], Hayden and Winter [25], and Nayak and Salzman [40]. Unfortunately this technique does not work for mixed states, since Schmidt decomposition is only for pure states. In fact, Lo and Chau proved that for pure state inputs, one-way protocols are as powerful as two-way protocols. On the other hand, Bennett et al. [10] showed that for certain mixed state inputs, two-way protocols are provable more powerful than one-way protocols. These results shows a distinct difference between pure state and mixed states.

Our technique, on the other hand, is designed to analysis protocols with mixed states as input. Roughly speaking, our technique works as follows. We consider both the reduced density matrix of Alice and Bob. When Alice sends a classical bit to Bob, this may cause Bob's density matrix to "split", since if Alice and Bob's states are entangled, then the bit sent by Alice may carry some information about Bob's state.[1] Our technique keeps track of the splitting reduced density matrix pair as the protocol proceeds, and builds a binary tree corresponding to the messages exchanged. By maintaining an invariant when traversing the tree of message history, we manage to prove our result. We discuss this in Section 5.

1.2 Related Work

To the best of our knowledge, the study of entanglement distillation protocols was initiated by Bennett et al. [8], who considered the problems of producing perfect EPR pairs from a large copy of identical pure states. From then on, the problem of entanglement distillation was studied by a number of researchers from different perspectives [9, 10, 27, 28, 44, 45, 50, 24]. All of which consider the situation where n identical copies of a state are given as input to an LOCC protocol, which then outputs m EPR pairs. They studied the asymptotic behavior of m/n as n approaches infinity.

Researchers also studied EDPs for a single copy of an arbitrary pure state; see Vidal [58], Jonathan and Plenio [30], Hardy [22], and Vidal et al. [59]. Much of the work was built on the result of majorization by Nielsen [41], who is the first one that studied conditions under which one pure state can be transformed into another one by LOCC. All the work above assumes that Alice and Bob know the explicit description of the state they share, and so they can act *optimally*.

Relatively less work was done on studying EDPs with incomplete information prior to this paper. See Bennett et al [9, 10]. The fidelity noise model was independently studied by a number of researchers: Lo and Chau [35] and Shor and Preskill [55] in proving security of the BB84 protocol; Barnum et al. [2] in the study of "purity-testing protocols" protocols; Ambainis et al. [1] in relating EDPs to classical randomness extractors.

Researchers have also studied the classical communication complexity of other quantum tasks. Lo and Popescu [37] observed that the "entanglement concentration protocol" in [8] does not require any classical communication, while the "entanglement dilution protocol" requires $O(n)$ bits of classical communication for producing n copies of the "diluted" state. They also constructed a new dilution protocol that only uses $O(\sqrt{n})$ bits of communication. This protocol was proven to be asymptotically optimal by Hayden and Winter [25], and Harrow and Lo [23]. Lo [34] studied the communication complexity for Alice and Bob to jointly *prepare* a large number of copies of arbitrary (known) pure states, and proved an non-trivial upper bound. All the results above focus on a relatively simple situation, where the input are n copies of a known pure state, and almost all are asymptotic results.

2 Notations and Definitions

All logarithms are base-2. We identify an integer with the 0-1 vector obtained from its binary representation. For a vector v, we write $v[j]$ to denote its j-th entry. For 0-1 vector x, we denote its *Hamming weight* by $|x|$, which is the number of 1's in x. For binary strings x and y, we use $x;y$ to denote the *concatenation* of these two strings.

Throughout the paper we are interested in finite, bipartite, symmetric quantum systems shared between Alice and Bob. We identify a "ket" $|\phi\rangle$ with a unit column vector. We assume there exists a canonical computational basis for any finite Hilbert space of dimension N, and we denote it by $\{|0\rangle, |1\rangle, \ldots, |N-1\rangle\}$. We use superscripts to indicate which "side" a qubit or an operation belongs to. For example, a general bipartite state $|\varphi\rangle$ can written as $|\varphi\rangle = \sum_{i,j} \alpha_{ij} |i\rangle^A |j\rangle^B$.

There are 4 *Bell states* for a pair of qubits shared between Alice and Bob, and we denote them as $\Phi^+ = \frac{1}{\sqrt{2}}(|0\rangle^A|0\rangle^B + |1\rangle^A|1\rangle^B)$, $\Phi^- = \frac{1}{\sqrt{2}}(|0\rangle^A|0\rangle^B - |1\rangle^A|1\rangle^B)$, $\Psi^+ = \frac{1}{\sqrt{2}}(|0\rangle^A|1\rangle^B + |1\rangle^A|0\rangle^B)$, and $\Psi^- = \frac{1}{\sqrt{2}}(|0\rangle^A|1\rangle^B - |1\rangle^A|0\rangle^B)$.

[1] On the other hand, assuming that Alice and Bob don't erase their information, if Alice and Bob's input states are not entangled, then the bit sent by Alice will not cause the split.

We denote the state $(\Phi^+)^{\otimes n}$, which represents n perfect EPR pairs, by Φ_n. We also abuse the notation to use Φ_n to denote *both* the vector Φ_n and its density matrix $|\Phi_n\rangle\langle\Phi_n|$, when there is no danger of confusion.

The Pauli Matrices X, Y, and Z are unitary operations over a single qubit defined as

$$\begin{aligned} X(\alpha|0\rangle+\beta|1\rangle) &= \beta|0\rangle+\alpha|1\rangle \\ Y(\alpha|0\rangle+\beta|1\rangle) &= i\beta|0\rangle-i\alpha|1\rangle \\ Z(\alpha|0\rangle+\beta|1\rangle) &= \alpha|0\rangle-\beta|1\rangle \end{aligned}$$

We use I to denote the identity operator.

For a unitary operator U, we can write it in a matrix form under the computational basis. Then we define its *conjugate*, U^*, to the entry-wise conjugate of U. Clearly U^* is still a unitary operation. An *error model* is simply a set of bipartite (mixed) states, and is often denoted by $\mathcal{M}$. We say a state ρ is *consistent* with $\mathcal{M}$, if $\rho \in \mathcal{M}$.

Fidelity is a measure of closeness between quantum states which we use to measure the quality of the output of an EDP. For two mixed states ρ and σ in the same Hilbert space their fidelity is defined as $F(\rho,\sigma) = \mathrm{Tr}^2(\sqrt{\rho^{1/2}\sigma\rho^{1/2}})$. If $\sigma = |\varphi\rangle\langle\varphi|$ is a pure state, the definition simplifies to $F(\rho,|\varphi\rangle\langle\varphi|) = \langle\varphi|\rho|\varphi\rangle$. A special case is when $|\varphi\rangle = \Phi_n$ for some n, such that ρ and Φ_n have the same dimension. In this case, we call the fidelity of ρ and $|\varphi\rangle$ the *fidelity of state* ρ, and the definition simplifies to $F(\rho) = \langle\Phi_n|\rho|\Phi_n\rangle$.

We are often interested in the fidelity of two states of unequal dimensions, and in particular, the fidelity of a general state ρ and the Bell state Φ^+. Then, we define the *base fidelity* of ρ to be the fidelity of the state obtained by tracing out all but the first qubit pair of ρ. We denote the base fidelity of ρ by $\widetilde{F}(\rho)$.

It is easy to verify that the fidelity is linear with respect to ensembles, so long as one of the inputs is a pure state.

Claim 1 *If ρ is the density matrix for a mixed state that is an ensemble $\{p_i, |\phi_i\rangle\}$, and σ is the density matrix of a pure state, then we have $F(\rho,\sigma) = \sum_i p_i \cdot F(|\phi_i\rangle\langle\phi_i|,\sigma)$.* ■

The fidelity is also monotone with respect to trace-preserving operations [43].

Claim 2 *For any states ρ and σ and any trace-preserving operator $\mathcal{E}$, we have $F(\mathcal{E}(\rho),\mathcal{E}(\sigma)) \geq F(\rho,\sigma)$.* ■

One useful fact about fidelity is that any completely disentangled state has base fidelity at most $1/2$.

Lemma 1 *If ρ is a completely disentangled state, then $\widetilde{F}(\rho) \leq 1/2$.*

Proof: By the definition of base fidelity, we may assume that ρ has dimension 2. By Claim 1, we only need to consider the case that ρ is a pure state $|\phi\rangle\langle\phi|$. Since $|\phi\rangle$ is disentangled, we may write it as

$$|\phi\rangle = (\alpha_0|0\rangle+\alpha_1|1\rangle)\otimes(\beta_0|0\rangle+\beta_1|1\rangle)$$

Then a direct calculation reveals that

$$\begin{aligned} \widetilde{F}(|\phi\rangle\langle\phi|) &= \frac{1}{2}|\alpha_0\beta_0+\alpha_1\beta_1|^2 \\ &= \frac{1}{2}\left(|\alpha_0|^2|\beta_0|^2+|\alpha_1|^2|\beta_1|^2+\alpha_0\beta_0\alpha_1^*\beta_1^*+\alpha_0^*\beta_0^*\alpha_1\beta_1\right) \\ &\leq \frac{1}{2}\left(|\alpha_0|^2|\beta_0|^2+|\alpha_1|^2|\beta_1|^2+|\alpha_0\beta_1^*|^2+|\alpha_1\beta_0^*|^2\right) \\ &= \frac{1}{2}, \end{aligned}$$

where the inequality is due to Cauchy-Schwartz. ■

2.1 Entanglement Distillation Protocols

We often denote an entanglement distillation protocol by $\mathcal{P}$. The protocol starts with a mixed state ρ shared between Alice and Bob. Alice and Bob can have their private ancillary qubits, originally all initialized to $|0\rangle$. A protocol is either deterministic or probabilistic. For *deterministic* protocols, Alice and Bob don't share any initial random bits; for *probabilistic* protocols, Alice and Bob share a (classical) random string. We say a protocol $\mathcal{P}$ is a t-bit protocol, if there are t bits of (classical) communication during the protocol. We don't allow protocols to have any initial entanglement as auxiliary inputs, nor do we allow quantum channels between Alice and Bob.

An the end of a protocol, both parties output m qubits, which form the output of the protocol. If σ is the density matrix of the output of protocol $\mathcal{P}$ on input ρ, we write it as $\mathcal{P}(\rho) = \sigma$. For an entanglement distillation protocol $\mathcal{P}$, we define its *fidelity* with respect to an error model $\mathcal{M}$, denoted by $F_{\mathcal{M}}(\mathcal{P})$, to be the minimal fidelity of its output over all input states consistent with $\mathcal{M}$. I. e.,

$$F_{\mathcal{M}}(\mathcal{P}) = \min_{\rho\in\mathcal{M}} F(\mathcal{P}(\rho)) \tag{1}$$

In the fidelity error model (Section 5), we allow protocols to fail with some probability. (As shown in [1], this is necessary for having good output fidelity in this model.) In this case, Alice also outputs a special symbol (either a SUCC or a FAIL). The *success probability* of a protocol $\mathcal{P}$ over an input state ρ is the probability that Alice outputs SUCC at the end of the protocol, and we write this as $P_{\mathcal{P}}^{\mathrm{SUCC}}[\rho]$. The *ideal success probability* of a protocol $\mathcal{P}$ is its success probability over the ideal input Φ_n. We say a protocol is *ideal*, if its ideal success probability is 1.

If τ is the density matrix of the output of protocol $\mathcal{P}$ on input ρ, *conditioned on* that Alice outputs SUCC, then we call τ the *conditional output* of protocol $\mathcal{P}$, and write this as $\mathcal{P}^c(\rho) = \tau$. We define the *conditional fidelity* to be the minimal fidelity of its conditional output:

$$F^c_{\mathcal{M}}(\mathcal{P}) = \min_{\rho \in \mathcal{M}} F(\mathcal{P}^c(\rho)) \tag{2}$$

When the error model $\mathcal{M}$ is clear from the context, it is often omitted.

3 The Bounded Measurement Model

We prove an upper bound on the fidelity of 0-bit EDPs with respect to the bounded measurement error model.

In the bounded measurement model, the input state of EDP consists of n EPR pairs r of which have been measured. That is, the input state is $|\phi_{\mathbf{v}}\rangle = \bigotimes_{j=0}^{n-1} |\phi_j\rangle$, where

$$|\phi_j\rangle = \begin{cases} |0\rangle^A|0\rangle^B & \text{if } v[j] = 0 \\ |1\rangle^A|1\rangle^B & \text{if } v[j] = 1 \\ \Phi^+ & \text{if } v[j] = * \end{cases}$$

and $\mathbf{v} \in \{0,1,*\}^n$. The state $|\phi_{\mathbf{v}}\rangle$ is called an *error state*, where $\mathbf{v}$ is called its *error indicator vector*. The degree of $\mathbf{v}$, denoted by $\deg(\mathbf{v})$, is the number of i's in $\{1,\ldots,N\}$ for which $\mathbf{v}_i \neq *$. The error model for the bounded measurement model, denoted by $\mathcal{M}^{\mathsf{m}}_{n,r}$, is defined to be

$$\mathcal{M}^{\mathsf{m}}_{n,r} = \{|\phi_{\mathbf{v}}\rangle \mid \deg(\mathbf{v}) = r\} \tag{3}$$

An n-dimensional 0-1 vector x is *consistent* with a binary indicator vector $\mathbf{v}$, if $x[j] = \mathbf{v}[j]$ for all j such that $\mathbf{v}[j] \neq *$. We write this as $x \sqsubseteq \mathbf{v}$. For any $\mathbf{v}$ of degree r, there are 2^{n-r} 0-1 vectors x consistent with $\mathbf{v}$. It is not hard to verify that

$$|\phi_{\mathbf{v}}\rangle = \frac{1}{2^{(n-r)/2}} \sum_{x \sqsubseteq \mathbf{v}} |x\rangle^A|x\rangle^B \tag{4}$$

3.1 Two Useful Lemmas

We prove two lemmas that would be useful for the proofs in this paper. Both lemmas are about how much "deviation" a quantum state undergoes when applied various unitary operations.

First, we consider the "deviation" of an arbitrary pure state under the operations $\{I,X,Y,Z\}$ over its first qubit. We have the following lemma:

Lemma 2 *Let $|\phi\rangle$ and $|\psi\rangle$ be two pure states of the same dimension, not necessarily bipartite. Let I, X, Y, and Z be the unitary operations over the first qubit of $|\phi\rangle$. Then we have*

$$\sum_{U \in \{I,X,Y,Z\}} |\langle\phi|U|\psi\rangle|^2 \leq 2 \tag{5}$$

Proof: We write $|\phi\rangle = \alpha_0|0\rangle|\phi_0\rangle + \alpha_1|1\rangle|\phi_1\rangle$ and $|\psi\rangle = \beta_0|0\rangle|\psi_0\rangle + \beta_1|1\rangle|\psi_1\rangle$

Then we have

$$\begin{aligned}
\langle\phi|I|\psi\rangle &= \alpha_0^*\beta_0\langle\phi_0|\psi_0\rangle + \alpha_1^*\beta_1\langle\phi_1|\psi_1\rangle \\
\langle\phi|X|\psi\rangle &= \alpha_1^*\beta_0\langle\phi_1|\psi_0\rangle + \alpha_0^*\beta_1\langle\psi_0|\phi_1\rangle \\
\langle\phi|Y|\psi\rangle &= -i\alpha_1\beta_0^*\langle\phi_1|\psi_0\rangle + i\alpha_0\beta_1^*\langle\phi_0|\psi_1\rangle \\
\langle\phi|Z|\psi\rangle &= \alpha_0^*\beta_0\langle\phi_0|\psi_0\rangle - \alpha_1^*\beta_1\langle\phi_1|\psi_1\rangle
\end{aligned}$$

Therefore

$$\begin{aligned}
&\sum_{U \in \{I,X,Y,Z\}} |\langle\phi|U|\psi\rangle|^2 \\
= \;& 2|\alpha_0\beta_0|^2|\langle\phi_0|\psi_0\rangle|^2 + 2|\alpha_1\beta_1|^2|\langle\phi_1|\psi_1\rangle|^2 + \\
& 2|\alpha_0\beta_1|^2|\langle\phi_0|\psi_1\rangle|^2 + 2|\alpha_1\beta_0|^2|\langle\phi_1|\psi_0\rangle|^2 \\
\leq \;& 2|\alpha_0|^2|\beta_0|^2 + 2|\alpha_1|^2|\beta_1|^2 + \\
& 2|\alpha_0|^2|\beta_1|^2 + 2|\alpha_1|^2|\beta_0|^2 \\
= \;& 2(|\alpha_0|^2 + |\alpha_1|^2)(|\beta_0|^2 + |\beta_1|^2) \\
= \;& 2
\end{aligned}$$

■

An immediate corollary is

Corollary 1 *Let $|\phi\rangle$ be a pure sate. We have $\sum_{U \in \{I,X,Y,Z\}} |\langle\phi|U|\phi\rangle|^2 \leq 2$.*

Next, we consider quantum states and operations over bipartite systems. In particular, we study the "deviation" of a general bipartite state under unitary operations of the form $U \otimes U^*$. We interpret $U \otimes U^*$ as Alice applies U to her first qubit and Bob applies U^* to his first qubit. Again, we consider $U \in \{I,X,Y,Z\}$.

We have the following lemma.

Lemma 3 *Let $|\phi\rangle$ be a pure state in a bipartite system shared between Alice and Bob. Let I, $X \otimes X^*$, $Y \otimes Y^*$, and $Z \otimes Z^*$ be the unitary operations over the first All these 4 operations work on the first qubit of Alice and the first qubit of Bob. Then we have*

$$\sum_{U \in \{I,X,Y,Z\}} \langle\phi|U \otimes U^*|\psi\rangle = 4\widetilde{F}(|\phi\rangle) \tag{6}$$

Proof: We first consider how the Bell states behave under these unitary operations. It is easy to verify the result, which we compile into the following Table 3.1.

It is easy to see that the state Φ^+ is invariant under any of the 4 operations, while other Bell states will change their signs under some operations.

Notice the 4 Bell states form an orthonormal basis for a bipartite system of 2 qubits. We decompose $|\phi\rangle$ into the Bell basis and write

$$|\phi\rangle = \alpha_0\Phi^+ \otimes |\psi_0\rangle + \alpha_1\Phi^- \otimes |\psi_1\rangle + \alpha_2\Psi^+ \otimes |\psi_2\rangle + \alpha_3\Psi^- \otimes |\psi_3\rangle$$

state	Φ^+	Φ^-	Ψ^+	Ψ^-
$I\otimes I^*$	Φ^+	Φ^-	Ψ^+	Ψ^-
$X\otimes X^*$	Φ^+	-Φ^-	Ψ^+	-Ψ^-
$Y\otimes Y^*$	Φ^+	-Φ^-	-Ψ^+	Ψ^-
$Z\otimes Z^*$	Φ^+	Φ^-	-Ψ^+	-Ψ^-

Table 1. The Bell States under operators

where $\sum_{j=0}^{3}|\alpha_j|^2=1$. Therefore we have

$$\begin{aligned}\langle\phi|\phi\rangle &= |\alpha_0|^2+|\alpha_1|^2+|\alpha_2|^2+|\alpha_3|^2\\ \langle\phi|(X\otimes X^*)|\phi\rangle &= |\alpha_0|^2-|\alpha_1|^2+|\alpha_2|^2-|\alpha_3|^2\\ \langle\phi|(Y\otimes Y^*)|\phi\rangle &= |\alpha_0|^2-|\alpha_1|^2-|\alpha_2|^2+|\alpha_3|^2\\ \langle\phi|(Z\otimes Z^*)|\phi\rangle &= |\alpha_0|^2+|\alpha_1|^2-|\alpha_2|^2-|\alpha_3|^2\end{aligned}$$

and thus $\langle\phi|\phi\rangle+\langle\phi|(X\otimes X^*)|\phi\rangle+\langle\phi|(Y\otimes Y^*)|\phi\rangle+\langle\phi|(Z\otimes Z^*)|\phi\rangle=4|\alpha_0|^2=4\widetilde{F}(|\phi\rangle)$. ■

Lemma 3 in fact gives an alternative definition of the base fidelity of a pure state.

We prove that the fidelity of 0-bit EDPs for the bounded measurement error model is at most $1-r/2n$, even if the protocols are only required to output one qubit-pair. Notice that fidelity is monotone. Therefore if no protocol can output a single qubit pair of fidelity at least $1-r/2n$, then no protocol can output multiple qubit pairs of fidelity at least $1-r/2n$.

Theorem 1 *For any probabilistic 0-bit protocol $\mathcal{P}$ that outputs one qubit pair, we have $F(\mathcal{P})\le 1-\frac{r}{2n}$ with respect to the bounded measurement model.*

Notice that there exists a very simple probabilistic 0-bit protocol of fidelity $1-\frac{r}{2n}$: Alice and Bob use their shared random string to uniformly pick an EPR pair and output it. If this pair is measured, (which happens with probability r/n), the fidelity is 1/2, and otherwise it is 1. So the overall fidelity is exactly $1-r/2n$ and thus our upper bound is tight.

Proof: We consider a slightly different error model, where a *random* r out of n EPR pairs are measured. This corresponds to the density matrix

$$\rho=\frac{1}{2^n\binom{n}{r}}\sum_{\mathbf{v}:\deg\mathbf{v}=r}|\phi_{\mathbf{v}}\rangle\langle\phi_{\mathbf{v}}|$$

Notice that this is the "average case" version of the bounded measurement model. Thus if we prove an upper bound on the fidelity of $\mathcal{P}$ over ρ, then it is also an upper bound with respect to the bounded measurement model.

We shall prove that no *deterministic* 0-bit protocol can have a fidelity higher than $1-r/2n$ if ρ is the input. Then, we conclude that no probabilistic protocol can have a fidelity higher than $1-r/2n$, too, since fidelity is linear.

Notice $\mathcal{P}$ is non-interactive, we can model it as Alice and Bob both applying a unitary operation to their share of qubits, outputs the first qubit and discard the rest.

Suppose the unitary operators of Alice and Bob are U_A and U_B. We denote the states under these operations by $U_A|x\rangle\longrightarrow|\phi_x\rangle$ and $U_B|x\rangle\longrightarrow|\psi_x\rangle$.

Notice that we use "$\longrightarrow$" instead of "=" since we allow Alice and Bob to use ancillary bits. Clearly, the vectors $\{|\phi_x\rangle\}_x$ are orthonormal, and so are the vectors $\{|\psi_x\rangle\}_x$.

We shall prove that

$$\frac{1}{2^r\binom{n}{r}}\sum_{\deg\mathbf{V}=r}\left[[\widetilde{F}((U_A\otimes U_B)|\phi_{\mathbf{v}}\rangle\langle\phi_{\mathbf{v}}|(U_A\otimes U_B)^\dagger)\right]\le 1-\frac{r}{2n},\tag{7}$$

which shall imply Theorem 1.

By Lemma 3, (7) is equivalent to

$$\begin{aligned}&\sum_{\deg\mathbf{V}=r}\left[\sum_{U\in\{I,X,Y,Z\}}\langle\phi_{\mathbf{v}}|(U_A\otimes U_B)^\dagger(U\otimes U^*)(U_A\otimes U_B)|\phi_{\mathbf{v}}\rangle\right]\\ \le\;& 2^r\binom{n}{r}\cdot 4(1-\frac{r}{2n})\end{aligned}$$

We expand the left hand side: Notice that

$$(U_A\otimes U_B)|\phi_{\mathbf{v}}\rangle=\frac{1}{2^{(n-r)/2}}\sum_{x\sqsubseteq\mathbf{V}}|\phi_x\rangle|\psi_x\rangle$$

where $x\sqsubseteq\mathbf{v}$ if x is *consistent* with $\mathbf{v}$ (that is, if $x[j]=\mathbf{v}[j]$ for all j such that $\mathbf{v}[j]\neq *$).

Therefore, we have

$$\begin{aligned}&\langle\phi_{\mathbf{v}}|(U_A\otimes U_B)^\dagger(U\otimes U^*)(U_A\otimes U_B)|\phi_{\mathbf{v}}\rangle\\ =\;&\frac{1}{2^{n-r}}\sum_{x\sqsubseteq\mathbf{V}}\sum_{y\sqsubseteq\mathbf{V}}\langle\phi_x|U|\phi_y\rangle\cdot\langle\psi_x|U^*|\psi_y\rangle\end{aligned}$$

for any unitary operation U. So we only need to prove that

$$\sum_{\deg\mathbf{V}=r}\sum_{x\sqsubseteq\mathbf{V}}\sum_{y\sqsubseteq\mathbf{V}}\sum_{U\in\{I,X,Y,Z\}}\langle\phi_x|U|\phi_y\rangle\cdot\langle\psi_x|U^*|\psi_y\rangle\le 2^n\binom{n}{r}\cdot 4(1-\frac{r}{2n})$$

However, by Cauchy-Schwartz, we have

$$\begin{aligned}&\sum_{\deg\mathbf{V}=r}\sum_{x\sqsubseteq\mathbf{V}}\sum_{y\sqsubseteq\mathbf{V}}\sum_{U\in\{I,X,Y,Z\}}\langle\phi_x|U|\phi_y\rangle\cdot\langle\psi_x|U^*|\psi_y\rangle\\ \le\;&\left(\sum_{\deg\mathbf{V}=r}\sum_{x\sqsubseteq\mathbf{V}}\sum_{y\sqsubseteq\mathbf{V}}\sum_{U\in\{I,X,Y,Z\}}|\langle\phi_x|U|\phi_y\rangle|^2\right)^{\frac{1}{2}}\cdot\\ &\left(\sum_{\deg\mathbf{V}=r}\sum_{x\sqsubseteq\mathbf{V}}\sum_{y\sqsubseteq\mathbf{V}}\sum_{U\in\{I,X,Y,Z\}}|\langle\psi_x|U^*|\psi_y\rangle|^2\right)^{\frac{1}{2}}\end{aligned}$$

Next, we estimate the terms on the right hand side:

$$\begin{aligned}
&\sum_{\deg \mathbf{V}=r}\sum_{x\sqsubseteq\mathbf{V}}\sum_{y\sqsubseteq\mathbf{V}}\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_x|U|\phi_y\rangle|^2 \\
= &\sum_x\sum_y\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_x|U|\phi_y\rangle|^2 \sum_{\deg\mathbf{V}=r\,:\,x_1\sqsubseteq\mathbf{V}\wedge x_2\sqsubseteq\mathbf{V}} 1
\end{aligned}$$

Notice that since $|\phi_x\rangle$'s are all orthonormal, we have $\sum_y |\langle\phi_x|U|\phi_y\rangle|^2 \le 1$ for all x's. Thus

$$\sum_x\sum_y\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_x|U|\phi_x\rangle|^2 \le 2^{n+2}$$

For any x and y, we have

$$\sum_{\deg\mathbf{V}=r\,:\,x\sqsubseteq\mathbf{V}\wedge y\sqsubseteq\mathbf{V}} 1 = \binom{n-|x\oplus y|}{n-r-|x\oplus y|}$$

The reason is simple: the only freedom for $\mathbf{v}$ is where to put the $(n-r)$ $*$'s. But for every position k such that $x[k]\neq y[k]$, we have to have $\mathbf{v}[k]=*$. Then we still have $(n-r-|x\oplus y|)$ $*$'s we can put anywhere. So if $x\neq y$,

$$\sum_{\deg\mathbf{V}=r\,:\,x\sqsubseteq\mathbf{V}\wedge y\sqsubseteq\mathbf{V}} 1 \le \binom{n-1}{n-r-1}$$

Also notice that by Lemma 2, we have $\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_x|U|\phi_x\rangle|^2 \le 2$ for any x.

Putting things together, we have

$$\begin{aligned}
&\sum_{\deg\mathbf{V}=r}\sum_{x\sqsubseteq\mathbf{V}}\sum_{y\sqsubseteq\mathbf{V}}\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_x|U|\phi_y\rangle|^2 \\
\le &\binom{n}{r}\cdot\sum_x\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_x|U|\phi_x\rangle|^2 + \\
&+\binom{n-1}{r-1}\cdot\sum_{x\neq y}\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_x|U|\phi_y\rangle|^2 \\
= &\left[\binom{n}{r}-\binom{n-1}{r-1}\right]\cdot\sum_x\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_x|U|\phi_x\rangle|^2 + \\
&\binom{n-1}{r-1}\cdot\sum_x\sum_y\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_x|U|\phi_y\rangle|^2 \\
= &\left[\binom{n}{r}-\binom{n-1}{r-1}\right]\cdot 2^{n+1} + \binom{n-1}{r-1}\cdot 2^{n+2} \\
= &\ 2^{n+2}\binom{n}{r}(1-\frac{r}{2n})
\end{aligned}$$

Similarly, we have

$$\sum_{\deg\mathbf{V}=r}\sum_{x\sqsubseteq\mathbf{V}}\sum_{y\sqsubseteq\mathbf{V}}\sum_{U\in\{I,X,Y,Z\}} |\langle\psi_x|U^*|\psi_y\rangle|^2 \le 2^{n+2}\binom{n}{r}(1-\frac{r}{2n})$$

too.

Thus we have

$$\begin{aligned}
&\sum_{\deg\mathbf{V}=r}\sum_{x\sqsubseteq\mathbf{V}}\sum_{y\sqsubseteq\mathbf{V}}\sum_{U\in\{I,X,Y,Z\}} \langle\phi_x|U|\phi_y\rangle\cdot\langle\psi_x|U^*|\psi_y\rangle \\
\le &\ 2^{n+2}\binom{n}{r}(1-\frac{r}{2n})
\end{aligned}$$

which proves the theorem. ■

4 The Depolarization Model

We prove an upper bound on the fidelity of 0-bit EDPs with respect to the depolarization model.

We first describe the depolarization channel. A depolarization channel $\mathcal{D}$ of parameter p is a super-operator defined as [43]

$$\mathcal{D}(\rho) = (1-p)\cdot\rho + p\cdot\frac{I}{2}$$

In other words, this channel behaves in the following manner: with probability $(1-p)$, it keeps the state untouched, and with probability p, it replaces that with the completely mixed state. After passing the second qubit through this channel, the state Φ^+ becomes a mixed state $\rho_p = (1-\frac{3p}{4})|\Phi^+\rangle\langle\Phi^+| + \frac{p}{4}(|\Phi^-\rangle\langle\Phi^-| + |\Psi^+\rangle\langle\Psi^+| + |\Psi^-\rangle\langle\Psi^-|)$.

The depolarization error model of n qubit pairs and parameter n, denoted as $\mathcal{M}^{\mathsf{d}}_{n,p}$, consists of a single state: $\mathcal{M}^{\mathsf{d}}_{n,p} = \{\rho_p^{\otimes n}\}$.

We prove that the maximal fidelity of 0-bit EDPs for the depolarization error model is $1-p/2$, even if the protocols are only required to output one qubit-pair.

Theorem 2 *For any probabilistic 0-bit protocol $\mathcal{P}$ that outputs one qubit pair, we have $F(\mathcal{P}) \le 1-\frac{p}{2}$ with respect to the depolarization model.* ■

There exists a very simple deterministic 0-bit protocol that has fidelity $1-\frac{3p}{4}$: Alice and Bob simply output the first qubit pair. It is very easy to verify that the fidelity of this protocol is $1-\frac{3p}{4}$. Therefore the bound in the theorem is almost-tight (up to a constant factor).

The proof to Theorem 2 is very similar to that to Theorem 1, except that it is more complicated. We omit the proof due to space limitations.

5 The Fidelity Model

We study the communication complexity of EDPs with respect to the fidelity error model.

First, we give the definition of the fidelity error model. For a bipartite system of n qubit pairs, we define the fidelity

error model of parameter ε to be the set of all bipartite systems of fidelity at least $1-\varepsilon$. We denote the error model by

$$\mathcal{M}^{\mathsf{f}}_{n,\varepsilon} = \{\rho \,|\, F(\rho) \geq 1-\varepsilon\} \tag{8}$$

Notice that this error model is very different from the two previous models we studied, since it provides much less information than the previous one. As a comparison, notice that in the bounded measurement model, all the error states have fidelity $1/2^r$, and in the depolarization model, the fidelity of the input is $(1-3p/4)^n$, both are very small. However, Alice and Bob have the additional information about the *structure* of the input states, and are able to use the information to do very well.

5.1 Two Useful Facts About Positive Operators

We present two useful facts about positive operators. used in the rest of the paper.

For two positive operators A and B, we say A *dominates* B, if $A-B$ is still a positive operator, and we write this as $A \succeq B$, or equivalently, $B \preceq A$.

Claim 3 *For any positive super-operator $\mathcal{E}$ and any positive operators A and B, if $A \succeq B$, then $\mathcal{E}(A) \succeq \mathcal{E}(B)$.* ■

This directly follows the fact that $\mathcal{E}$ is linear and preserves the positivity of operators: If $A-B$ is a positive operator, then $\mathcal{E}(A)-\mathcal{E}(B) = \mathcal{E}(A-B)$ is also a positive operator.

Claim 4 *Let ρ and σ be density matrices such that $\rho \succeq a \cdot \sigma$, for some positive number a. For any POVM $\{E_m\}$, let $p_m = \mathrm{Tr}(\rho E_m)$ and and $q_m = \mathrm{Tr}(\sigma E_m)$ be the probabilities the measurement result being m for ρ and σ, respectively. Then we have $p_m \geq a \cdot q_m$.* ■

This is obvious, since we have $p_m - a \cdot q_m = \mathrm{Tr}((\rho - a \cdot \sigma)E_m) \geq 0$.

5.2 Bounds for the Fidelity Model

Ambainis et al. [1] proved that in the fidelity error model of parameter ε (which they called the "general error model"), the maximal fidelity of a protocol is $1 - \frac{2^m-2^k}{2^m}\frac{2^n}{2^n-1}\varepsilon$. If the protocol has n qubit pairs as input, k perfect EPR pairs as auxiliary input, and outputs m qubit pairs. In a special case where $k=0$ (no auxiliary input) and $m=1$ (only one pair is output), the maximal fidelity is $1 - \frac{2^n}{2^n-1}\frac{\varepsilon}{2} < 1-\varepsilon/2$. In other words, no "interesting" entanglement distillation protocols exist for the fidelity error model. Their result is tight, in that they also constructed a protocol, namely the "Random Permutation Protocol", which achieves a fidelity of $1 - \frac{2^m-2^k}{2^m}\frac{2^n}{2^n-1}\varepsilon$.

One can modify this protocol to eliminate communication. The resulting protocol has fidelity about $1-\frac{3}{4}\varepsilon$ (therefore communication almost doesn't help at all in this case). We also have a lower bound that matches the protocol up to exponentially small terms.

Theorem 3 *(a) There exists a probabilistic 0-bit entanglement distillation protocol of fidelity $1 - \frac{3}{4}\frac{2^n-\frac{2}{3}}{2^n-1}\varepsilon$ with respect to the fidelity model of parameter ε.*

(b) If $\frac{2^{2n}}{2^{2n}-1}\varepsilon \leq \frac{1}{2}$, then any probabilistic 0-bit entanglement distillation protocol has fidelity at most $1 - \frac{3}{4}\frac{2^{2n}}{2^{2n}-1}\varepsilon$ with respect to the fidelity model of parameter ε.

The proof to this theorem is postponed to Appendix A.

The situation for conditional fidelity is very different. Ambainis et al. proved that good protocols exist with high *conditional fidelity*. In particular, the following result can be easily derived from [1]:

Theorem 4 *[1] For every n and $s < n$, there exists probabilistic s-bit entanglement distillation protocols of conditional fidelity $1-2^{-s}/(1-\varepsilon)$ with respect to the fidelity model of parameter ε.*

Proof's sketch: Consider the "Simple Random Hash" protocol in [1]. The original construction for this protocol in [1] has $(2n+2)$ bits of two-way communication. But a close examination reveals that 1 bit of one-way communication suffices. In the original construction, Alice sends $2n$ bits to Bob to establish a common random string, which are not needed for a probabilistic protocol. In the original protocol, Bob also sends 1 bit of his measurement result back to Alice. This bit can also be eliminated in our model, since we allow one player (normally Alice) to output a SUCC or FAIL symbol at the end of the protocol. We then repeat the simplified 1-bit protocol for s rounds sequentially, and obtain an s-bit protocol of conditional fidelity $1-2^{-s}/(1-\varepsilon)$. ■

Notice that this protocol only consists of one-way communication. Also notice this protocol is ideal, in that if the input is the perfect EPR pairs Φ_n, then the protocol always succeeds.

Therefore, to achieve a conditional fidelity of $1-\delta$, only $\log(\frac{1}{\delta}) - \log(1-\varepsilon)$ bits of communication is needed in the fidelity error model. Next, we prove a lower bound on the communication complexity.

Theorem 5 *For any probabilistic s-bit protocol of ideal success probability p, its conditional fidelity is at most $1-\varepsilon p/2^{s+1}$ with respect to the fidelity model of parameter ε.*

Immediately from the theorem, we obtain a $\log(\frac{1}{\delta}) - \log(\frac{1}{\varepsilon}) - 1$ lower bound on the communication complexity for ideal protocols of conditional fidelity $1-\delta$. In the usual setting where ε is a constant, our lower bound matches the upper bound from Theorem 4, up to an additive constant. Interestingly, the theorem is proven for protocols that only output 1 qubit pair. However, this lower bound matches the upper bound of the Simple Random Hash protocol, which in fact outputs many qubit pairs. In this sense, the communication complexity is quite independent from the yield of the EDPs.

Proof: WLOG we assume the protocol only outputs one qubit pair. Consider a particular input state

$$\rho_0 = (1-\varepsilon')\Phi_n + \varepsilon' \cdot \frac{I}{2^{2n}} \tag{9}$$

It is a mixture of the perfect EPR pairs Φ_n (with probability $1-\varepsilon'$) and the completely mixed state $\frac{I}{2^{2n}}$ (with probability ε'). Notice that $F(\frac{I}{2^{2n}}) = \frac{1}{2^{2n}}$. So if we set $\varepsilon' = \frac{2^{2n}}{2^{2n}-1}\varepsilon$, then we have $F(\rho) = 1-\varepsilon$. We shall prove that no deterministic, s-bit protocol has fidelity more than $1-2^{-(s+1)}\varepsilon p$ over state ρ_0, which will imply that no probabilistic protocol can have fidelity more than $1-2^{-(s+1)}\varepsilon p$, too.

We fix a deterministic protocol $\mathcal{P}$. WLOG, we assume it proceeds in *rounds*: in each round, one of the two parties (Alice or Bob) applies a super-operator $\mathcal{E}$ to his or her share of qubits, and then sends one (classical) bit to the other party. The protocol consists of s rounds, with one bit in each round. Finally, Alice outputs the special symbol, determining if the protocol succeeds or fails.

To analyze the behavior of the protocol $\mathcal{P}$ over the input ρ_0, we consider how $\mathcal{P}$ behaves over state Φ_n and state $\frac{I}{2^{2n}}$, respectively. We use p (resp. q) to denote the probabilities that $\mathcal{P}$ succeeds over state Φ_n (resp. $\frac{I}{2^{2n}}$). Notice p is in fact the ideal success probability of protocol $\mathcal{P}$. Then it is easy to see that

$$F^c(\mathcal{P}(\rho_0)) = \frac{(1-\varepsilon')p \cdot F^c(\mathcal{P}(\Phi_n)) + \varepsilon' q \cdot F^c(\mathcal{P}(\frac{I}{2^{2n}}))}{(1-\varepsilon')p + \varepsilon' q} \tag{10}$$

Notice that we always have $F^c(\mathcal{P}(\Phi_n)) \le 1$. Since $\frac{I}{2^{2n}}$ is a disentangled state, $\mathcal{P}(\frac{I}{2^{2n}})$ is also disentangled.

By Lemma 1, we have $F^c(\mathcal{P}(\frac{I}{2^{2n}})) \le 1/2$. We shall prove that

$$q \ge p^2/2^s, \tag{11}$$

which will imply that

$$\begin{aligned} F(\mathcal{P}(\rho_0)) &\le \frac{(1-\varepsilon') + \varepsilon' p/2^{s+1}}{(1-\varepsilon') + \varepsilon' p/2^s} \\ &= 1 - \frac{\varepsilon' p}{2^{s+1}(1-\frac{2^s}{2^s-1}\varepsilon' p)} \\ &\le 1 - \varepsilon p/2^{s+1} \end{aligned}$$

Now we prove that $q \ge p^2/2^s$. We analyze two cases separately: in case I, the state Φ_n is the input to the protocol; in case II, the state $\frac{I}{2^{2n}}$ is the input to the protocol. For each case, we keep track of the reduced density matrices of Alice and Bob. In case I, we use $\tau_k^{\mathrm{I},A}$ and $\tau_k^{\mathrm{I},B}$ to denote the reduced density matrices of Alice and Bob after the k-th round; in case II, we use $\tau_k^{\mathrm{II},A}$ and $\tau_k^{\mathrm{II},B}$, respectively. For $k=0$, we define the $\tau_0^{\mathrm{I},A}$, $\tau_0^{\mathrm{I},B}$, $\tau_0^{\mathrm{II},A}$, and $\tau_0^{\mathrm{II},B}$ to be the density matrices at the moment that protocol starts.

We give more definitions: after the k-th round, there are 2^k possibilities depending on the first k bits communicated. For any binary string $t \in \{0,1\}^k$, we use $\sigma_t^{\mathrm{I},A}$ (resp. $\sigma_t^{\mathrm{I},B}$) to denote the reduced density matrix of Alice (resp. Bob) after the k-th round in case I, conditioned on that the first k bits communicated so far are $t[0], t[1], \ldots, t[k-1]$. We use p_t^{I} to denote the probability that this happens (that the first k bits are $t[0], t[1], \ldots, t[k-1]$). Obviously we have $p_t^{\mathrm{I}} = p_{t;0}^{\mathrm{I}} + p_{t;1}^{\mathrm{I}}$ for any $t \in \{0,1\}^k$. Furthermore, we have the following equalities

$$\sum_{t\in\{0,1\}^k} p_t^{\mathrm{I}} = 1 \tag{12}$$

$$\sum_{t\in\{0,1\}^k} p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A} = \tau_k^{\mathrm{I},A} \tag{13}$$

$$\sum_{t\in\{0,1\}^k} p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},B} = \tau_k^{\mathrm{I},B} \tag{14}$$

We define $\sigma_t^{\mathrm{II},A}$, $\sigma_t^{\mathrm{II},B}$, and p_t^{II} for case II, similarly.

We use ξ to denote the empty string. So we have $p_\xi^{\mathrm{I}} = p_\xi^{\mathrm{II}} = 1$.

One important observation is that when the protocol starts, the reduced density matrices for Alice and Bob are identical in both cases:

$$\sigma_\xi^{\mathrm{I},A} = \sigma_\xi^{\mathrm{I},B} = \sigma_\xi^{\mathrm{II},A} = \sigma_\xi^{\mathrm{II},B} = \frac{I}{2^n} \tag{15}$$

When the protocol proceeds, the reduced density matrices in two cases will become different, since the state Φ_n is an entangled state, while $\frac{I}{2^{2n}}$ is not. However, they cannot differ "too far", as we shall prove in the following lemma. (proof postponed to Appendix A).

Lemma 4 *For all $k = 0,1,\ldots,s-1$ and $t \in \{0,1\}^k$, $p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A} \preceq \sigma_t^{\mathrm{II},A}$ and $p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},B} \preceq \sigma_t^{\mathrm{II},B}$.*

Now we are ready to prove (11). After s bits are sent, Alice will decide whether to succeed or fail. In case I,

we use r_t to denote the probability that Alice choose to succeed conditioned on that the bits communicated are $t[0], t[1], \ldots, t[s-1]$. Notice we have $p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A} \preceq \sigma_t^{\mathrm{II},A}$, and thus by Lemma 4, we know that in case II, the success probability is at least $p_t^{\mathrm{I}} \cdot r_t$.

Therefore, we have

$$p = \sum_{t \in \{0,1\}^s} r_t \cdot p_t^{\mathrm{I}} \tag{16}$$

$$q \geq \sum_{t \in \{0,1\}^s} r_t \cdot p_t^{\mathrm{I}} \cdot p_t^{\mathrm{I}} \tag{17}$$

which implies that

$$q \geq \sum_{t \in \{0,1\}^s} r_t \cdot \left(p_t^{\mathrm{I}}\right)^2 \tag{18}$$

$$\geq \frac{1}{2^s} \left(\sum_{t \in \{0,1\}^s} r_t \right) \cdot \left[\sum_{t \in \{0,1\}^s} r_t \cdot \left(p_t^{\mathrm{I}}\right)^2 \right] \tag{19}$$

$$\geq \frac{1}{2^s} \left(\sum_{t \in \{0,1\}^s} r_t \cdot p_t^{\mathrm{I}} \right)^2 \tag{20}$$

$$= \frac{p^2}{2^t} \tag{21}$$

This proves the theorem. ■

References

[1] A. Ambainis, A. Smith, and K. Yang. Extracting Quantum Entanglement (General Entanglement Purification Protocols). In *IEEE CCC'02*, pp. 103–112, 2002.

[2] H. Barnum, C. Crépeau, D. Gottesman, A. Smith and A. Tapp. Authentication of Quantum Messages. In *FOCS 2002*, also available at `quant-ph/0205128`.

[3] G. Brassard. Quantum communication complexity (a survey). Available at `quant-ph/0101005`.

[4] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, J. Smolin. Experimental quantum cryptography. In *Journal of Cryptology*, 5:3–28, 1992.

[5] C. H. Bennett and G. Brassard. Quantum Cryptography: Public-key Distribution and Coin Tossing. In *IEEE Int. Conf. on Computers, Systems and Signal Processing*, pp.175 – 179, 1984.

[6] C. H. Bennett, G. Brassard, C. Crépeau, R. Josza, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. In *Phys. Rev. Lett.*, pp. 1895, 1993.

[7] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf,. Quantum Lower Bounds by Polynomials. In *FOCS'98*, pp. 352–361., 2001. Also available at `quant-ph/9802049`.

[8] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. In *Phys. Rev. A*, vol. 53, No. 4, April 1996.

[9] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters. Purification of Noisy Entanglement and Faithful Teleportation via Noisy Channels. In *Phys. Rev. Lett.*, vol. 76, pages 722-725, 1996.

[10] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. In *Phys. Rev. A*, vol. 54, No. 5, pages 3824-3851, November 1996.

[11] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. In *Phys. Rev. Lett.* **69**, 2881 (1992).

[12] D. S. Bethune and W. P. Risk. An autocompensating fibre-optic quantum cryptography system based on polarization splitting of light. In *IQEC'98 Digest of Postdeadline Papers*, pp. QPD 12-2, Optical Society of America, 1998.

[13] D. S. Bethune and W. P. Risk. An autocompensating fibre-optic quantum cryptography system based on polarization splitting of light. In *J. Quantum Electronics*, 36(3):100, 2000.

[14] E. Biham, M. Boyer, G. Brassard, J. van de Graaf, and T. Mor. Security of quantum key distribution against all collective attacks. In *LANL e-print* `quant-ph/9801022`, 1998.

[15] E. Biham, M. Boyer, P. O. Boykin, T. Mor, and V. Roychowdhury. A proof of the security of quantum key distribution. In *FOCS '00*, pp. 512–724, 2000. Also available at *LANL e-print* `quant-ph/9912053`.

[16] H. Buhrman, R. Cleve, and A. Wigderson Quantum vs. classical communication and computation. *Proceedings of STOC'98*, pp. 63-68, quant-ph/9702040.

[17] A. Ekert and C. Macchiavello. Error correction in quantum communication. In *Phys. Rev. Lett.*, 77:2585, 1996.

[18] A. Einstein, B. Podolsky, and N. Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? In *Phys. Rev.* **47**, 777 (1935)

[19] D. Gottesman. Class of quantum error correcting codes saturating the quantum Hamming bound. In *Phys. Rev. A*, 54:1862, 1996.

[20] D. Gottesman. Stabilizer Codes and Quantum Error Correction. *Ph.D. thesis*, California Institute of Technology, 1997.

[21] D. Gottesman. Theory of fault-tolerant quantum computation. In *Phys. Rev. A*, vol. 57(1):127–137, 1998.

[22] L. Hardy. Method of areas for manipulating the entanglement properties of one copy of a two-particle pure entangled state. In *Phys. Rev. A*, **60**, 1912 (1999). also available at `quant-ph/9903001`.

[23] A. Harrow and H. K. Lo. A tight lower bound on the classical communication cost of entanglement dilution. in `quant-ph/0204096`.

[24] M. Hayashi and K.Matsumoto. Universal distortion-free entanglement concentration. in `quant-ph/0209030`.

[25] P. Hayden and A. Winter. On the communication cost of entanglement transformations. In `quant-ph/0204092`.

[26] A. Holevo. Some estimates of the information transmitted by quantum communication channels. In *Problems of Information Transmission*, 9:177–183, 1973.

[27] M. Horodecki, P. Horodecki, and R. Horodecki. Distillability of Inseparable Quantum Systems. In *LANL e-print* quant-ph/9607009.

[28] M. Horodecki, P. Horodecki, and R. Horodecki. Asymptotic entanglement manipulations can be genuinely irreversible. In *Phys. Rev. Lett.*, 84:4260–4263, 2000. See errata at *LANL e-print* quant-ph/9912076.

[29] R. J. Hughes, D. M. Alde, P. Dyer, G. G. Luther, G. L. Morgan, and M. Schauer. Quantum cryptography. In *Contemp. Phys.*, 36(3):149–163, 1995. Also available at *LANL e-print* quant-ph/9504002.

[30] D. Jonathan and M. Plenio. Minimal conditions for local pure-state entanglement manipulation. In *Phys. Rev. Lett.* **83**, 1455 (1999), also available at *LANL e-print* quant-ph/9903054.

[31] E. Kushilevitz and N. Nisan. Communication Complexity. *Cambridge University Press*, 1997.

[32] D. W. Leung, M. A. Nielsen, I. L. Chuang and Y. Yamamoto. Approximate quantum error correction can lead to better codes. In *Phys. Rev. A*, 56:2567–2573, 1997, also available at *LANL e-print* quant-ph/9704002.

[33] H. K. Lo. A simple proof of the unconditional security of quantum key distribution. available at *LANL e-print* quant-ph/9904091.

[34] H. K. Lo. Classical communication cost in distributed quantum information processing — a generalization of quantum communication complexity. available at *LANL e-print* quant-ph/9912009.

[35] H. K. Lo and H. F. Chau. Unconditional Security of Quantum Key Distribution Over Arbitrary Long Distances. In *Science* **283**, 2050-2056 (1999), also available at *LANL e-print* quant-ph/9803006.

[36] H. K. Lo and S. Popescu. Concentrating local entanglement by local actions — beyond mean values. Available at*LANL e-print* quant-ph/9707038.

[37] H. K. Lo and S. Popescu. The classical communication cost of entanglement manipulation: Is entanglement an inter-convertible resource? In *Phys. Rev. Lett.*, **83**, pp. 1459 – 1462, 1999, also available at*LANL e-print* quant-ph/9902045.

[38] D.Mayers. Unconditional security in quantum cryptography. In *LANL e-print* quant-ph/9802025, 1998.

[39] A. Muller, H. Zbinden, and N. Gisin. Quantum cryptography over 23km in installed under-lake telecom fibre. In *Europhys. Lett.*, 33:334–339, 1996.

[40] A. Nayak and J. Salzman. On communication over an entanglement-assisted quantum channel. In *STOC 2002* and *IEEE CCC 2002*.

[41] M. Nielsen. Conditions for a class of entanglement transformations. In *Phys. Rev. Lett.* **83** (2), pp 436–439 (1999), also available at quant/ph/9811053.

[42] M. Nielsen, Probability distributions consistent with a mixed state. available at quant/ph/9909020.

[43] M. Nielsen and I. Chuang. Quantum Computation and Quantum Information. *Cambridge University Press*, 2000.

[44] E. M. Rains. Rigorous treatment of distillable entanglement. In *Phys. Rev. A*, 60(1):173–178, 1999, also available at eprint quant-ph/9809078.

[45] E. M. Rains. Bound on distillable entanglement. In *Phys. Rev. A*, 60(1):179–184, 1999, Errata 63(1), 2001, also available at eprint quant-ph/9809082.

[46] E. M. Rains. Quantum weight enumerators. In *IEEE Trans. Inf. Theory*, 44(4):1388–1394, 1998.

[47] E. M. Rains. Monotonicity of the quantum linear programming bound. In *IEEE Trans. Inf. Theory*, 45(7):2489–2492, 1999.

[48] E. M. Rains. Quantum shadow enumerators. In *IEEE Trans. Inf. Theory*, 45(7):2361–2366, 1999.

[49] E. M. Rains. Nonbinary quantum codes. In *IEEE Trans. Inf. Theory*, 45(6):1827–1832, 1999.

[50] E. M. Rains. A semidefinite program for distillable entanglement. eprint quant-ph/0008047.

[51] R. Raz. Exponential separation of quantum and classical communication complexity. *STOC 1999*, 358-367.

[52] A. Razborov. Quantum communication complexity of symmetric predicates, *Izvestiya of the Russian Academy of Science, mathematics*, 2002. quant-ph/0204025.

[53] C. Shannon. A Mathematical Theory of Communication. In *Bell Sys. Tech. Journal*, 27:379–423, 623–656, 1948.

[54] P. W.Shor and R. Laflamme. Quantum analog of the MacWilliams identities for classical coding theory. In *Phys. Rev. Lett.*, 78(8):1600–1602, 1997.

[55] P. W. Shor and J. Preskill. Simple Proof of Security of the BB84 Quantum key Distribution Protocol. In *Phys. Rev. Lett.* 85 (2000). 441-444, also available at quant-ph/0003004.

[56] P. W. Shor. Schemes for Reducing Decoherence in Quantum Computer Memory. In *Phys. Rev. A*, **52**, 2493 (1995).

[57] A. M. Steane. Error Correcting Codes in Quantum Theory. In *Phys. Rev. A*, **77**, 793 (1996).

[58] G. Vidal. Entanglement of pure states for a single copy. in *Phys. Rev. Lett.* 83 (1999) 1046-1049, quant-ph/9902033.

[59] G. Vidal, D. Jonathan, and M. Nielsen Approximation Transformations and Robust Manipulation of Bipartite Pure State Entanglement. in *Phys. Rev. A* **62**, 012304 (2000) Also available at quant-ph/9910099.

[60] A. Yao. Some complexity questions related to distributed computing. In *STOC'79*, pp 209 – 213, 1979.

[61] A. Yao. Quantum circuit complexity. In *FOCS'93*, pp. 352 – 361, 1993.

A Proofs to the Results in the Fidelity Model

Proof: **[to Theorem 3,a]** Our protocol is a modification of the random permutation protocol of [1].

No-communication Random Permutation Protocol.

1. Using the shared random string, Alice and Bob generate a uniformly random permutation $\pi \in S_{2^n}$ and $x_1 \in \{-1,1\}$, $x_2 \in \{-1,1\}, \ldots, x_{2^n} \in \{-1,1\}$.
2. Alice and Bob apply the transformation U mapping $U|i\rangle = (-1)^{x_i}|\pi(i)\rangle$ to their qubits.
3. They output the first EPR pair and trace out the rest.

Note that if they are given the perfect state Φ_n, then $U \otimes U|\Phi_n\rangle = \Phi_n$ and the output is a perfect EPR pair. If the starting state is not perfect, then the first two steps "symmetrize" it.

Claim 5 *Let ρ be the mixed state obtained after the first two steps. Then,*

$$\rho = p_0|\Phi_n\rangle\langle\Phi_n| + p_1\rho_1 + p_2\rho_2 + p_3\rho_3$$

where ρ_1 is a uniform mixture of 2^n states $|i\rangle|i\rangle$, ρ_2 is a uniform mixture of $2^n(2^n-1)$ states $\frac{1}{\sqrt{2}}(|i\rangle|j\rangle + |j\rangle|i\rangle)$, $j \neq i$, ρ_3 is a uniform mixture of $2^n(2^n-1)$ states $\frac{1}{\sqrt{2}}(|i\rangle|j\rangle - |j\rangle|i\rangle)$, $j \neq i$ and p_0, p_1, p_2, $p_3 \in R$.

Proof: We divide the transformation into two parts: $U = U''U'$, $U'_{|i\rangle} = (-1)^{x_i}$, $U''|i\rangle = |\pi(i)\rangle$. Let ρ' be the intermediate density matrix after applying U'. Then, the only nonzero entries in ρ' are $|i\rangle|i\rangle\langle i|\langle i|$, $|i\rangle|i\rangle\langle j|\langle j|$, $|i\rangle|j\rangle\langle i|\langle j|$, $|i\rangle|j\rangle\langle j|\langle i|$. Applying U'' after that makes all entries of each type equal.

Let a,b,c,d be their values. Then, we can set $p_0 = 2^n a$, $p_1 = 2^n(b-a)$, $p_2 = 2^n(2^n-1)(c+d)$, $p_3 = 2^n(2^n-1)(c-d)$. ■

We have $F(\rho_0) = 1$, $F(\rho_1) = \frac{1}{2^n}$ and $F(\rho_2) = F(\rho_3) = 0$. We note that

$$p_0 + \frac{1}{2^n}p_1 \geq 1 - \varepsilon \tag{22}$$

because each of states $U \otimes U|\psi\rangle$ has the same fidelity as $|\psi\rangle$ and fidelity is convex. We can rewrite (22) as $\frac{2^n-1}{2^n}p_1 + p_2 + p_3 \leq \varepsilon$.

Outputting the first EPR pair and tracing out the rest transforms ρ_0 into a state of fidelity 1, ρ_1 into a state of fidelity 1/2 and ρ_2 and ρ_3 into states of fidelity $(2^{n-1}-1)/2(2^n-1)$. Thus, the final fidelity is $1-\delta$,

$$\delta = \frac{1}{2}p_1 + \frac{3\cdot 2^{n-1}-1}{2(2^n-1)}(p_2+p_3) \leq \frac{3\cdot 2^{n-1}-1}{2(2^n-1)}\varepsilon = \frac{3}{4}\frac{2^n-2/3}{2^n-1}\varepsilon.$$

■

Proof: **[to Theorem 3,b]** Let ρ be the mixture of $|\Phi_n\rangle\langle\Phi_n|$ with probability $1 - \frac{2^{2n}}{2^{2n}-1}\varepsilon$ and the completely mixed state in $2^n \times 2^n$ dimensions with probability $\frac{2^{2n}}{2^{2n}-1}\varepsilon$. Since the perfect state has fidelity 1 and the completely mixed state has fidelity $\frac{1}{2^{2n}}$, this state has fidelity $1-\varepsilon$.

W.l.o.g., a no-communication protocol consists of Alice applying U_A, Bob applying U_B and each of them outputting the first qubit.

Let ρ_A be the density matrix of Alice's first qubit if she starts with her system in 2^n-dimensional completely mixed state. As any density matrix on one qubit, ρ_A has can be decomposed into mixture of two orthogonal one-qubit states (its eigenstates)

$$\rho_A = \lambda_1|\psi_A\rangle\langle\psi_A| + \lambda_2|\psi_A^\perp\rangle\langle\psi_A^\perp|$$

where $\lambda_{1,2}$ are the eigenvalues of ρ_A. Since eigenvalues of a density matrix must sum up to 1, we can assume that $\lambda_1 = \frac{1}{2} + \delta_A$ and $\lambda_2 = \frac{1}{2} - \delta_A$, $\delta_A \geq 0$. Let ρ_B be the density matrix of Bob's first qubit if he starts with his system in 2^n-dimensional completely mixed state. We define $|\psi_B\rangle$, $|\psi_B^\perp\rangle$, δ_B similarly. Let $\delta = \max(\delta_A, \delta_B)$.

Claim 6 *If the starting state is Φ_n, the fidelity of the final state is at most $1-\delta^2$.*

Proof: W.l.o.g. assume that $\delta = \delta_A$.

Consider Alice's part of Φ_n. It is the completely mixed state on Alice's 2^n dimensional system. Therefore, Alice's output qubit will be in the state ρ_A. This means that the fidelity of the state output by Alice+Bob and $|00\rangle + |11\rangle$ is at most the fidelity between ρ_A and $\frac{1}{2}I$ (density matrix of Alice's part of $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$).

Let U be the unitary transformation that maps $|0\rangle$ to $|\psi_A\rangle$ and $|1\rangle$ to $|\psi_A^{\perp}\rangle$. Then,

$$F(\rho_A, \frac{1}{2}I) = F(U^{-1}\rho_A U, \frac{1}{2}I) = F\left(\begin{pmatrix} \frac{1}{2}+\delta & 0 \\ 0 & \frac{1}{2}-\delta \end{pmatrix}, \frac{1}{2}I\right)$$

$$= \left(\frac{1}{\sqrt{2}}\sqrt{\frac{1}{2}+\delta} + \frac{1}{\sqrt{2}}\sqrt{\frac{1}{2}-\delta}\right)^2 = \frac{1}{2} + \sqrt{\frac{1}{4}-\delta^2} \leq \frac{1}{2} + \left(\frac{1}{2}-\delta^2\right) = 1-\delta^2.$$

■

Claim 7 *If the starting state is the completely mixed state in 2^{2n} dimensions, the fidelity of the final state is at most $\frac{1}{4}+\varepsilon$.*

Proof: Since the completely mixed state is the tensor product of completely mixed states of Alice and Bob, the final state of output qubits is $\rho_A \otimes \rho_B$. This state is a mixture of $|\psi\rangle \otimes |\psi'\rangle$, where $|\psi\rangle$ (or $|\psi'\rangle$) is one of $|\psi_A\rangle$ and $|\psi_A^{\perp}\rangle$ (or $|\psi_B\rangle$ and $|\psi_B^{\perp}\rangle$) with probabilities $(\frac{1}{2} \pm \delta_A)(\frac{1}{2} \pm \delta_B)$.

Notice that

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|\psi\rangle|\psi^*\rangle + |\psi^{\perp}\rangle|(\psi^{\perp})^*\rangle)$$

for any one qubit state $|\psi\rangle$. In particular, we can take $|\psi\rangle = |\psi_A\rangle$. Let $a = |\langle\psi_A^*|\psi_B\rangle|^2$. Then, the fidelity of states $|\psi_A\rangle \otimes |\psi_B\rangle$ and $|\psi_A^{\perp}\rangle \otimes |\psi_B^{\perp}\rangle$ is $\frac{a}{2}$ and the fidelity of states $|\psi_A\rangle \otimes |\psi_B^{\perp}\rangle$ and $|\psi_A^{\perp}\rangle \otimes |\psi_B\rangle$ is $\frac{1-a}{2}$. Therefore, the overall fidelity of the final state is

$$\frac{a}{2}\left((\frac{1}{2}+\delta_A)(\frac{1}{2}+\delta_B) + (\frac{1}{2}-\delta_A)(\frac{1}{2}-\delta_B)\right) + \frac{1-a}{2}\left((\frac{1}{2}+\delta_A)(\frac{1}{2}-\delta_B) + (\frac{1}{2}-\delta_A)(\frac{1}{2}+\delta_B)\right)$$

$$= \frac{a}{2}\left(\frac{1}{2}+2\delta_A\delta_B\right) + \frac{1-a}{2}\left(\frac{1}{2}-2\delta_A\delta_B\right) \leq \frac{1}{2}\left(\frac{1}{2}+2\delta_A\delta_B\right) \leq \frac{1}{4}+\delta^2.$$

■

Therefore, the fidelity of the protocol on ρ_4 is at most

$$(1-\frac{2^{2n}}{2^{2n}-1}\varepsilon)(1-\delta^2) + \frac{2^{2n}}{2^{2n}-1}\varepsilon(\frac{1}{4}+\delta^2) \leq 1 - \frac{3}{4}\frac{2^{2n}}{2^{2n}-1}\varepsilon. \tag{23}$$

If Alice and Bob share randomness, we can fix one value r for randomness and take U_A and U_B for this r. The bound of equation (23) applies for any particular r, Therefore, it also applies on the average over all r. ■

Proof: [to Lemma 4] By induction. The base case is obvious. Now the inductive case. Consider the situation at the end of the k-th round. Suppose the first k bits sent are $t[0], t[1], \ldots, t[k-1]$. WLOG we assume that in the $(k+1)$-th round, Alice applies a super-operator $\mathcal{E}$ to her share of qubits, and send one bit a to Bob.

First we consider the density matrix for Alice. Notice that in general, a is the result of the measurement from $\mathcal{E}$. Therefore, we can "split" $\mathcal{E}$ into two positive super-operators $\mathcal{E}_0$ and $\mathcal{E}_1$, such that

$$\mathcal{E}_0(\sigma_t^{\mathrm{I},A}) = \frac{p_{t;0}^{\mathrm{I}}}{p_t^{\mathrm{I}}} \cdot \sigma_{t;0}^{\mathrm{I},A} \tag{24}$$

$$\mathcal{E}_1(\sigma_t^{\mathrm{I},A}) = \frac{p_{t;1}^{\mathrm{I}}}{p_t^{\mathrm{I}}} \cdot \sigma_{t;1}^{\mathrm{I},A} \tag{25}$$

$$\mathcal{E}_0(\sigma_t^{\mathrm{II},A}) = \frac{p_{t;0}^{\mathrm{II}}}{p_t^{\mathrm{II}}} \cdot \sigma_{t;0}^{\mathrm{II},A} \tag{26}$$

$$\mathcal{E}_1(\sigma_t^{\mathrm{II},A}) = \frac{p_{t;1}^{\mathrm{II}}}{p_t^{\mathrm{II}}} \cdot \sigma_{t;1}^{\mathrm{II},A} \tag{27}$$

Intuitively, $\mathcal{E}_0$ corresponds to the case that $a = 0$ is sent, and $\mathcal{E}_1$ corresponds to the case that $a = 1$ is sent.

By inductive hypothesis, we have

$$p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A} \preceq \sigma_t^{\mathrm{II},A} \tag{28}$$

Combining (28), (24) and (26) with Claim 3 yields that

$$p_{t;0}^{\mathrm{I}} \cdot \sigma_{t;0}^{\mathrm{I},A} = \mathcal{E}_0(p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A}) \preceq \mathcal{E}_0(\sigma_t^{\mathrm{II},A}) = \frac{p_{t;0}^{\mathrm{II}}}{p_t^{\mathrm{II}}} \cdot \sigma_{t;0}^{\mathrm{II},A} \preceq \sigma_{t;0}^{\mathrm{II},A} \tag{29}$$

Combining (28), (25) and (27) with Claim 3 yields that

$$p_{t;1}^{\mathrm{I}} \cdot \sigma_{t;1}^{\mathrm{I},A} = \mathcal{E}_1(p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A}) \preceq \mathcal{E}_1(\sigma_t^{\mathrm{II},A}) = \frac{p_{t;1}^{\mathrm{II}}}{p_t^{\mathrm{II}}} \cdot \sigma_{t;1}^{\mathrm{II},A} \preceq \sigma_{t;1}^{\mathrm{II},A} \tag{30}$$

Now we consider the reduced density matrix for Bob. In case I, the qubits between Alice and Bob are entangled. Therefore, the bit Alice sends to Bob carries some information about his state. In terms of the density matrix, Bob's reduced density matrix will "split" from $\sigma_t^{\mathrm{I},B}$ to $\sigma_{t;0}^{\mathrm{I},B}$ and $\sigma_{t;1}^{\mathrm{I},B}$. Notice that Bob doesn't perform any operation to his qubits, and thus we have

$$\sigma_t^{\mathrm{I},B} = \frac{p_{t;0}^{\mathrm{I}}}{p_t^{\mathrm{I}}} \cdot \sigma_{t;0}^{\mathrm{I},B} + \frac{p_{t;1}^{\mathrm{I}}}{p_t^{\mathrm{I}}} \cdot \sigma_{t;1}^{\mathrm{I},B} \tag{31}$$

In case II, the qubits between Alice and Bob are disentangled. Therefore, the bit sent by Alice carries no information about Bob's own state. Thus Bob's reduced density matrix remains unchanged.[2] Thus we have

$$\sigma_t^{\mathrm{II},B} = \sigma_{t;0}^{\mathrm{II},B} = \sigma_{t;1}^{\mathrm{II},B} \tag{32}$$

By inductive hypothesis, we have

$$p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},B} \preceq \sigma_t^{\mathrm{II},B} \tag{33}$$

Combining (31), (32), and (33), we have

$$p_{t;0}^{\mathrm{I}} \cdot \sigma_{t;0}^{\mathrm{I},B} \preceq p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},B} \preceq \sigma_t^{\mathrm{II},B} = \sigma_{t;0}^{\mathrm{II},B} \tag{34}$$

$$p_{t;1}^{\mathrm{I}} \cdot \sigma_{t;1}^{\mathrm{I},B} \preceq p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},B} \preceq \sigma_t^{\mathrm{II},B} = \sigma_{t;1}^{\mathrm{II},B} \tag{35}$$

So the inductive case is proved. ■

[2] We assume that Alice and Bob don't erase any information during the protocol.