On the Communication Complexity of Correlation and Entanglement Distillation

Ke Yang

May 4th, 2004

CMU-CS-04-136

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee

Steven Rudich, Chair
Avrim Blum
Robert Griffiths
Andris Ambainis

Submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

This work is partially supported by NSF grant NSF-ITR 0122581 through the Aladdin Center and the CMU SCS Alumni Fellowship.

Keywords: correlation distillation, entanglement distillation, communication complexity, EPR pairs, quantum key distribution

Abstract

One of the recurring themes in information theory and quantum information theory is correlation corruption and correlation recover. Correlation corruption refers to the situation where Alice and Bob share information that is not perfectly correlated (or perfectly entangled, if they share quantum information). Correlation corruption arises in many natural situations, including transmitting information through a noisy channel, measuring a noisy signal source, quantum decoherence, and adversarial distortion. Correlation recovery refers to the action Alice and Bob takes to "restore" the correlation/entanglement by agreeing on some perfectly correlated/entangled information.

Traditionally correlation repair is done via a preventive strategy, namely error correction. Using this strategy, Alice encodes her information using an error correcting code or a quantum error correcting code before sending it through a noisy channel to Bob, who then decodes and recovers the original information. Error correcting codes and quantum error correcting codes are extremely useful objects in information theory with numerous applications in many other areas of science and technology. They are well studied and well understood. However they have limitations. We shall show that some assumptions used by error correction are not sound in many scenarios and make the preventive strategy unsuitable.

We study the alternative strategy of correlation repair, known as the reparative strategy. Using this strategy, Alice and Bob start by sharing imperfectly correlated (raw) information, and then engage in a protocol to "distill" the correlation/entanglement via communication. We call these protocols (classical) correlation distillation protocols and (quantum) entanglement distillation protocols. We show that such a reparative strategy can be as efficient as the preventive strategy. Furthermore, the reparative strategy is more flexible, in that it doesn't have the limitations suffered by error correction. We also point out that in particular, quantum entanglement distillation

protocols play a very important role in quantum information theory. Despite the significance of these protocols, they have received much less attention than error correcting codes and are much less well understood.

We focus on the communication complexity of the correlation and entanglement distillation protocols. In designing error correcting codes, efficiency is one of the main concerns. One wants to construct an error correcting code with the least possible redundancy while being able to withhold the highest rate of noise. In correlation and entanglement distillation protocols, the efficiency is measured by the amount the communication between Alice and Bob, and thus it is important to design protocols with minimal amount of communication. Our study is concerned with the minimal amount the communication needed for distillation.

We present a number of results concerning communication complexity for protocols over various noise models, which are mathematically models for different types of correlation corruption. These results span both classical and quantum information theory, and have connections to other areas of computer science, including cryptography and computational complexity.

Acknowledgment

I owe a lot to my advisor, Steven Rudich, who first introduced me to the fascinating area of theoretical computer science, which I enjoy greatly ever since. Steven taught me how to do research: how to understand a problem from a fundamental level, how to refine the answers relentlessly, how to spell out the core idea of a solution, and how to strike a balance between being intuitive and being rigorous. An excellent speaker himself, Steven also taught me how to give good talks, which I found extremely useful. Steven is more than just an academic advisor to me, I found his advices on many other topics very enlightening as well: philosophy, life, cooking, magic, to name a few.

Manuel Blum is "personally responsible" for bringing me to Carnegie Mellon and I am very grateful for it. I am fortunate to be in the group of people who work with him and are deeply inspired by his unique way of thinking and conducting research. I can only describe his style as "magic."

Avrim Blum is unique in his own way. Avrim seems to possess the super-human ability to look through a seemingly messy and complicated problem and see its mathematical essence, and then explain it in a very intuitive way, as if the problem were indeed that simple. He is also a master in finding the connection between apparently unrelated problems. I will forever remember the things I learned from him.

I started to be interested in the field of quantum information theory and quantum computation because of Andris Ambainis, whose talents in mathematics serves as a strong inspiration. Two of the three papers this thesis is based on are collaboration with him, and I cherish the wonderful experience. It is an honor to have him in my thesis committee. My formal learning of the quantum world would be impossible without Bob Griffiths, who taught a course at CMU. Bob is one of the very few physicists I know that can explain quantum in such a clear way such that even a computer

scientist can understand. The central idea of this thesis grows from a course project I did in his class. It is only appropriate to have Bob in my thesis committee, and I am very happy that he agrees.

I thank my collaborators, both inside CMU and out, for the advices and ideas they contributed selflessly: Luis von Ahn, Leemon Baird, Nina Balcan, Alex Gray, Nick Hopper, Ning Hu, Russell Impagliazzo, Jeff Jackson, Adam Kalai, Lea Kissner, John Langford, Andrew Moore, Alina Oprea, Bartosz Przydatek, Mike Reiter, Rachel Rue, Adam Smith, Dawn Song, Wei Xu, Li Zhang (at HP labs), Li Zhang (at University of Washington), and Jerry Xiaojin Zhu.

I remember that in year 1998, I spent a lot of time deciding which graduate school to attend, and I finally chose CMU. Now, six years after, I feel only how lucky I was in making the right choice. Indeed, CMU has the ideal atmosphere for me — nice and friendly people, open and relaxing environment, tons of free food and numerous foosball games in the lounge... People say CMU students are so happy that they don't want to graduate — well, they are right to some extent. People do graduate, but yes, most of us are very, very happy here. I will surely miss this place, as well as its people: Sharon Burks, Jeanette Wing, Catherine Copetas, Nikhil Bansal, Ashwin Bharambe, Mihai Budiu, Shuchi Chawla, Kedar Damdhere, Mihim Mishra, Amit Manjhi, Yue Pan, Francisco Pereira, Minglong Shao, Mengzhi Wang, Weng-Keen Wong, Hua Yu, Joy Zhang, Antonia Zhai, and Hua Zhong, to name a few. They are the ones responsible for my personal happiness at CMU.

I spent one summer at Akamai and two summers at Bell Labs, Lucent Technologies as an intern. These experiences are eye-openers to me, offering to me a unique opportunity to see how people outside universities do "real" work and do research. I enjoy all my intern experiences and I thank my co-workers there that make it possible: Gabe Loh, Harald Prokop, Ramesh Sitaraman, Bin Song, Juan Garay, Eric Grosse, Phil MacKenzie, Seny Kamara, and Gabe Plunk.

Finally, my infinite thanks to Ting Liu, who is obviously my most significant discovery in life.

Contents

List of Figures

List of Theorems

Chapter 1

Introduction

We introduce the notion of correlation distillation and entanglement distillation. We also discuss their motivations and related work.

1.1 Correlation Corruption and Correlation Repair

Information theory, since its inception in 1948 by Claude Shannon in his groundbreaking paper [82], has developed into a rich field of research, with applications in a broad spectrum of areas, including electrical engineering, computer science, statistics, and physics. From the 1970s, as researchers started to understand quantum mechanics, the field of quantum information theory emerged as a natural extension to the classical information theory. Exciting (and sometimes confusing) results were discovered, such as the EPR paradox (that two quantum states can be space-separated yet entangled, such that their measurements will be correlated), the non-cloning theorem (that quantum information cannot be duplicated), and teleportation (that Alice and transmit an unknown quantum state to Bob by sending two classical bits). Not only did quantum information theory contribute to the development of quantum mechanics, it also found applications in "traditional" areas, such as cryptography.

One of the most recurring themes in information theory is *correlation corruption* and *correlation recovery*. Correlation corruption refers to the situation where Alice and Bob share some information which is not perfectly "correlated". Classically this means that with positive probability, Alice's bits doesn't agree with Bob's. Quantum mechanically, this means that Alice's quantum state isn't perfectly entangled with Bob's quantum state. Researchers have striven to understand the nature of correlation corruption and constructed various mathematical models for it; we call them *noise models*. On the other hand, correlation recovery refers to the action Alice and Bob take to "restore" the correlation (or entanglement) to the maximum. Naturally, the goal is to **perform correlation repair, using as little resource as possible**.

We discuss some situations where the theme of correlation corruption and correlation recovery occurs naturally.

1.1.1 Information Transmission

Perhaps the most well-known problem in information theory is to transmit information through a noisy channel. In fact, it was considered in Shannon's original paper [82] and remains one of the most important topics in information theory. When Alice sends information to Bob through a *noisy channel*, the channel can "distort" the information. More concretely, suppose Alice sends classical bits to Bob, a classical noisy channel may flip some of the bits (a bit "0" becomes "1", and a bit "1" becomes "0"), or erase some of the bits (a bit becomes "$\perp$", a special symbol indicating the loss of the bit); suppose Alice sends qubits to Bob, a quantum noisy channel may apply a "bit-flip" (normally denoted by X) which switches $|0\rangle$ and $|1\rangle$, a "phase-shift" (normally denoted by Z), which keeps $|0\rangle$ unchanged but changes $|1\rangle$ to $-|1\rangle$, or a bit-flip composed with a phase-shift (normally denoted by Y). If Alice keeps a copy of the information she sends to Bob, then the noisy channel certainly can corrupt the correlation between Alice and Bob. A large part of information theory is to understand the nature of these noisy channels and devise mechanisms to fight the noise, namely, to perform correlation recovery.

1.1.2 Random Beacons

A random beacon is an entity that broadcasts uncorrelated unbiased random bits. The concept of random beacons were first introduced in 1983 by Rabin [74], who showed how they can be used to solve problems in cryptography. Bennett, DiVincenzo, and Linsker [26] proposed to use a random beacon to authenticate video recording. Maurer [60], Aumann and Rabin [6], and Ding [33] proposed to use a random beacon of extremely high rate to build information-theoretically secure cryptographic primitives, e.g., key exchange, encryption, and oblivious transfer. von Ahn et. al. [2] discusses various applications of random beacons, including verifiable lotteries and proof of ignorance.

There are many proposals to construct a *public, verifiable* random beacon, among them are the ones that use the signals from a cosmic source [2, 63]. In these proposals, Alice (as the beacon owner) and Bob (as a verifier) both point a radio telescope to some extraterrestrial objects, e.g. pulsars, and then measure the signal from them, which presumably contains enough randomness. However, it is inevitable that Alice and Bob have discrepancy in their results, due to measurement errors. Nevertheless, Alice and Bob still wish to agree on some common random bits, or, in other words, to recover the correlation between them. Notice that the random bits they wish to agree on are not necessarily the "raw data" from the measurement. Alice and Bob are free to apply any transformation to their measurement results.

1.1.3 Distilling EPR Pairs

An EPR pair, or an Einstein-Podolsky-Rosen pair [35], is a qubit pair in the state $\frac{1}{\sqrt{2}}(|\,00\rangle + |\,11\rangle)$ shared by two parties, with one party (Alice) holding one quantum bit and the other party (Bob) holding the second bit. EPR pairs are maximally entangled states and play a very important role in quantum information theory. Using an EPR pair, Alice and Bob can perform quantum teleportation. By performing only local operations and classical communication (often abbreviated as "LOCC"), Alice can "transport" a qubit to Bob, who could be miles away from Alice [18]. So EPR pairs, along with a classical communication channel, effectively constitute a quantum channel. Conversely, "superdense coding" is possible with EPR pairs: if Alice and Bob share an EPR pair, then Alice can transport two classical bits to Bob by just sending one qubit [29]. Therefore, it is very desirable for Alice and Bob to pre-manufacture a large number of EPR pairs and store them. In this way, they only need to maintain a classical channel between them, which is much more economical than a quantum channel, to transmit quantum information.

However, it is very hard to store qubits, since they can easily become entangled with the environment and *decohere*. Moreover, the decoherence happen continuously with time, and it is hard to prevent with current technology. This poses a serious problem to teleportation, since teleportation needs perfect EPR pairs, and if EPR pairs cannot be stored almost perfectly, teleportation would not be useful. Therefore, Alice and Bob need to "distill" almost perfect EPR pairs from the noisy ones, or, in other words, to "recover" the entanglement.

1.1.4 Quantum Key Distribution

Consider the quantum key distribution protocols by Bennett and Brassard [16], and by Bennett [13]. In these protocols, Alice randomly produces a sequence of qubits and send them to Bob, who then measures these qubits. If Alice keeps a copy of the qubits she sends to Bob, then Alice and Bob will share a number of perfectly entangled states. Next, Alice and Bob can exchange information to agree on some random bits, which then will be used as their shared key. However, Eve, the eavesdropper, might intercept some of the qubits Alice sent and distort them. This distortion caused by Eve will result in imperfectly entangled states between Alice and Bob. Therefore, they need to recover from the imperfect entanglement and agree on almost perfectly entangled states, or EPR pairs.

1.2 Error Correction: the Preventive Strategy

The most popular strategy to correlation repair is though the means of Error Correcting Codes (ECCs) and Quantum Error Correcting Codes (QECCs). Consider the situation of transmitting information through a noisy channel. Alice can *encode* her information using an *error correcting code*, or a *quantum error correcting code* into a *code-word*, before sending it to Bob. Then Bob can *decode* the noisy code-word and recover the information. See Figure 1.1. We call this the "preventive" strategy, since preventive measures are taken before the corruption takes place.

Error correcting codes and quantum error correcting codes have long been central objects of study in the field of information theory, and they have received tremendous amount of attention. Moreover, not only are they extremely useful in information theory, they also found numerous applications in other fields, including combinatorics, cryptography, and computational complexity. However, they have their limitations, and we discuss some of these limitations below.

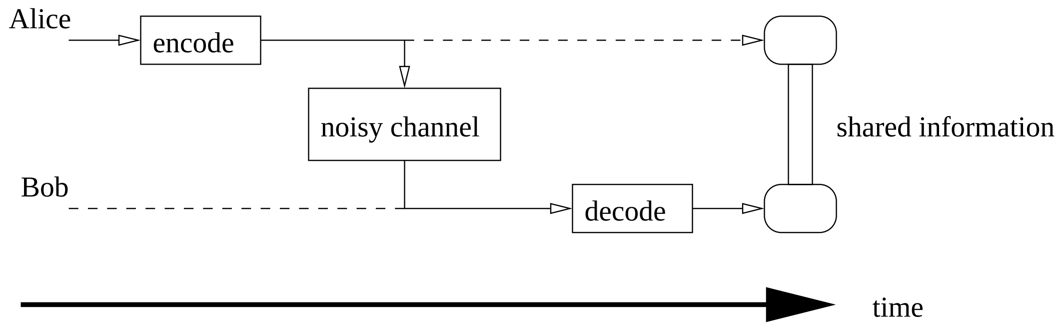

Figure 1.1: The preventive strategy for correlation repair.

Timing Constraint

First of all, there is the *timing constraint.* Error correction codes only work if Alice can encode the information *before* the noise takes place, which is not always possible. Consider the random beacon where Alice and Bob measure the noisy signals from a pulsar. In this case, it is impossible to encode the signal from the pulsar and thus error correction becomes totally useless.

Assumptions on Noise Model

Moreover, almost all research work on error correcting codes focuses on a relatively limited noise model, which we call the *identical independent distortion (IID).* In this model, the information is transmitted in units (e.g. bits or qubits) through a noisy channel, which applies a "distortion" process to each of the units independently. Examples of the deformation process include "flip a bit with probability ϵ" (which corresponds to the Binary Symmetric Channel), "change a bit to $\perp$ with probability ϵ" (which corresponds to the Erasure Channel), and "replace a qubit by a a completely mixed state with probability ϵ" (which corresponds to the Depolarization Channel). Two important assumptions in the IID model is that: 1) the deformation processes are identical to each unit; 2) the processes are independent. These two assumptions greatly simplify the problem of error correction, since the Law of the Large Numbers can be used. One can thus separate the so-called "typical error syndromes" from the "atypical" ones, and only focus on the typical syndromes. However, it is not always realistic to assume the IID model. This is best illustrated by the case of quantum key distribution protocols. Recall in this situation, Eve may intercept some qubits sent by Alice and cause distortion. Notice Eve is adversarial in nature and there is no reason to assume the the noise she causes is IID. Therefore, quantum error correction is not suitable in this case.

As a comment, we point out that Shor and Preskill [84] in fact used a particular class of quantum error correcting codes (known as CSS codes) in the analysis of security of the BB84 protocol. In particular, they showed that this class of QECCs, which were originally designed to work in a so-called "bounded corrupt" noise model, work in the so-called "fidelity" noise model as well. Here, the fidelity model is adversarial and is suitable for the quantum key distribution protocol. However, this appears to be a coincidence, and there is no evidence that an arbitrary QECC designed for a non-adversarial model will automatically work for an adversarial one.

Assumptions on Noise Rate

Finally, error correcting assumes that the *noise rate* is known at the time of encoding, so that an appropriate encoding scheme with appropriate redundancy can be designed. Notice that the noise rate has to be determined before the noise actually takes place, and therefore one often has to *guess* the rate. If the guess is too high, then too much redundancy would be added and bandwidth wasted; if the guess is too low, then too little redundancy may cause the loss of information. Furthermore, there are situations where there simply is not a fixed noise rate. Take the decohering EPR pairs as an example. The decoherence happens continuously with time, and thus the noise rate is varying with time (more precisely, increases with time). In this case, it is rather inefficient and inflexible to use an quantum error correcting code of a fixed rate.

1.3 Correlation Distillation: the Reparative Strategy

Correlation Distillation Protocols (CDPs) and Entanglement Distillation Protocols (EDPs) provide an alternative strategy for correlation repair. In this strategy, Alice and Bob start by sharing imperfectly consistent information, and then "distill" near-perfect information via communication and local operations. See Figure 1.2. If it is the classical information Alice and Bob are to distill, we call the process a "correlation distillation protocol"; if it is the quantum information, we call it an "entanglement distillation protocol". Overall, we call the strategy the "reparative strategy".

As a technical note, we always assume that the communication in the protocols is classical and noise-free. It is a standard assumption that only classical communication is allowed in quantum

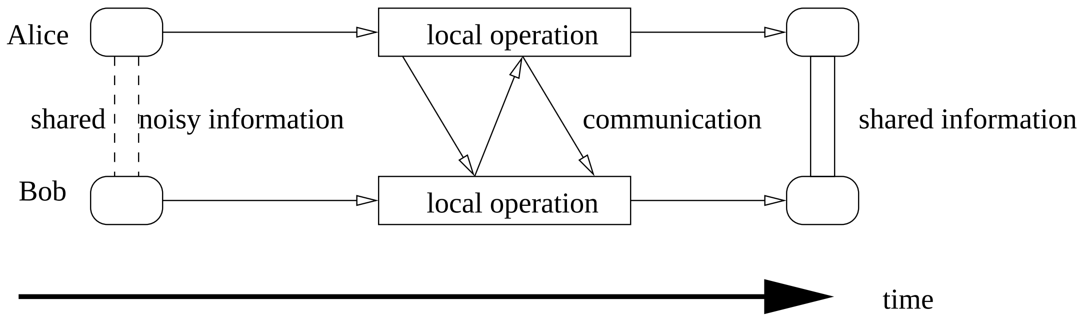

Figure 1.2: Reparative strategy for correlation repair.

entanglement distillation protocols, since quantum communication is considerably more expensive. These protocols that only involve local operations and classical communications are called "LOCC protocols", standing for "Local Operation Classical Communication". The assumption of noise-free communication can be justified in the following ways. First, the amount of communication is normally much smaller than the amount of the information Alice and Bob share, and thus they can afford to protect their communication either using a communication channel of higher quality or using error correcting codes of high redundancy. In this way, Alice and Bob can virtually assume noiseless communication. Second, much of the study in this thesis focus on the question of how much information Alice and Bob need to exchange in order to perform correlation/entanglement distillation, and the assumption of noiseless communication greatly simplifies the analysis. Finally, in the case of entanglement distillation, classical communication is used to distill quantum entanglement, and it is reasonable to assume a noise-free classical channel while the quantum channel might be noisy.

Correlation distillation protocols and entanglement distillation protocols solve several problems with error correcting codes and quantum error correcting codes. First, since the distillation takes place after the noise, there is no timing constraint for correlation/entanglement distillation. Therefore, CDPs are suitable for situations such as random beacons. Furthermore, since Alice and Bob perform distillation only after the correlation corruption, they can measure the noise rate first, and then choose the appropriate distillation protocol. This is more flexible and some times more desirable than error correction, which needs to guess the noise rate (for example, in the case of decohering EPR pairs). In fact, as we shall exhibit later in the thesis, there exist situations (both in classical and in quantum) where error correction almost completely fails while it is still possible to do correlation distillation (see Section 4.3). Finally, as we shall discuss later, CDP/EDPs admit a broader range of noise models, and in particular, noise models that are not identical independent distortion. In particular, while QECCs are not appropriate for quantum key distribution protocols, where the noise model is adversarial, EDPs turned out to be the perfect solution, as pointed out by Lo and Chau [57] and Shor and Preskill [84] (they used the term "entanglement purification protocols" for EDPs). [1]

Besides the "practical" advantages of EDPs, they have great theoretical importance in quantum information theory. Quantum entanglement plays a crucial role in quantum information and researchers have striven to understand entanglement, and in particular, ways to measure the amount the entanglement as a physical resource. Among various proposals is the concept of *distillable entanglement*[25]. For a quantum state ρ, its distillable entanglement is defined to be asymptotically the ratio of the amount of EPR pairs that can be produced by the optimal EDP from n copies of state ρ over n, as n increases. Clearly, the study of entanglement distillation protocols is closely related to that of entanglement.

If we compare the two approaches to information agreement, ECC/QECC and CDP/EDP, perhaps the most salient difference between them is that ECC/QECCs are algorithms performed by a single party (Alice for encoding and Bob for decoding), while CDP/EDP are two-party protocols that involve communication. In designing ECC/QECCs, the *overhead* is one of the main concerns and the goal is to design ECC/QECCs with as low as possible overhead that can withstand an as high as possible noise rate. For CDP/EDPs, the overhead is the amount of communication between Alice and Bob, i.e., the number of bits exchanged between them. Therefore, the *communication complexity* of CDP/EDPs is one of their most important parameters.

1.4 Our Contributions

In this thesis, we study the communication complexity of correlation and entanglement distillation protocols. Since CDP/EDPs are protocols, they are more complicated objects than ECC/QECCs. For example, with protocols, one might want to distinguish *one-way* communication, where only

[1] In fact, Shor and Preskill used CSS codes, which are a special class of quantum error correcting codes, in their proof. See the discussion before.

Alice sends information to Bob, who never sends anything back, from *two-way* communications, where Alice and Bob exchange bits. A protocol can be *deterministic*, where both Alice and Bob are deterministic, *randomized*, where Alice and Bob can have their own supply of random bits, or *randomized public-coin*, where Alice and Bob share a common random source.[2] It is the focus of this thesis to study various type of CDP/EDPs over a large range of noise models.

We briefly summarize a collection of results contained in this thesis. The ones marked with a star ($\star$) are the major results.

1. **A Relation Between ECC/QECCs and CDP/EDPs**

   We relate a large class of error correcting codes and quantum error correcting codes to correlation distillation protocols and entanglement distillation protocols. More precisely, we point out that every linear ECC corresponds to a CDP over the same noise model with the same overhead, and every stabilizer QECC corresponds to an EDP over the same noise model with the same overhead. See Theorem 4.1 and Theorem 4.4. Furthermore, we prove that their exist natural noise models where CDP/EDPs overperform ECC/QECCs. See Theorem 4.5, Theorem 4.7, and the discussions in Section 4.3.

2. ($\star$) **Impossibility Resulta for Non-Interactive Correlation Distillation**

   We show several general impossibility result for non-interactive correlation distillation over a number of natural noise models, including the binary symmetric model, the binary erasure model, and the extensions. We also show how this result is related to various research areas, including random beacon and information reconciliation. See Theorem 5.1, Theorem 5.2, Theorem 5.3, and Theorem 5.4.

3. **A Positive Result on One-bit Correlation Distillation**

   We present a positive result where Alice and Bob, by exchanging one bit of information, can perform correlation repair, which would be impossible without communication. This shows that even the minimal amount of communication can help in correlation repair. See Theorem 6.1.

[2]We are using the notations from Kushilevitz and Nisan [52].

4. ($\star$) **An Impossibility Result of Non-Interactive Entanglement Distillation**

   We show several impossibility results for non-interactive entanglement distillation, where Alice and Bob wish to produce near-EPR pairs without communication. These are the first results in the area of communication complexity of EDPs, and they provide the first step in understanding entanglement distillation protocols. See Theorem 7.1, Theorem 7.2, Theorem 7.3.

5. **An Impossibility Result of EDPs over the Entanglement Noise Model**

   We prove an impossibility result on entanglement distillation over the so-called "entanglement noise model". We show that it is impossible to distill EPR pairs from an arbitrarily entangled quantum state. We show how this result is related to classical randomness extractors. See Theorem 8.1.

6. ($\star$) **A Complete Characterization of EDPs over the Fidelity Noise Model**

   We completely characterize the communication complexity of entanglement distillation protocols over the so-called "fidelity noise model". We present a protocol that distills near-perfect EPR pairs very efficiently, and prove such a protocol is in fact optimal (up to an additive constant). We also show how this noise model is related to other areas of quantum information theory, including purity-testing protocols [23] and quantum key-distribution protocols [57, 84]. See Theorem 8.2, Theorem 8.3, Theorem 8.4 and Theorem 8.5.

These results appear in the following publications.

1. A. Ambainis, A. Smith, —.

   Extracting Quantum Entanglement (General Entanglement Purification Protocols).

   Appeared in the *IEEE Conference of Computational Complexity (CCC 2002)*, Montréal, Quebéc, Canada, pp. 103-112, 2002.

2. —.

   On the (Im)possibility of Non-interactive Correlation Distillation.

   Appeared in the *Latin American Theoretical INformatics (LATIN 2004)*, Buenos Aires, Argentina, 2004.

3. A. Ambainis, —.

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

To appear in the *19th Annual IEEE Conference of Computational Complexity (CCC 2004)*, Amherst, MA.

	noise model	communication: 0	communication: 1	communication: many
classical	bounded corruption			L
classical	binary symmetric	😐 U	☺ L	L
classical	binary erasure	☺ U		L
classical	tensor product	☺ U		
quantum	bounded corruption	☺ U		L
quantum	bounded measurement	☺ U		L
quantum	depolarization	☺ U		L
quantum	entanglement	😐 U	😐 U	😐 U
quantum	fidelity	☺ L U	☺ L U	☺ L U

L =	lower bound
U =	upper bound
☺ =	my orignal result
😐 =	independent result

Figure 1.3: Summary of known results

We summarize all the results in this thesis in a table in Figure 1.3. Each row in the table corresponds to a noise model, and each column corresponds to the amount of the communication allowed for a protocol. In each cell, we put the known upper and lower bounds on the "quality" of the best known CDP/EDPs. The ones with a smiley face indicates my original contributions, and the ones with a non-smiley face indicates my discovery that are independent from other researchers. The blanks indicate open problems.

1.5 Related Work

We discuss some related work on correlation distillation and communication complexity.

1.5.1 Error Correction

As we discussed before, error correction is closely related to correlation distillation protocols. Error correction is the preventive strategy for correlation recover, and correlation distillation is the reparative strategy.

Not only are error correcting codes extremely useful in information theory, they have also found numerous applications in other fields, including combinatorics, cryptography, and computational complexity.

Error correction has received a tremendous amount of attention. Because of its sheer volume, it is impossible to give an (even remotely) comprehensive list of the literature on this topic. I only list a few items. Shannon [82] is the first one to consider the problem of error correction, and his paper marked the beginning of the field of information theory. Blahut [12] has a wonderful book completely dedicated to error correcting codes and contains abound resources. Sudan [86] has a very nice survey on ECCs that is more tailor-made for audiences in computational complexity. Shor [83] and Steane [85] are the first to study quantum error correcting codes and to actually construct them. Gottesman's thesis [36] is a great source for the theory behind quantum error correcting codes with many results. Nielsen and Chuang's book [69] also gives a nice description on both classical and quantum error correction.

1.5.2 Two-party Coin-flipping

Two-party coin-flipping is a classical problem in cryptography, where Alice and Bob wish to establish some commonly agreed random bits by communication. Blum [10] is the first to study the setting where Alice and Bob initially don't share any information and one of them could be cheating. He suggested protocols that are secure against a computationally-limited adversary, based on number-theoretical assumptions. Following Blum's work, Lindell [54] studied the parallel version of the problem under the same setting. Barak [9] consider the two-party coin-tossing resistant to the man-in-the-middle attack. On the other hand, researchers have studied quantum coin-flipping, where Alice and Bob exchange quantum information and agree on a classical bit. For results in this area, see [56, 62, 1, 4, 88, 53]. Classical two-party coin-flipping is a special version of correlation distillation protocols with the assumption that: 1) the players do not share any prior information; 2) they are polynomial-time bounded; and 3) they don't necessarily collaborate and are liable to cheating. As a result, the protocols for two-party coin-flipping rely on cryptographic assumptions and the communication complexity is higher than the number of coin flips they agreed on. Quantum two-party coin-flipping, however, does not fit into this thesis, since it requires a quantum channel between Alice and Bob.

1.5.3 Information Reconciliation

Information reconciliation is an extensively studied concept [17, 61, 27, 30, 31] with applications in quantum cryptography and information-theoretical cryptography. In this setting, Alice and Bob each possesses a sequence of random bits that agree "most of the time". Here the "agreement" between Alice's bits (denoted by A) and Bob's bits (denoted by B) is described by the mutual information $I(A;B)$. Moreover, Eve, the eavesdropper, also possess some information (denoted by Z) about the bits held by Alice and Bob, which is quantified by the mutual information $I(Z;AB)$. Alice and Bob wish to "reconcile" their information (namely, to agree on some random information) by communicating in a public channel (which is noiseless but readable by Eve). Their goal is to agree on a common random string U with very high probability, while ensuring that Eve gains little information from U. In terms of the entropy, let C be the communication between Alice and Bob, then we should have $H(U|AC) \approx 0$, $H(U|BC) \approx 0$, and $I(U;ZC) \approx 0$. Information reconciliation and correlation distillation operate in similar models: Alice and Bob share noisy information, and then communicate to agree on something with higher correlation. However, the primary concern for information reconciliation is *privacy*, i.e., that Eve gains little information about the information agreed upon, while this thesis focus on the communication complexity.

1.5.4 Quantum Entanglement Distillation

As we mentioned before, quantum entanglement distillation protocols are two-party protocols involving only local (quantum) operation and classical communication. These protocols generally takes some entangled bipartite states as input and output near-perfect EPR pairs. The process of entanglement distillation was also known as "entanglement concentration" or "entanglement purification".

There have been a lot of research efforts on studying entanglement distillation protocols [21, 22, 25, 43, 44, 75, 76, 77, 7]. Different "noise" models on the imperfect EPR pairs are presented and studied.

To the best of our knowledge, Bennett, Bernstein, Popescu, and Schumacher are the first to consider the problem of producing EPR pairs from "less entangled" states. In their seminal paper [21], they give a protocol that converts many identical copies of pure state $|\phi\rangle = (\cos\theta|01\rangle + \sin\theta|10\rangle)$ to perfect EPR pairs. They call this process "entanglement concentration". In the same year, Bennett, Brassard, Popescu, Schumacher, Smolin, and Wootters [22] studied the problem of "extracting" near-perfect EPR pairs from identical copies of mixed entangled states. This is the first time that the notion "entanglement purification protocols" was presented, which were renamed to "entanglement distillation protocols" later. They also pointed out that EDPs can be used to send quantum information through a noisy channel. Later, Bennett, DiVincenzo, Smolin and Wootters [25] improved the efficiency of the protocols in [22] and proved a result that closely related EDPs to quantum error correcting codes, which is an alternative means to transmit quantum information reliably through a noisy channel. Horodecki, Horodecki, and Horodecki [42, 45] and Rains [75, 76, 77] give various asymptotic bounds on distillable entanglement for arbitrary entangled states. They considered 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, for example, Vidal [90], Jonathan and Plenio [49], Hardy [41], and Vidal, Jonathan, and Nielsen [91]. Much of the work was built on the result of majorization by Nielsen [67], who is the first one that studied conditions under which one pure state can be transformed into another one by LOCC.

From another direction, researchers have studied EDPs with *incomplete information*, where Alice and Bob do not know the exact state they share. The state is in a mixed state, or is prepared adversarially. In this case we cannot hope that Alice and Bob would act optimally. However, there still exist protocols that do reasonably well. Bennett *et. al* [22, 25] studied the model where Bob's share in the EPR pairs underwent a noisy channel, resulting in a mixed state. They showed that their protocol would "distill" near-perfect EPR pairs even when Alice and Bob do not have the complete knowledge of the shared state. Under another circumstance, "purity-testing protocols" were studied implicitly by Lo and Chau [57], Shor and Preskill [84], and later explicitly by Barnum, Crépeau, Gottesman, Smith, and Tapp [23]. Purity-testing protocols are LOCC protocols that approximately distinguish the state of perfect EPR pairs from the rest states. Ambainis, Smith, and Yang [7] pointed out that purity-testing protocols are indeed EDPs where Alice and Bob only know the *fidelity* of the state they share. Using constructions from [23], Ambainis, Smith and Yang constructed a "Random Hash" protocol that produces $(n - s)$ EPR pairs of conditional fidelity at least $1 - \frac{2^{-s}}{1-\epsilon}$ on any n qubit-pair input state of fidelity $1 - \epsilon$. Their protocol would fail with probability ϵ, and the conditional fidelity of its output is the fidelity *conditioned on* the protocol not failing.

Much of previous work assumes that Alice and Bob have the complete information about the state they share, and thus they can act *optimally*. The main focus of the majority of the previous work is the *yield* of the protocols, i.e., the question "how many EPR pairs can be extracted from the input state, using *unlimited* classical communication?" Lately, there has been work that start to study the communication complexity of EDPs, started by Lo and Popescu [58] and followed by Ambainis and Yang [8]. Here the question is "how many bits need to be exchanged in order to distill n EPR pairs?" In the thesis, I continue this line of research on the communication complexity of EDPs with the focus on the situation where Alice and Bob have *incomplete* information about their shared states.

1.5.5 Communication Complexity

Classical communication complexity studies the minimal amount of classical information (typically measured in bits) needed to be transmitted between multiple parties in order to collectively perform a certain computation. The results are typically information theoretical, and do not rely on any un-proven assumptions. The field of communication complexity was pioneered by Yao [94], and now is a very rich field in theoretical computer science, and has found applications in many areas, like network analysis, VLSI design, data structure, and computational complexity. The readers are referred to [52] for a nice introduction and tutorial.

Quantum communication complexity mostly studies the minimal amount of quantum information (typically measured in qubits) needed to be exchanged in order to perform some (classical or quantum) task. This field was also first studied by Yao [95], and now it is becoming one of the main topics in quantum information theory. It is a very successful area, and numerous results have emerged. In fact, most known lower bounds in quantum computation can be regarded as communication complexity results. We refer the readers to [15] for a nice survey, and [19, 50, 51, 78] for some important techniques and results.

Despite the numerous results emerging from classical and quantum communication complexity, another class of problem, namely the *classical* communication complexity for *quantum* protocols, has being largely ignored until very recently. This class of problem is concerned with the minimal number of classical bits needed to be communicated to perform certain quantum tasks. An example is the classical communication complexity for EDPs. One may ask "how many bits do Alice and Bob need to exchange in order to distill n EPR pairs?" One reason that not many researchers pay too much attention to this problem might be the conception that classical communication is "cheap" compared to quantum communication, and thus one can assume they are free. However, as pointed by Lo and Popescu [58], there are situations where classical communication can not be justifiably ignored. One example is the super-dense coding [29]. Alice and Bob can use n qubits to transmit $2n$ bits of classical information, if they previously share n EPR pairs. Nevertheless, if it takes more than n bits of classical communication to distill the n EPR pairs, it would completely destroy the purpose of super-dense coding. Furthermore, in the study of LOCC protocols over quantum states, no quantum communication takes place, and it is therefore interesting to study the classical communication complexity of these (quantum) protocols.

The history of classical communication complexity for quantum protocols can probably traced back to the seminal paper by Bennett and Wiesner [29], which discussed teleportation and constructed a protocol that uses $2n$ classical bits to transmit n qubits. However, this topic was largely overlooked until the work by Lo and Popescu [58] and Lo [55]. Lo and Popescu [58] discussed the classical communication complexity of various protocols by Bennett et. al. [21]. They observed that the "entanglement concentration protocol" in [21] does not require any classical communication. However, the "entanglement dilution protocol", which transforms m EPR pairs into n copies of less entangled qubit pairs, requires $O(n)$ bits of classical communication. Lo and Popescu then constructed a new dilution protocol that only uses $O(\sqrt{n})$ bits of communication. This protocol was proven to be asymptotically optimal independently by Hayden and Winter [47], and Harrow and Lo [46], who proved matching lower bounds for general entanglement dilution protocols. Lo [55] studied the communication complexity for Alice and Bob to jointly *prepare* many copies of arbitrary (known) pure states, and proved a non-trivial upper bound.

All the previous results focus on a relatively simple situation, where the input are identical copies of a known pure state, and only the asymptotic results are known. In this thesis, I study the communication complexity of EDPs with *incomplete information*. In this setting, Alice and Bob do not have the complete knowledge about the input state they share. Rather, the input state is a mixed state, or is adversarially prepared. I also study 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 a certain quality?"

Chapter 2

Quantum Mechanics and Quantum Information Theory

We introduce the notions and concepts in quantum mechanics and quantum information theory.

2.1 Quantum Mechanics

We briefly summarize the laws and conventions in quantum mechanics used in this thesis. This summary is by no means complete and we refer the reader to Peres [73] and Nielsen and Chuang [69] for a more comprehensive treatise.

2.1.1 The Quantum States and the Dirac Notation

A quantum system is described in a *Hilbert space*, i.e., a linear space with a well-defined inner product. In this thesis we only consider Hilbert spaces of finite dimension. We use $\mathcal{H}_N$ to denote a Hilbert space of dimension N. A *pure state* is described by a unit (column) vector in a Hilbert space $\mathcal{H}_N$ and is normally denoted in the so-called *Dirac notation* as $|\,\phi\rangle$. A *qubit* is a two-state quantum system (and is thus in a 2-dimensional space $\mathcal{H}_2$), and is also the smallest quantum state possible. A general qubit can be written as $|\,\phi\rangle = \alpha|\,0\rangle + \beta|\,1\rangle$, where α and β are complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$. We can view this general state $|\,\phi\rangle$ as a *superposition* of the two basis states $|\,0\rangle$ and $|\,1\rangle$. In general, a system of n qubits is described in a Hilbert space of dimension 2^n, which can be conveniently viewed as a tensor product of n two-state subspaces, i.e., $\mathcal{H}_{2^n} = \mathcal{H}_2 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_2$. We always assume the existence of a fixed, canonical *computational basis* in an N-dimensional Hilbert space, denoted as $\{|\,0\rangle, |\,1\rangle, ..., |\,N-1\rangle\}$, and a general pure state can be written as $|\,\phi\rangle = \sum_{x=0}^{2^n-1} \alpha_x |\,x\rangle$, where $\sum_{x=0}^{2^n-1} |\alpha_x|^2 = 1$. Naturally we have $N = 2^n$. Again, it is in general a superposition of 2^n basis states.

A "bra" is a unit row vector, defined as $\langle\phi\,| = (|\,\phi\rangle)^\dagger$, where $x^\dagger$ denotes the operation of applying transpose followed by the complex conjugate to x. For pure states $|\,\phi\rangle$ and $|\,\psi\rangle$, their inner product can be conveniently written as $(|\,\phi\rangle, |\,\psi\rangle) = \langle\phi\,| \cdot |\,\psi\rangle = \langle\phi\,|\,\psi\rangle$.

An *outer product* of two pure states $|\,\phi\rangle$ and $|\,\psi\rangle$ is a matrix defined as $|\phi\rangle\langle\psi| = |\,\phi\rangle \cdot \langle\psi\,|$.

The outer product and the inner product are conveniently related by the trace of a matrix.

$$\mathrm{Tr}(|\phi\rangle\langle\psi|) = \langle\psi\,|\,\phi\rangle \tag{2.1}$$

2.1.2 The Density Matrix and Mixed States

An alternative way to describe a pure state $|\,\phi\rangle$ is by its outer product with itself, $|\phi\rangle\langle\phi|$. This is known as the *density matrix* notation, and $|\phi\rangle\langle\phi|$ is the density matrix representing state $|\,\phi\rangle$. One advantage for the density matrix notation is that it can conveniently represent *mixed states*. A mixed state emerges when we do not have the complete information about a quantum system but only partial knowledge represented as a probabilistic distribution. More precisely, a mixed state is a probabilistic ensemble (mixture) of pure states. In Dirac notation, one writes a mixed state as $\{p_i, |\,\phi_i\rangle\}$, which means this state is in state $|\,\phi_i\rangle$ with probability p_i. Naturally, we have that $\sum_i p_i = 1$. In the density matrix notation, such a state is simply represented as

$$\rho = \sum_i p_i \cdot |\phi_i\rangle\langle\phi_i|. \tag{2.2}$$

It is easy to see that all density matrices are positive operators (i.e., they are Hermitians and all their eigenvalues are non-negative) and have trace 1. In fact, one can define a density matrix as one that is positive and have trace 1. Notice any such matrix can be written in the form of Eq. (2.2) by spectral decomposition.

Notice that there might exist two very different ensembles of pure states that yield the same density matrix. For example, consider an ensemble A which is state $|0\rangle$ with probability 0.5, and state $|1\rangle$ with probability 0.5. Its density matrix is $\rho_A = 0.5 \cdot |0\rangle\langle 0| + 0.5 \cdot |1\rangle\langle 1| = I/2$. Consider another ensemble B that is state $|\phi_+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ with probability 0.5 and state $|\phi_-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ with probability 0.5. It density matrix is $\rho_B = 0.5 \cdot |\phi_+\rangle\langle\phi_+| + 0.5 \cdot |\phi_-\rangle\langle\phi_-| = I/2$. So these two ensembles have the same density matrix, although they are formed very differently. However, by the laws of quantum mechanics, all the information one can obtain from a quantum system can be derived from its density matrix. Therefore, if two systems have identical density matrices, then there is no way to distinguish them. So the two ensembles A and B describe the same quantum system.

When studying a large quantum system, sometimes it is convenient to focus on a smaller "subsystem" within the large system. One can derive the *reduced density matrix* for the subsystem from the density matrix of the large system. Suppose the smaller system is in a Hilbert space $\mathcal{H}_A$ and the large system is in a Hilbert space $\mathcal{H}_{AB}$ with density matrix ρ. Then the density matrix ρ_A for the subsystem can be obtained by "tracing out" the system B, denoted by $\rho_A = \mathrm{Tr}_B(\rho)$. Here Tr_B is a linear operator defined as

$$\mathrm{Tr}_A(|a_0\rangle\langle a_1|^A \otimes |b_0\rangle\langle b_1|^B) = \langle b_0 \mid b_1\rangle \cdot |a_0\rangle\langle a_1| \tag{2.3}$$

Here we use superscript to denote the subsystem a state is in: $|a_0\rangle\langle a_1|^A$ is a state in subsystem A and $|b_0\rangle\langle b_1|^B$ is a state in subsystem B. It is possible that ρ is a pure state in the large quantum system AB, while the local density matrix ρ_A corresponds to a mixed state. In this case we say that state AB is *entangled.* Entanglement is one of the most important features in quantum mechanics and quantum information theory.

2.1.3 Quantum Operations

There are two types of operations that can be applied to a quantum system, namely unitary operations and measurements.

A unitary operation is a linear operator. For a quantum system of dimension N, such a linear operator can be naturally described as an $N \times N$ matrix U that maps a pure state $|\phi\rangle$ to $U|\phi\rangle$, and (equivalently) a mixed state ρ to $U\rho U^\dagger$. Such a matrix is *unitary*, if and only if $UU^\dagger = I$.

The laws of quantum mechanics stipulate that all unitary operations are allowed. Some of the most important unitary operations are single-qubit operators known as Pauli operators or Pauli matrices, denoted by X, Y, and Z, respectively, and defined as

$$X(\alpha|\,0\rangle + \beta|\,1\rangle) \;=\; \beta|\,0\rangle + \alpha|\,1\rangle \tag{2.4}$$

$$Y(\alpha|\,0\rangle + \beta|\,1\rangle) \;=\; -i\beta|\,0\rangle + i\alpha|\,1\rangle \tag{2.5}$$

$$Z(\alpha|\,0\rangle + \beta|\,1\rangle) \;=\; \alpha|\,0\rangle - \beta|\,1\rangle \tag{2.6}$$

The simplest version of measurements is a *projective measurement.* A *projector* is a linear operator P such that $P^2 = P$. An *observable* is an orthogonal decomposition of the identity operation. In other words, an observable is a collection of projectors $\{P_i\}$ satisfying $\sum_i P_i = I$. If one applies an observable $|\,\phi\rangle$ to a state $|\,\phi\rangle$, we have a projective measurement. A measurement is generally probabilistic: the resulting state is $\frac{P_i|\,\phi\rangle}{\sqrt{\langle\phi\,|P_i|\,\phi\rangle}}$ with probability $\langle\phi\,|P_i|\,\phi\rangle$. A measurement on a mixed state can be naturally generalized. A more general version of measurement, known as *POVM* ("Positive Operator-Valued Measurement"), is more conveniently described using the density matrix notation. A POVM is a collection of *measurement operators* $\{E_i\}$, where each E_i is a positive operation and we have $\sum_i E_i = I$. One may write $E_i = M_i^{\dagger} M_i$ for each i. The result of such a measurement on a quantum state ρ is state $\frac{M_i \rho M_i^{\dagger}}{\mathrm{Tr}(M_i^{\dagger} M_i \rho)}$ with probability $\mathrm{Tr}(M_i^{\dagger} M_i \rho)$. To see that POVM is indeed a more general notion, observe that it includes unitary operations as a special case. It can be shown, however, that any POVM can be realized by unitary operator and projective measurements with ancillary qubits.

The formalism of *super-operators* is used to describe how a quantum system evolve when interacting with its environment. A super-operator, normally denoted by $\mathcal{E}$, is a linear operator over density matrices defined as

$$\mathcal{E}(\rho) = \sum_i E_i \rho E_i^{\dagger} \tag{2.7}$$

where $\sum_i E_i^{\dagger} E_i \leq I$. We say $\mathcal{E}$ is *trace-preserving*, if $\sum_i E_i^{\dagger} E_i = I$.

2.2 Quantum Information Theory

We review some of the basic notions in quantum information theory. We do not attempt to give a complete or comprehensive survey on this topic. Again, the readers are refereed to Nielsen and Chuang [69] for more comprehensive treatise.

2.2.1 Entropy

The *entropy* of a quantum state ρ is denoted by $S(\rho)$ and known as the *von Neumann entropy.* It is defined as

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho) \tag{2.8}$$

where the logarithm is base-2.

It is not hard to derive from the definition that all pure states have entropy zero and the maximum entropy of an n-qubit system is n, which is achieved by the completely mixed state $\frac{I}{2^n}$.

2.2.2 Entanglement

In this thesis we will be mainly interested in bipartite systems shared between Alice and Bob. In such a bipartite system, the *entanglement* of a normalized pure state $|\,\phi\rangle$, denoted by $\mathsf{E}(|\,\phi\rangle)$, is defined to be the von Neumann entropy of the mixed state obtained by tracing out Bob's subsystem. In other words,

$$\mathsf{E}(|\,\phi\rangle) = S(\mathrm{Tr}_B(|\phi\rangle\langle\phi|)) \tag{2.9}$$

A pure state is *entangled* if its entanglement is non-zero, and is otherwise *disentangled* or *separable.* A mixed state is disentangled if it can be expressed as an ensemble $\{p_i, |\,\phi_i\rangle\}$ where each $|\,\phi_i\rangle$ is disentangled. All other mixed states are entangled. However, there isn't an agreed-upon definition on the amount of entanglement of a mixed state.

For a bipartite system consisting of n qubit pairs (or $2n$ qubits in total), its maximum possible entanglement is n. The most important among the maximally entangled states are the four *Bell* *states*, defined as

$$\Phi^+ = \frac{1}{\sqrt{2}}(|\,0\rangle^A|\,0\rangle^B + |\,1\rangle^A|\,1\rangle^B) \tag{2.10}$$

$$\Phi^- = \frac{1}{\sqrt{2}}(|\,0\rangle^A|\,0\rangle^B - |\,1\rangle^A|\,1\rangle^B) \tag{2.11}$$

$$\Psi^+ = \frac{1}{\sqrt{2}}(|\,0\rangle^A|\,1\rangle^B + |\,1\rangle^A|\,0\rangle^B) \tag{2.12}$$

$$\Psi^- = \frac{1}{\sqrt{2}}(|\,0\rangle^A|\,1\rangle^B - |\,1\rangle^A|\,0\rangle^B) \tag{2.13}$$

These are maximally entangled two-qubit pure states.

The Bell states are closely related to the Pauli matrices. In particular, it is easy to verify that unitary operators of the form $I \otimes U$, where $U \in \{X, Y, Z\}$ translates one Bell state to another. For example, we have $(I \otimes X)\,\Phi^+ = \Psi^+$, $(I \otimes Y)\,\Phi^+ = \Psi^-$, and $(I \otimes Z)\Phi^+ = \Phi^-$.

An *EPR pair*, or an Einstein-Podolsky-Rosen pair, refers to the Bell state Φ^+.[1] 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 $|\,\Phi_n\rangle$ and its density matrix $|\Phi_n\rangle\langle\Phi_n|$, when there is no danger of confusion.

2.2.3 Fidelity

The fidelity is a measure of the "closeness" of two quantum states. For two (mixed) states ρ and σ of equal dimension, their fidelity if defined as

$$\mathsf{F}(\rho, \sigma) = \mathrm{Tr}^2(\sqrt{\rho^{1/2}\sigma\rho^{1/2}}). \tag{2.14}$$

Notice we are using a different definition as in [NC00], where the *square root* of (2.14) is used.

If $\sigma = |\varphi\rangle\langle\varphi|$ is a pure state, the definition simplifies to

$$\mathsf{F}(\rho, |\varphi\rangle\langle\varphi|) = \langle\varphi\,|\rho|\,\varphi\rangle \tag{2.15}$$

A special case for the fidelity is when $|\,\varphi\rangle = \Phi_n$ for some n. In this case, we call the fidelity of

[1] There exist contexts where an EPR pair refers to the state Ψ^-. See, for example, Bohm [14]. But in this thesis, we use the convention of Φ^+.

ρ and $|\varphi\rangle$ simply the *fidelity of state* ρ, denoted as $\mathsf{F}(\rho)$. In other words, we have

$$\mathsf{F}(\rho) = \langle \Phi_n | \rho | \Phi_n \rangle \tag{2.16}$$

We are often interested in the fidelity of two states of unequal dimensions, and in particular, the fidelity of a general bipartite state ρ, and the Bell state Φ^+. If ρ has dimension 2, then this is simply $\mathsf{F}(\rho)$. However, when ρ has a higher dimension, we need to define its *base fidelity* as 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 $\mathsf{F}^{\mathsf{b}}(\rho)$. Mathematically, we have $\mathsf{F}^{\mathsf{b}}(\rho) = \mathsf{F}(\mathrm{Tr}_1(\rho))$.

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, as in the following claim.

Claim 2.1 *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*

$$\mathsf{F}(\rho, \sigma) = \sum_i p_i \cdot \mathsf{F}(|\phi_i\rangle\langle\phi_i|, \sigma). \tag{2.17}$$

The fidelity is also monotone with respect to trace-preserving operations [69].

Claim 2.2 *For any states ρ and σ and any trace-preserving operator $\mathcal{E}$, we have*

$$\mathsf{F}(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq \mathsf{F}(\rho, \sigma). \tag{2.18}$$

One useful fact is that the base fidelity of any completely disentangled state is at most 1/2.

Lemma 2.1 *If ρ is a completely disentangled state, then $\mathsf{F}^{\mathsf{b}}(\rho) \leq 1/2$.*

Proof: By the definition of base fidelity, we may assume that ρ has dimension 2. By Claim 2.1, 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

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

■

2.3 Some Useful Results

2.3.1 The Deviation of Pure States over Unitary Operations

We study how much "deviation" a quantum state undergoes when applied various unitary operations. In particular, we will prove two lemmas that would be useful in the rest of the thesis.

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

Lemma 2.2 *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*

$$\sum_{U\in\{I,X,Y,Z\}} |\langle\phi\,|U|\,\psi\rangle|^2 \leq 2 \tag{2.19}$$

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 2.1 *For any pure state* $|\,\phi\rangle$, $\sum_{U\in\{I,X,Y,Z\}} |\langle\phi\,|U|\,\phi\rangle|^2 \leq 2.$

Next, we consider quantum states and operations over bipartite systems, and study the "deviation" of a general bipartite state under unitary operations of the form $U\otimes U^*$, where U^* is defined as the complex conjugate of U, i.e., one simply takes the conjugate of each entry in U^*. Alternatively, U^* is defined as the unique unitary operation that satisfies that $U^*|\,\phi^*\rangle = (U|\,\phi\rangle)^*$. We interpret $U\otimes U^*$ as Alice applies U to her first qubit and Bob applies U^* to his first qubit. Again, we consider $U\in\{I,X,Y,Z\}$.

Lemma 2.3 *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*

$$\langle\phi\,|\,\phi\rangle + \langle\phi\,|(X\otimes X^*)|\,\phi\rangle + \langle\phi\,|(Y\otimes Y^*)|\,\phi\rangle + \langle\phi\,|(Z\otimes Z^*)|\,\phi\rangle = 4\mathsf{F}^{\mathsf{b}}(|\,\phi\rangle) \tag{2.20}$$

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 figure.

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

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^-

Figure 2.1: The Bell States under Pauli Operators

decompose $|\,\phi\rangle$ into the Bell basis and write

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

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 so,

$$\langle\phi\,|\,\phi\rangle + \langle\phi\,|(X \otimes X^*)|\,\phi\rangle + \langle\phi\,|(Y \otimes Y^*)|\,\phi\rangle + \langle\phi\,|(Z \otimes Z^*)|\,\phi\rangle = 4|\alpha_0|^2 = 4\mathsf{F}^{\mathsf{b}}(|\,\phi\rangle)$$

■

2.3.2 Positive Operators

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, $B \preceq A$.

Lemma 2.4 *For any 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.

Lemma 2.5 *Let ρ and σ be density matrices such that $\rho \succeq a \cdot \sigma$, for some positive number a. For any POVM $\{E_i\}$, where $E_i = M_i^\dagger M_i$, let $p_i = \mathrm{Tr}(\rho M_i)$ and and $q_i = \mathrm{Tr}(\sigma M_i)$ be the probabilities the measurement result being i for ρ and σ, respectively. Then we have $p_i \geq a \cdot q_i$.* ■

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

Chapter 3

Preliminaries and Notations

3.1 General Notations

We present some general notations, both classical and quantum, to be used throughout the thesis.

All logarithms are base-2. All vectors are column vectors by default. We use $[n]$ to denote the set $\{0, 1, ..., n-1\}$. If A and B are two sets, then $A \times B$ denotes the Cartesian product between sets A and B.

We often work with symbols from a particular *alphabet*, which is a finite set and is normally denoted by Σ. We always assume the existence of a canonical one-to-one correspondence between an alphabet Σ of size q and the set $[q]$, and often identify Σ with $[q]$.

A *string* is a sequence of symbols from an alphabet. We often identify a string with a vector and shall use them interchangeably. For a string x of length n, we use $x[j]$ to denote its j-th entry, for $j = 0, 1, ..., n-1$. We often also use a tuple to index an entry in a vector. For example, We index an $(a \cdot b)$-dimensional vector by (x, y), where $x \in [a]$ and $y \in [b]$. In this case, we assume there exists a canonical mapping from $[a] \times [b]$ to $[ab]$. We use $\mathbf{0}_n$ to denote the all-zero vector (whose each entry is 0) of dimension n, and $\mathbf{1}_n$ to denote the all-one vector (whose each entry is 1) of dimensional n. When the dimension is clear from the context, it is often omitted.

The *Hamming distance* between 2 strings x and y of equal length is the number of positions that these 2 strings differ, and is denoted by $\mathsf{dist}(x, y)$. For strings x and y, we use $x; y$ to denote the *concatenation* of these 2 strings.

A *binary string* or *binary vector* is a string over alphabet $\{0,1\}$. We identify an integer with the binary vector obtained from its binary representation. For a binary vector x, we denote its *Hamming weight* by $|x|$, which is the number of 1's in x. Obviously the Hamming distance between 2 binary strings x and y is simply $|x \oplus y|$, where $x \oplus y$ denote the string obtained by entry-wise XORing x and y.

A classical probabilistic distribution for some alphabet Σ, normally denoted by $\mathcal{D}$, is a mapping from Σ^* to $[0,1]$, such that $\sum_{x\in\Sigma^*} \mathcal{D}(x) = 1$. A *uniform distribution* over a set S is denoted by $\mathcal{U}_S$, and is defined to be $\mathcal{U}_S(x) = 1/|S|$ for all $x \in S$.

We identify a random variable with its distribution and shall use the terms "random variable" and "probabilistic distribution" interchangeably.

The *correlation* of a pair of random variables X and Y over a distribution $\mathcal{D}$, denoted by $\mathsf{Cor}_{\mathcal{D}}[(X,Y)]$, is the probability they agree minus the probability they disagree.

$$\mathsf{Cor}_{\mathcal{D}}[(X,Y)] = \mathsf{Prob}_{\mathcal{D}}[X = Y] - \mathsf{Prob}_{\mathcal{D}}[X \neq Y]. \tag{3.1}$$

The *statistical distance* between two distributions X and Y is

$$\mathsf{SD}(X,Y) = \frac{1}{2}\sum_x |\mathsf{Prob}\,[X = x] - \mathsf{Prob}\,[Y - x]| \tag{3.2}$$

If the statistical distance between X and Y is ϵ, then we say that they are ϵ-*close*.

For any function over a finite set, we identify this function with its truth table, which can be written as a vector. For example, we regard a function over $\{0,1\}^n$ also as a 2^n-dimensional vector. We assume a canonical ordering of n-bit strings.

3.2 Protocols

We focus on two-party protocols executed between Alice and Bob. A protocol is normally denoted by $\mathcal{P}$. Classical protocols can be modeled by two interactive Turing machines as by Goldreich [40]. Quantum protocols can be modeled by two quantum circuits connected by classic wires, as defined by Yao [95]. The actual model of computation isn't essential for this thesis, since all the lower bounds I shall prove are information-theoretical, and therefore are independent from the actual computation model being used, and all the algorithms I present would be efficiently realizable in any of the reasonable computation models.

Next, we will give formal definitions on various aspects of the correlation distillation protocols. However, first we discuss different types of these protocols

Classical vs. Quantum The classical version of correlation distillation protocols work with classical information. At the beginning of a protocol, Alice and Bob share information that is not perfectly correlated, and at the end of the protocol, they output classical information that is almost perfectly correlated.

The quantum version of correlation distillation protocols is more appropriately called entanglement distillation protocols. Here, Alice and Bob start with qubits that are imperfectly entangled, and at the end, they output qubits that are almost perfectly entangled.

Recovering vs. Refreshing Intuitively, the *recovering protocols* are the ones that try to recover the information that is "corrupted" by a noisy channel. A bit more formally, a protocol is a recovering protocol, if Alice directly outputs her local input. Consider the situation where Alice sends some information A through a noisy channel, and when Bob receives B from the channel, A and B are not perfectly correlated (or entangled). In a recovering protocol, Alice and Bob try to reconstruct the information A Alice sent out. At the end of the protocol, Alice will output A, and Bob tries to output $\hat{A}$ that is as "close" to A as possible.

Protocols that are not recovering protocols are called *refreshing protocols*. These protocols, on the other hand, aim to generate fresh information that is not necessarily the original shared information. At the end of a refreshing protocol, Alice and Bob each outputs some information, which we denote as X and Y. The goal is to have X and Y be as correlated (or entangled) as possible.

Non-interactive, One-way, and Two-way Depending on the amount of communication, a protocol can be *non-interactive*, *one-way*, or *two-way*. A non-interactive protocol is one where Alice and Bob don't communicate at all. They are perhaps the simplest protocols in their class. For interactive protocols, we say a protocol $\mathcal{P}$ is a k-bit protocol, if it contains k bits of communication. In a *one-way* protocol, only one of the players sends information to the other party. We always assume that in this case it is Alice that sends information to Bob, and Bob sends nothing back. In a *two-way* protocol, Alice and Bob both send information to each other.

Deterministic, Randomized, and Randomized Public-Coin A distillation protocol is either *deterministic* or *randomized*. Deterministic protocols refer to ones where both Alice and Bob are deterministic. In a randomized protocol, both Alice and Bob are randomized. They both have their own supply of random bits, but they do not share any randomness. A protocol is *randomized public-coin*, if Alice and Bob have read access to a *shared* random string.

Clearly a randomized public-coin protocol is more powerful than a randomized one, which in turn is more powerful than a deterministic protocol. In fact, refreshing protocols with shared randomness are trivial, since Alice and Bob can simply discard the imperfectly shared information and use the shared randomness entirely. However, shared randomness does not trivialize quantum entanglement distillation protocols. In fact, it proves very useful in constructing EDPs.

Absolute vs. Conditional We assume that protocols always terminate. However, we make a distinction between a *successful termination* and an *abort*. Protocols that always successfully terminate are called *absolute protocols*; protocols that may abort are called *conditional protocols*. For a conditional protocol, we assume that besides the normal output, Alice will output a special symbol (either SUCC or FAIL) that indicates if the protocol successfully terminates or aborts. We assume that this special symbol is output in a special tape (in the Turing Machine notation) or a special wire (in the circuit notation), so that it will not be confused with the "normal" output of Alice. We also assume that the special symbol is a piece of classical information.

A classical correlation distillation protocol $\mathcal{P}$ works over a fixed alphabet Σ. Both the input and the output of $\mathcal{P}$ are pairs of strings in Σ.[1] A *string pair* $S \in \Sigma^n \times \Sigma^n$ is written as $S = (S^A, S^B)$, indicating that S^A belongs to Alice and S^B belongs to Bob.

[1] In fact, in some of the protocols we study in the thesis, the input and the output alphabets are different. However, they can be viewed as a natural extension to our convention here.

We say $\mathcal{P}$ is a (Σ, n, m)-protocol, if the input string pairs have length n, and the output pairs have length m. We call m the *yield* of the protocol $\mathcal{P}$. Formally we may write this as

$$\mathcal{P}(I) = O \tag{3.3}$$

where $I \in \Sigma^n \times \Sigma^n$ is the input string pair, and $O \in \Sigma^m \times \Sigma^m$ is the output string pair. At the beginning of the protocol, Alice receives I^A as her local input, and Bob receives I^B as his. At the end of the protocol, Alice outputs O^A as her local output, and Bob outputs O^B. Notice that if $\mathcal{P}$ is randomized, then O can be a random variable.

A quantum entanglement distillation protocol $\mathcal{P}$ works over qubits. The shared quantum state between Alice and Bob can be described by a mixed state ρ. Suppose Alice and Bob share a state consisting of n qubit pairs, then ρ is a mixed state in a Hilbert space of dimension 2^{2n}. The reduced density matrices of Alice and Bob represent the local information they possess regarding the state ρ. We denote them by ρ^A and ρ^B. In other words, we have $\rho^A = \mathrm{Tr}_B[\rho]$ and $\rho^B = \mathrm{Tr}_A[\rho]$.

We say $\mathcal{P}$ is an (n, m)-protocol, if its input consists n qubit pairs and it outputs m qubit pairs. We call m the *yield* of $\mathcal{P}$. Formally we write this as

$$\mathcal{P}(\rho) = \sigma \tag{3.4}$$

where ρ is a density matrix of dimension 2^{2n} and σ a density matrix of dimension 2^{2m}.

3.3 Noise Models

For both classical and quantum protocols, noise models are used to describe the inputs to the protocols. A noise model is normally denoted by N, and is either classical or quantum, and is either adversarial or probabilistic.

Definition 3.1 (Adversarial Classical Noise Model) *An* adversarial classical noise model *over an alphabet Σ, often denoted by $\mathsf{N}^{\mathsf{ca}}_{\Sigma,n}$, is a set of string pairs.*

$$\mathsf{N}^{\mathsf{ca}}_{\Sigma,n} = \{I_1, I_2, ..., I_M\} \tag{3.5}$$

where $I_k \in \Sigma^n \times \Sigma^n$ for $k = 1, 2, ..., M$. When there is no danger of confusion, the subscripts Σ and/or n are omitted.

Definition 3.2 (Probabilistic Classical Noise Model) *A* probabilistic classical noise model *over an alphabet Σ, often denoted by $\mathsf{N}^{\mathsf{cp}}_{\Sigma,n}$, is a probabilistic distribution over $\Sigma^n \times \Sigma^n$. When there is no danger of confusion, the subscripts Σ and/or n are omitted.*

Definition 3.3 (Adversarial Quantum Noise Model) *An* adversarial quantum noise model, *often denoted by $\mathsf{N}^{\mathsf{qa}}_n$, is a set of quantum (mixed) states in a 2^{2n}-dimensional Hilbert space.*

$$\mathsf{N}^{\mathsf{qa}}_n = \{\rho_0, \rho_1, ..., \rho_{M-1}\} \tag{3.6}$$

When there is no danger of confusion, the subscript n is omitted.

Definition 3.4 (Probabilistic Quantum Noise Model) *A* probabilistic quantum noise model, *often denoted by $\mathsf{N}^{\mathsf{qp}}_n$, is a single density matrix ρ of dimension 2^{2n}. When there is no danger of confusion, the subscript n is omitted.*

All our definitions on noise models (classical/quantum, adversarial/probabilistic) can be naturally extended to *families* of noise models.

Definition 3.5 (Noise Model Family) *A* noise model family *is an infinite sequence of noise models over a fixed alphabet Σ.*

$$\mathcal{N} = (\mathsf{N}_1, \mathsf{N}_2, ..., \mathsf{N}_n, ...) \tag{3.7}$$

3.4 Quality of the Protocols

We define measures for the quality of correlation distillation protocols.

3.4.1 Classical Correlation Distillation Protocols

The quality of a classical protocol is measured by the *correlation* of the string pair it outputs.

Definition 3.6 (Correlation of Classical Protocols) *If a classical correlation distillation protocol $\mathcal{P}$ produces a string pair $O = (O^A, O^B)$ on input I, then its* correlation *on input I is the correlation between O^A and O^B, and it written as as $\mathsf{Cor}[\mathcal{P}(I)]$. The correlation of $\mathcal{P}$ over an adversarial noise model N^{ca}, denoted by $\mathsf{Cor}_{\mathsf{N}^{\mathsf{ca}}}[\mathcal{P}]$, is the minimal correlation of $\mathcal{P}$ over all inputs in N^{ca}*

$$\mathsf{Cor}_{\mathsf{N}^{\mathsf{ca}}}[\mathcal{P}] = \min_{I \in \mathsf{N}^{\mathsf{ca}}} \{\mathsf{Cor}[\mathcal{P}(I)]\} \tag{3.8}$$

The correlation of $\mathcal{P}$ over a probabilistic noise model N^{cp}, denoted by $\mathcal{P}[\mathsf{N}^{\mathsf{cp}}]$, is the expected correlation of $\mathcal{P}$ over all inputs in N^{ca}

$$\mathsf{Cor}_{\mathsf{N}^{\mathsf{cp}}}[\mathcal{P}] = E_{I \in \mathsf{N}^{\mathsf{cp}}} \{\mathsf{Cor}[\mathcal{P}(I)]\} \tag{3.9}$$

Definition 3.7 (Perfect Classical Protocol) *A classical correlation distillation protocol $\mathcal{P}$ is* perfect *for a classical noise model N^{c}, if $\mathsf{Cor}_{\mathsf{N}^{\mathsf{cp}}}[\mathcal{P}] = 1$.*

Often there are other constraints on the output besides the correlation. In a recovering protocol, Alice needs to output the original information she sent over. In a refreshing protocol, both Alice and Bob need to output (locally) uniformly distributed bits. The performance of a protocol is measured both in its yield and the correlation of its output with the constraints.

3.4.2 Quantum Entanglement Distillation Protocols

The quality of a quantum protocol is measured by the fidelity of its output and the perfect EPR pairs.

Definition 3.8 (Fidelity of Quantum Protocols) *The fidelity of an entanglement distillation protocol $\mathcal{P}$ on input state ρ is the fidelity of its output, written as $\mathsf{F}(\mathcal{P}(\rho))$. The fidelity of $\mathcal{P}$ over an adversarial noise model N^{qa}, denoted by $\mathsf{F}_{\mathsf{N}^{\mathsf{qa}}}(\mathcal{P})$, is the minimal fidelity of $\mathcal{P}$ on all inputs in N^{qa}*

$$\mathsf{F}_{\mathsf{N}^{\mathsf{qa}}}(\mathcal{P}) = \min_{\rho \in \mathsf{N}^{\mathsf{qa}}} \{F(\mathcal{P}(\rho))\}. \tag{3.10}$$

The fidelity of a protocol $\mathcal{P}$ over N^{qp} is simply $\mathsf{F}(\mathcal{P}(\mathsf{N}^{\mathsf{qp}}))$.

Definition 3.9 (Perfect Quantum Protocol) *A quantum correlation distillation protocol $\mathcal{P}$ is* perfect *for a quantum noise model $\mathsf{N^q}$, if $\mathsf{F}_{\mathsf{N^q}}(\mathcal{P}) = 1$.*

Definition 3.10 (Conditional Fidelity) *For a conditional protocol $\mathcal{P}$, its* conditional fidelity *over a noise model $\mathsf{N^q}$ is its fidelity conditioned on that $\mathcal{P}$ succeeds (i.e., outputs "SUCC"), and is denoted by $\mathsf{F}^{\mathsf{c}}_{\mathsf{N^q}}(\mathcal{P})$.*

Chapter 4

Error Correcting Codes and Correlation Distillation Protocols

We discuss the relation between error correcting codes and correlation distillation protocols. In particular, we shall establish several results. The first result relates classical linear error correcting codes to classical correlation distillation protocols by proving that every linear ECC corresponds to a CDP of the same overhead with respect to the same noise model; the second result relates quantum stabilizer codes to entanglement distillation protocols by proving a similar result, that any stabilizer QECC corresponds to an EDP of the same overhead with respect to the same noise model.[1] The last result separates the power of error correction from correlation distillation. In particular, we present two noisy channels (one classical and one quantum) of such high noise rates that error correction becomes useless (for noiseless transmission of information), but there exist correlation distillation protocols that can achieve a positive rate of noiseless information transmission.

The results in this Chapter relative to this thesis are summarized in Figure 4.1.

[1]In fact, we prove that for any stabilizer QECC with an overhead of ℓ *qubits*, there exists an EDP with an overhead of ℓ *bits*. Thus, in some sense EDPs are *much more* efficient than QECC, since classical bits are much cheaper than qubits.

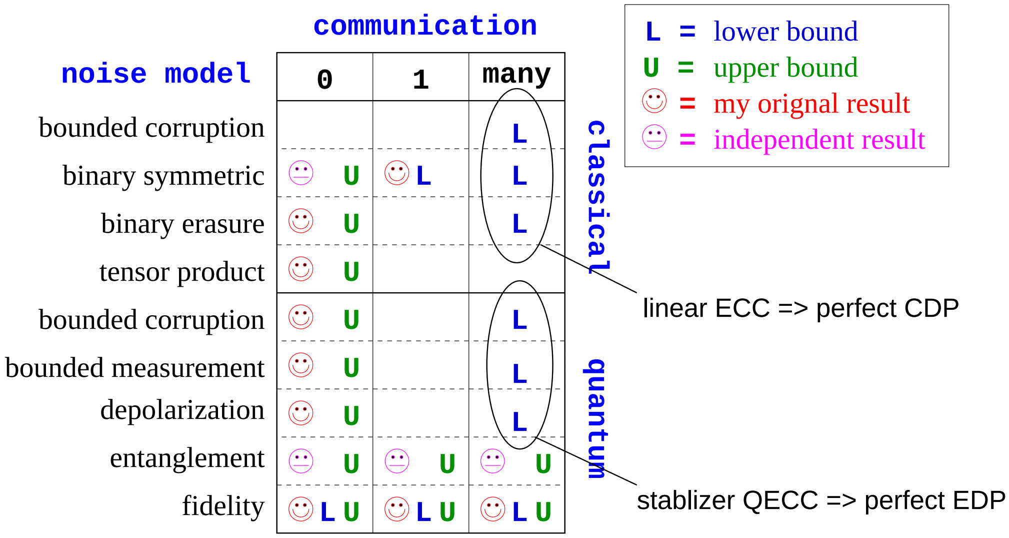

Figure 4.1: Results in Chapter 4.

4.1 Classical Error Correcting Codes and Correlation Distillation Protocols

Here we prove a very general result that relates a very large class of error correcting codes to correlation distillation protocols.

4.1.1 Error Correcting Codes

We describe the notion of Error Correcting Codes very briefly. Generally, an error correcting code is a systematic way of adding redundancy to the information, so that the redundant information is resilient to "small" disturbances. In this thesis we only focus on *block codes* that encode messages of a fixed length into code-words of a fixed length.

Definition 4.1 (Classical Error Correcting Code) *A* (classical) error correcting code *of parameter* (n, k, d) *over an alphabet* Σ *is function* $E : \Sigma^k \mapsto \Sigma^n$*, such that for any* $x, y \in \Sigma^k$*,* $x \neq y$*,* $\mathsf{dist}(E(x), E(y)) \geq d$*. The function* E *is called an* encoder*. A string* $x \in \Sigma^k$ *is called a* message*, and its image* $E(x) \in \Sigma^n$ *is called its* code-word.

This definition implicitly defines a *decoder* D as well. Consider an (n, k, d)-code. For any string $t \in \Sigma^n$, there can be at most one code-word of Hamming distance less than or equal to $(d-1)/2$ from t. If such a code-word exists, and suppose it is $E(x)$, then t will naturally be decoded to message x. If no such code-word exists, the decoding of t is *undefined*. More formally, $D : \Sigma^n \mapsto \Sigma^k$ is defined as

$$D(t) = \begin{cases} x & \text{if there exists an } x \text{ s.t. } \mathsf{dist}(E(x), t) \leq (d-1)/2 \\ \perp & \text{otherwise} \end{cases} \tag{4.1}$$

We stress that we focus on the properties of the code-words, rather than *computational complexity* of encoding/decoding. For example, we don't require the encoding and decoding algorithms of the codes to be efficient. Neither do we consider *list decoding*, where some strings more than $(d-1)/2$ away from any code-words may be decoded to a list of "candidate" messages (interested readers are referred to Guruswami's Ph.D. thesis [38] for a comprehensive survey).

4.1.2 Linear Codes

Perhaps the most important class of error correcting codes is the class of *linear codes*. Linear codes are of particular interest because of their simplicity and beautiful mathematical structures. In fact, most of the known good codes belong to the class of linear codes. The alphabet of a linear code is a finite field $\mathbb{F}$, and the encoder E for a linear code is a linear mapping from $\mathbb{F}^k$ to $\mathbb{F}^n$. Therefore E can be succinctly described as an $n \times k$ *generator matrix* G, and the encoding is simply a matrix multiplication: a message x, a k-dimensional vector, is mapped to code-word $G \cdot x$. All the code-words form a k-dimensional subspace in $\mathbb{F}^n$, which is the column space of G[2]. An (n, k, d)-linear code is often denoted as a $[n, k, d]$-code. The square brackets replaces the round parentheses to indicate that it is a linear code.

Given two linear codes E and E', represented by generator matrices G and G', we say they are *equivalent*, if G' can be obtained from G by row permutations and elementary column operations. Intuitively, if E and E' are equivalent, then one is only trivially different from the other, and there exists a very simple correspondence between the code-words of E and E'.

Next, we describe a special form of linear codes, known as the *systematic* codes. The definition is taken from [12, Definition 3.2.4, page 49].

[2]The column space of G is the subspace generated by the columns of G.

Definition 4.2 (Systematic Code) *A linear code E is a* systematic code*, if its generator matrix G is of the form $G = \begin{bmatrix} I \\ P \end{bmatrix}$, where I is an $k \times k$ identity matrix and P a $(n-k) \times k$ matrix.*

Intuitively, a systematic code is one where a code-word is the message it encodes concatenated with $(n-k)$ so-called "parity-check symbols".

It is a standard exercise in linear algebra that any linear code is equivalent to a systematic code [12, Theorem 3.2.5, page 80].

4.1.3 The Classical Bounded Corruption Model

We describe a classical noise model that is used by most error correcting codes, namely, the *classical bounded corruption model.*

Definition 4.3 (Classical Bounded Corruption Model) *A* classical bounded corruption model *of parameter (n,t) over alphabet Σ, denoted by $\mathcal{B}^{\mathsf{c}}_{n,t}$, is an adversarial model consisting of all the pairs (a,b), where both a and b are elements of Σ^n and the Hamming distance between a and b is at most t. In other words,*

$$\mathcal{B}^{\mathsf{c}}_{n,t} = \{(a,b) \mid a,b \in \Sigma^n, \mathsf{dist}(a,b) \leq t\} \tag{4.2}$$

Intuitively, the classical bounded corruption model adversarially corrupts (modifies) up to t symbols in a string of length n.

Now we are ready to state a positive result. We show a relation between systematic linear codes and correlation distillation protocols over the bounded corruption noise model.

Theorem 4.1 (From ECC to CDP) *For every systematic linear code E of parameter $[n,k,d]$ over alphabet Σ, there exists a perfect recovering, one-way, (Σ,k,k)-protocol $\mathcal{P}_E$ over a classical bounded corruption noise model $\mathcal{B}^{\mathsf{c}}_{k,(d-1)/2}$ that uses $(n-k)$ bits of communication.*

Proof: The idea behind this proof is in fact very simple. Let the generator matrix of the systematic linear code E be $G = \begin{bmatrix} I \\ P \end{bmatrix}$. Then $\mathcal{P}_E$ proceeds as follows. When $\mathcal{P}_E$ starts, Alice and Bob each possesses a length-k string, I^A and I^B, respectively. Alice then computes $C = P \cdot I^A$, an $(n-k)$-dimensional vector, sends it over to Bob, and output $O^A = I^A$. Bob then applies the decoding function D and compute $O^B = D(I^B; C)$.

We prove that $\mathcal{P}_E$ is perfect with respect to $\mathcal{B}^{\mathsf{c}}_{k,(d-1)/2}$. In fact, $I^A; C$ is the code-word for the message I^A, and since the channel $\mathcal{B}^{\mathsf{c}}_{k,(d-1)/2}$ only changes at most $(d-1)/2$ symbols, we have that $\mathsf{dist}([I^A; C], [I^B; C]) \leq (d-1)/2$. Therefore, the decoding function D will correctly decode $I^B; C$ to I^A. In other words, we have $O^A = O^B$, and thus $\mathcal{P}_E$ is perfect. ■

We present this positive result as a link to relate error correction to correlation distillation. As the result shows, in general, correlation distillation is at least as efficient as error correction, if not more efficient, for the majority of the error correction codes.

4.2 Quantum Error Correcting Codes and Entanglement Distillation Protocols

We relate the notion of quantum error correcting codes (QECCs) to entanglement distillation protocols (EDPs), with the focus on their efficiencies.

4.2.1 Quantum Error Correcting Codes

Like their counterparts in classical information theory, quantum error correcting codes are systematic ways of adding redundancy to the quantum information, so that the encoded information is resilient to "small" noises. However, quantum error correction is more complicated. First of all, unlike in the classical case, quantum information cannot be duplicated, due to the No-cloning Theorem [93]. So the redundancy added by QECCs is limited, and measurement of the error syndrome should not yield any information about the encoded message. Second, the noise model is more complicated: one qubit can suffer from a bit flip (an X operator), a phase shift (a Z operator), a bit flip *combined with* a phase shift (a Y operator), or a superposition of them. There are infinitely many (in fact, uncountably many) possible ways to "corrupt" a code-word, and a QECC needs to correct all of them. Indeed, less than one decade ago, it was not even clear if QECC was possible at all, and a positive answer by Shor [83] and Steane [85] caused quite a surprise in the quantum information community. In a nutshell, QECC is possible because of the following reasons. First, for properly designed codes, the measurement of the error syndrome will only yield information about the *errors* on a code-word, and no information about the encoded message, thus not violating the no-cloning theorem. Second, due to the linearity of quantum mechanics, it suffices to correct the *basis errors*, and all other errors will be automatically corrected (by "collapsing" into one of the basis errors), thus solving the problem of infinitely many errors.

We now formally define QECCs. We always assume that these codes work over qubits, and they are *block codes*.

Definition 4.4 (Quantum Error Correcting Code) *An* error correcting code *of parameter* (n, k, r) *is a pair of quantum algorithms* (E, D)*, both over* n *qubits as input (they can have ancillary qubits, initialized to state* $|\,0^m\rangle$*), such that for every* $x \in \{0,1\}^k$*,* $|\,\phi_x\rangle = E|\,x\rangle|\,0^{n-k}\rangle$*, and for any state* $|\,\psi\rangle$ *that can be obtained from* $|\,\phi_x\rangle$ *by (arbitrarily) modifying at most* r *qubits, we have* $D|\,\psi\rangle = |\,x\rangle \otimes \rho$ *for some mixed state* ρ *of* $n-k$ *qubits. We write such a code a* $[\![n, k, r]\!]$*-code.*

4.2.2 The Quantum Bounded Corruption Model

We describe the quantum bounded corruption model, which is the quantum counterpart of the classical bounded corruption model. Correspondingly, this model is used by most quantum error correcting codes.

Before giving the formal definition, we need some additional notations. Recall that X, Y, and Z denote the Pauli operators, while I denotes the identity operator, all over a single qubit. We define $X^0 = Y^0 = Z^0 = I$. We use X_k, Y_k, and Z_k to denote these operators over the k-th qubit. Given a $2n$-bit vector $v = (x_0, x_1, ..., x_{n-1}, z_0, z_1, ...z_{n-1})$, which we call a *Pauli vector*, we can associate it with a unique *multi-qubit Pauli operator* U_v, defined as

$$P_v = X_0^{x_0} Z_0^{z_0} \otimes \cdots \otimes X_{n-1}^{x_{n-1}} Z_{n-1}^{z_{n-1}} \tag{4.3}$$

which is a unitary operator over n qubits. Notice that since $X \cdot Z = -iY$, we have $X^0 Z^0 = I$, $X^0 Z^1 = Z$, $X^1 Z^0 = X$, and $X^1 Z^1 = -iY$. In other words, a Pauli vector designates a unitary operator formed by applying one of the four operators in $\{I, X, Y, Z\}$ to each of the n qubits. We define the *degree* of a Pauli vector to be the number of k's where x_k and z_k are not both 0, and we denote this by $\deg(v)$.

We use $[A,B]$ to denote $AB-BA$, and we say operators A and B *commute*, if $[A,B]=0$. We use $\{A,B\}$ to denote $AB+BA$, and we say operators A and B *anti-commute*, if $\{A,B\}=0$. It is not hard to see that any two Pauli operators either commute or anti-commute.

Definition 4.5 (Quantum Bounded Corruption Model) *A* quantum bounded corruption model *of parameter (n,r), denoted by $\mathcal{B}^{\mathsf{q}}_{n,r}$, is an adversarial quantum noise model consisting of all states of the form $(I\otimes P_v)\Phi_n$, where v is a Pauli vector of degree at most k. In other words,*

$$\mathcal{B}^{\mathsf{q}}_{n,r}=\{(I\otimes P_v)\ \Phi_n \mid \deg(v)\leq r\} \tag{4.4}$$

Intuitively, the quantum bounded corruption model adversarially corrupts up to r EPR pairs. The corruption appears quite limited, since it only allows applying one of the Pauli operators to Bob's share of the qubit (we call them "Pauli corruptions"). There are certainly more ways to corrupt the qubits; in fact there are uncountably many. However, since Pauli matrices, along with the identity operator, form a basis for one-qubit operations, any corruption can be decomposed into a linear superposition of the Pauli corruptions (or a mixture of them, if the corruption involves measurements).

4.2.3 An Equivalence between QECCs and One-way EDPs

Bennett et. al. [25] showed that every QECC corresponds to a one-way EDP with the same "efficiency". We review their results here.

Theorem 4.2 (From QECC to EDP [25]) *For every $[\![n,k,r]\!]$-code, there exists a corresponding perfect, deterministic, one-way, (n,k)-protocol over a quantum bounded corruption model $\mathcal{B}^{\mathsf{q}}_{n,r}$ that uses $2n$ bits of communication.*

Proof's sketch: Let (E,D) be an $[\![n,k,r]\!]$-code. We construct a protocol $\mathcal{P}$ as follows. First Alice generates k fresh EPR pairs locally, keeps half of them, and encodes the other half using E. Next, Alice sends these n qubits to Bob by teleportation, using the shared n EPR pairs. Finally Bob decodes the n qubits received using D. Since at most r out of the n original EPR pairs are corrupted, the n qubits Bob receives from the teleportation contains at most r errors, and they can be recovered by the decoding algorithm D. ■

Theorem 4.3 (From EDP to QECC [25]) *For every perfect, one-way (n, k)-protocol over a quantum bounded corruption model $\mathcal{B}^{\mathsf{q}}_{n,r}$, there exists a corresponding $[\![n, k, r]\!]$-code.*

Proof's sketch: First, we show how Alice and Bob can turn the EDP protocol into a error-correcting protocol with one-way communication. This is simple: Alice and Bob first use the EDP to distill k perfect EPR pairs, and then Alice teleport k qubits to Bob using the distilled EPR pairs. In both the EDP and the teleportation, only one-way communication is used. Finally, it was proven that any error-correcting protocol with one-way communication corresponds to a QECC with the same rate but no communication [25]. ■

4.2.4 Stabilizer Codes and EDPs

Theorems 4.2 and Theorem 4.3 establishes the equivalence between QECCs and EDPs over the quantum bounded corruption model. In particular, Theorem 4.2 shows a positive result on the power of EDPs. However, the construction of the EDPs in this theorem is not very efficient. Since n teleportation procedures are used, a total of $2n$ bits of communication is needed. Can we do better than this? The answer is "yes" for a large class of QECCs, namely the stabilizer codes.

Stabilizer Code

The class of stabilizer codes is a very general class of quantum error correcting codes, and is the analogue of the class of linear codes in classical error correction. We briefly describe the properties, and the readers are referred to Gottesman [37] and Nielsen and Chuang [69] for a comprehensive tutorial. Informally, a stabilizer code S is a collection of "parity check" operators $S = \{M_0, M_1, ..., M_{\ell-1}\}$, where each M_i is a Pauli operator, and a state $|\, x\rangle$ is a code-word, if and only if $M_i|\, x\rangle = |\, x\rangle$ for all $i = 1, 2, ..., \ell - 1$. We use $\langle S\rangle$ to denote the subgroup generated by S, and $N(S)$ the normalizer of S, which consists of all Pauli operators P such that $P \cdot S \cdot P^{\dagger} = S$. We say a subspace L is *stabilized* by S, if every element $|\, \phi\rangle \in L$ is invariant under all elements in S. In other words, $L = \{|\, \phi\rangle \mid \forall i \in [\ell], M_i|\, \phi\rangle = |\, \phi\rangle\}$, and we write this as $L = C(S)$. Then $C(S)$ is also precisely the subspace spanned by all the code-words.

Definition 4.6 (Stablizer Code) *A* $[\![n,k,r]\!]$-stabilizer code S *is an independent set of* $(n-k)$ *Pauli vectors of dimension* $2n$*, denoted by* $S = \{M_0, M_1, ..., M_{n-k-1}\}$*, such that for any two Pauli vectors* P_0*,* P_1 *of degree at most* r*,* $P_0^\dagger P_1 \notin N(S) - \langle S \rangle$*.*

It is known that an $[\![n,k,r]\!]$-stabilizer code is an $[\![n,k,r]\!]$-QECC [37, 69]. In other words, there exists generic constructions of the encoding/decoding circuit pair (E, D) from any stabilizer code. In particular, the decoding circuit D takes the following form. First, a unitary operator M is applied to all n qubits, which, intuitively, computes the $(n-k)$ "parity checks" defined by the $(n-k)$ operators $M_0, M_1, ..., M_{n-k-1} \in S$. Then, $(n-k)$ qubits are measured in the computational basis, resulting an "error syndrome" e. Finally, an appropriate "correction" circuit U_e is applied to the remaining k qubits. In particular, if the error syndrome is 0^{n-k}, then the correction circuit is the identity circuit.

Theorem 4.4 (From Stabilizer QECC to EDP) *For every* $[\![n,k,r]\!]$*-stabilizer code, there exists a corresponding perfect, one-way,* (n,k)*-protocol over a quantum bounded corruption model* $\mathcal{B}^{\mathsf{q}}_{n,r}$ *that uses* $(n-k)$ *bits of communication.*

Comparing this result to Theorem 4.2, we see a large improvement for communication complexity (from $2n$ to $n-k$). Notice that there exists $[\![n,k,r]\!]$-stabilizer codes where c is a constant and $k = n - c\log n$. In this case, Theorem 4.4 yields an exponential improvement over Theorem 4.2. This result appears to be a folk-lore in the quantum information theory community and in particular, appeared as an exercise in Nielsen and Chuang [69, pp.597].

We present a sketch of the proof for completeness.

Proof's sketch: Let $S = \{M_0, M_1, ..., M_{n-k-1}\}$ be an $[\![n,k,r]\!]$-stabilizer code, and (E, D) be the corresponding encoding/decoding circuit pair. In particular, we assume that D takes the form of a parity check circuit M (which is a linear mapping modulo 2) followed by measuring $(n-k)$ qubits and then a family of correction circuits U_e. We construct a corresponding EDP $\mathcal{P}_S$ as follows. Alice applies the decoding circuit D to her share of qubits, i.e., she applies the parity check circuit M followed by a measurement and the corresponding correction circuit. She then outputs the remaining k qubits and sends the $(n-k)$ bits of the measurement result, denoted by e_A, to Bob. Bob performs the same parity check circuit M to his share of qubits, followed by the same measurement and obtain $(n-k)$ bits, denoted by e_B. Then Bob computes $e = e_A \oplus e_B$ and applies the correction circuit U_e to his remaining k qubits and output them.

We prove that protocol $\mathcal{P}_S$ is perfect. The main observation is that a stabilizer code is linear. The entire space of n-qubit states can be decomposed into 2^{n-k} subspaces, each of dimension 2^k and denoted by L_e, where $e \in \{0,1\}^{n-k}$, such that each subspace L_e is stabilized by the group generated by $S_e = \{(-1)^{e_0} \cdot M_0, ..., (-1)^{e_{n-k-1}} \cdot M_{n-k-1}\}$. In particular, L_0 is the subspace spanned by all code-words. Naturally, all these 2^{n-k} subspaces are isomorphic to each other.

Now consider the operation in $\mathcal{P}_S$. If the input to the protocol is $(I \otimes P_v)|\,\phi\rangle^A|\,\phi\rangle^B$, then this state becomes $(M \otimes MP_v)|\,\phi\rangle^A|\,\phi\rangle^B$ after both Alice and Bob have applied their parity check circuits. If $|\,\phi\rangle$ is a code-word, then it is clear that Alice's measurement would yield $e_A = 0$ and Bob will apply the correction circuit $U_e = U_{e_B}$, which will correct the "corrupted code-word" $P_v|\,\phi\rangle$, and result in state $|\,\psi\rangle^A|\,\psi\rangle^B$, where $|\,\psi\rangle$ is the decoding of state $|\,\phi\rangle$. Now, we fixed P_v and consider the case $|\,\phi\rangle \in L_a$ is not a code-word. In this case, it is not hard to see that $M|\,\phi\rangle$ will yield a measurement result of a. Furthermore, the measurement of $MP_v|\,\phi\rangle^B$ will give a result of $e \oplus a$, since $|\,\phi\rangle$ is stabilized by $\langle S_e\rangle$, which has the same commute/anti-commute property with P_v as $\langle S_e\rangle$, since a phase change does not affect commutability. Therefore, Bob will still apply the correcting circuit U_e, effectively "remove" the affect of P_v — this is by the isomorphism between L_0 and L_a.

Finally, notice that the state Φ_n can is a superposition of 2^n states of form $|\,\phi_x\rangle^A|\,\phi_x\rangle^B$, with 2^k x's from each subspace L_e. Overall, we conclude that the output of the protocol $\mathcal{P}_S$ is Φ_k. ∎

4.3 Separating Error Correction from Correlation Distillation

We present two (very) noisy channels, one classical, one quantum. In both channels the error correction almost completely fails to transmit information noiselessly (because of the high noise rate), while there exist correlation distillation protocols promising a positive rate of noiseless information transmission. These results show a separation between the power of error correction and that of correlation distillation.

4.3.1 Separation Result for Classical Channels

Consider a classical bounded corruption model $\mathcal{B}^{\mathsf{c}}_{n,n/3}$. It is a classical result that a perfect error correcting code can only encode two bits of information for such a channel. So error correction is almost useless.

Theorem 4.5 (Limits on ECCs) *A perfect error correcting code for $\mathcal{B}^{\mathsf{c}}_{n,n/3}$ can only encode 2 bits of information.*

Proof: We prove that there can be at most 4 n-bit vectors such that any two of them have Hamming distance at least $2n/3$. This will imply the theorem.

We write these vectors as $v_1, v_2, ..., v_m$ and define m new vectors $u_1, u_2, .., u_m$ as follows: $u_i[j] = 2v_i[j] - 1$. Thus each entry of u_i is ± 1 and we have $(u_i, u_j) \leq -1/3$, where (x, y) denotes the scaled inner product $(x, y) = \frac{1}{n}\sum_i x[i] \cdot y[i]$. This is because any u_i and u_j must differ at at least $2n/3$ of their entries. Now, let $u = \sum_i u_i$, and we compute (u, u). We have

$$(u,u) \quad = \quad \sum_i (u_i, u_i) + \sum_{i \neq j} (u_i, u_j) \leq m - m(m-1)/3 = m(4-m)/3$$

Since we have $(u, u) \geq 0$, we have $m \leq 4$. ∎

However, using correlation distillation protocols, we can do much better. To show the result, we need to introduce some notions from [72, 65].

Consider a cooperative game played by two players, the "sender" S and the "receiver" R. At the beginning of the game, S receives a private input x and R receives a private input y, where the pair (x, y) is drawn from a pre-determined set $T \subseteq \{(x, y) | x, y \in \{0,1\}^*\}$. Here we call T the *support set.* Furthermore, we define the *projection* of T on the sender S to be $T_S = \{x : (x, y) \in T \text{ for some } y\}$, and T_R similarly. During the game, S and R communicate, using a pre-determined protocol $\mathcal{P}$. At the end of the game, R outputs x', and they win if $x' = x$. A *winning* protocol is one that always wins over all inputs in T. The *communication complexity* of the protocol $\mathcal{P}$ is the maximum number of bits exchanged between S and R over all possible inputs $(x, y) \in T$. Clearly this game is closely related to correlation distillation protocols, and in particular, if the support set T is an adversarial classical noise model, then the protocol is precisely a perfect recovering classical protocol.

For a fixed support set T and a string y, we define the *ambiguity* of y (with respect to T) to be

$$\mu(y,T) = |\{x \,:\, (x,y) \in T\}|\,, \tag{4.5}$$

and the *maximum ambiguity* of T to be

$$\hat{\mu}(T) = \max_{y \in T_R}\{\mu(y,T)\}. \tag{4.6}$$

An element $y \in T_R$ defines a *hyperedge*

$$E(y) = \{x \,:\, (x,y) \in T\}. \tag{4.7}$$

Finally, we define the *edge count* of T is defined as

$$\sigma(T) = |\{E(y) \,:\, y \in T_R\}|\,. \tag{4.8}$$

Naor *et al.*[65] proved the following theorem.

Theorem 4.6 (Communication Compleiry Result [65]) *For any support set T, there exists a four-round winning protocol with communication complexity at most*

$$\log\log\sigma(T) + \log\hat{\mu}(T) + 3\log\log\hat{\mu}(T) + 7 \tag{4.9}$$

Now we bring our attention to the case where $T = \mathcal{B}^{\mathsf{c}}_{n,t}$ is a classical bounded corruption model. It is straightforward to compute the maximum ambiguity and the edge count of T. In fact, we have $\hat{\mu}(T) = \sigma(T) = \sum_{i=0}^{n/3}\binom{n}{i} \leq 2^{n\cdot(H(1/3)+o(1))}$ (we refer the readers to, for example, Sudan's course note [87] for proofs). Plugging in this to Theorem 4.6, we have

Theorem 4.7 (Separating ECC from CDP) *There exists a four-round perfect correlation distillation protocol for the noise model $\mathcal{B}^{\mathsf{c}}_{n,t}$ with communication complexity $n \cdot (H(1/3) + o(1))$.*

Here $H(x)$ is defined as $H(x) = -x \cdot \log(x)$. Notice that $\log\hat{\mu}(T)$ dominates all other terms in (4.9). By investing about $n \cdot H(1/3) \approx 0.918n$ bits of communication, Alice and Bob are able transmit n bits of information. So the saving is about $0.082n$ bits. By contrast, the error correction approach can only manage to get two bits through.

4.3.2 Separation Results for Quantum Channels

The separation results for quantum channels is in fact given by Bennett *et al.* [25]. We briefly sketch a slight variation of their result here for completeness.

Consider a quantum bounded corruption model $\mathcal{B}^{\mathsf{q}}_{n,n/2}$. In other words, the model corrupts up to half of the qubits transmitted. One can easily prove that there does not exist perfect QECC for such a channel. Here is a brief sketch. Assuming otherwise, then we can feed a k-qubit state $|\,\phi\rangle$ into the encoding algorithm and obtain an n-qubit state $|\,\psi\rangle$. The decoding algorithm would be able to recover $|\,\phi\rangle$ from the first $n/2$ qubits of $|\,\psi\rangle$, as well as the last $n/2$ qubits of $|\,\psi\rangle$. If we do both, we can effectively clone the state $|\,\phi\rangle$, which contradicts the No-cloning Theorem. Therefore no QECC can be used here to even transmit a single qubit perfectly.

On the other hand, there exists a two-round entanglement distillation protocol for $\mathcal{B}^{\mathsf{q}}_{n,n/2}$ that produces a constant fraction (0.00457) of perfect EPR pairs that can then be used to transmit quantum information through teleportation. The detailed protocol can be found in [25].

Chapter 5

Non-Interactive Correlation Distillation

Here we demonstrate a series of negative results that aim to understand one of the most basic problems in the communication complexity of correlation distillation, i.e., how well Alice and Bob can do *if there is no communication at all*? We call this process *non-interactive correlation distillation* (NICD).

The results in this Chapter relative to this thesis are summarized in Figure 5.1.

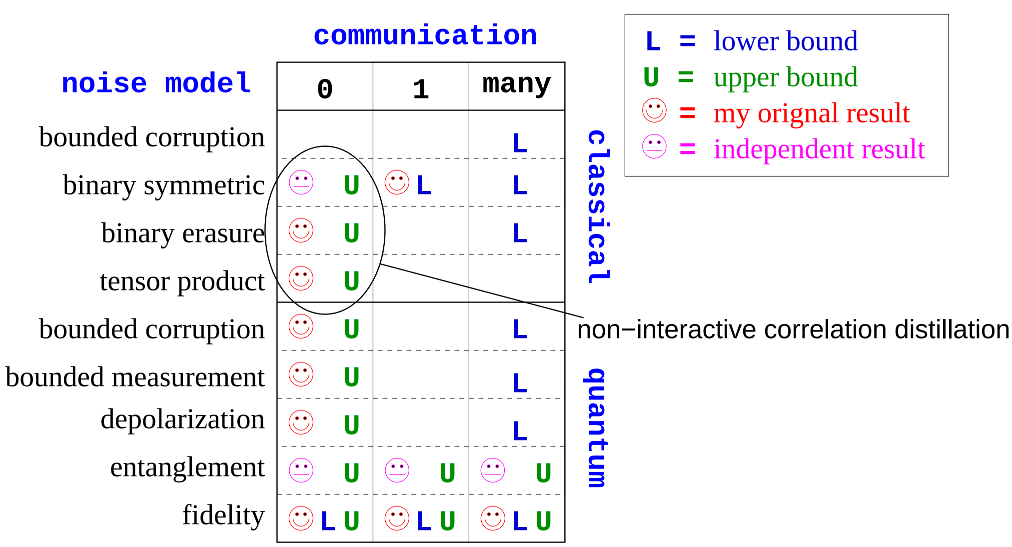

Figure 5.1: Results in Chapter 5.

At the first glimpse of the problem, it may be tempting to answer "nothing interesting". Intuitively, it makes sense; if Alice and Bob do not communicate at all, they have no knowledge about the other party, and how would they possible "recover" the information?

This intuition is in some sense correct for recovering protocols. Recall that in a recovering protocol, Alice simply outputs her input ($O^A = I^A$), and Bob wishes to output a O^B that is as close to O^A as possible. For an adversarial noise model, the optimal behavior of Bob is determined by the minimax theorem. For a probabilistic noise model, Bob knows I^B and the joint distribution (I^A, I^B), and therefore his optimal strategy is to "guess" I^A according to the Bayes rule. In other words, Bob needs to choose X such that

$$X = \operatorname{argmax}_x \left\{ \frac{\mathcal{D}(x, I^B)}{\sum_y \mathcal{D}(y, I^B)} \right\} \tag{5.1}$$

where $\mathcal{D}$ is the distribution of (I^A, I^B) according to the noise model. Therefore, the noise model essentially determines the optimal strategy of Alice and Bob for non-interactive recovering protocols.

However, the situation is quite different for refreshing protocols over a probabilistic noise model. In a refreshing protocol, Alice and Bob share a probabilistic noise model, which is a distribution over the string pairs. Alice does not need to output her input string verbatim. Rather, Alice and Bob have the liberty to output *anything*. Furthermore, Alice and Bob may gather a large collection of the samples, all from the same distribution, and then hope to "concentrate" the correlation down to a small number of symbols. In this case, the problem of whether Alice and Bob can distill highly correlated bits without communication is not intuitively clear.

In fact, this problem of non-interactive correlation distillation has been considered by various researchers from different perspectives.

Consider the study of information reconciliation. In information reconciliation, Alice and Bob each possess some information that are not perfectly correlated. They wish to distill highly correlated bits by communication, yet maintaining privacy. In this model, Eve, the eavesdropper, can see all the communication between Alice and Bob. Therefore, if Alice and Bob could distill correlated bits non-interactively, this would be ideal for information reconciliation. Moreover, only after having an impossibility result on non-interactive distillation should one consider interactive information reconciliation. In this sense, the problem of non-interactive correlation distillation is the underlying problem of the study of information reconciliation, and only a negative answer to this problem can justify the existence of this study.

A similar situation exists in the study of random beacons. In this setting, Alice (the beacon owner) and Bob (the verifier) each possesses the measurement data from an extraterrestrial object. Due to the measurement error, their data are correlated but not perfectly so. Alice would convert her measurement into a sequence of random bits and publish these bits. The goal of the study of random beacons is to construct a *publicly verifiable* random source, and prevent Alice (the beacon owner) from cheating, i.e., affecting the outcome of the bits. If it is possible to distill highly correlated bits non-interactively, then the random beacon problem would be perfectly solved. Alice distills her bits from the measurement and publishes them. Then Bob can apply his part of the distillation, and with very high probability the result would agree with the bits Alice publishes. If the bits do not agree, Bob announces that Alice is cheating. In this way Alice would have no motives to cheat, since Bob can catch her cheating with very high probability. Therefore, here again, the problem of non-interactive distillation underlies the study of random beacons, and a negative answer to this problem lies at the foundation of this study.

Given the importance of this problem, it is not surprising that many researchers have considered it. In fact, a basic version of the problem was discovered and proven independently by several researchers beginning in 1991, including Alon, Maurer, Wigderson [3], Mossel and O'Donnell [63], and Yang [96].

We shall prove a sequence of negative answers to various versions of this problem. We assume that in all the protocols considered in this section, Alice and Bob only output one bit each. We make this assumption, since it seems to be the minimal requirement for a useful refreshing protocol. In some of the results, we will consider protocols whose output alphabets differ from their input alphabets.

5.1 Tensor Product Noise Models

The noise models we discuss in this section are of a special form, which we call the "tensor product noise models". First, we review the definitions of the tensor product.

Definition 5.1 (Tensor Product of Vectors) *The* tensor product *of an n-dimensional vector v and an m-dimensional vector u is an $(n \cdot m)$-dimensional vector, denoted by w, such that $w[(x, y)] =$*

*$v[x] \cdot u[y]$, for $x \in [n]$ and $y \in [m]$. We use $v^{\otimes k}$ to denote the vector obtained by taking the tensor product of k copies of v, and call it the n-*th tensor power *of v.*

Definition 5.2 (Tensor Product of Matrices) *The* tensor product *of an $a \times c$ matrix A and a $b \times d$ matrix B is an $(ab) \times (cd)$ matrix P, such that $P_{(x,z),(y,w)} = A_{x,y} \cdot B_{z,w}$ for $x \in [a]$, $y \in [b]$, $z \in [c]$, and $w \in [d]$. We write this as $P = A \otimes B$. We use $A^{\otimes k}$ to denote the matrix obtained by taking the tensor product of k copies of A, and call it the n-*th tensor power *of A.*

Definition 5.3 (Tensor Product of Probabilistic Distributions) *The* tensor product *of a probabilistic distribution $\mathcal{D}_A$ over set A and a distribution $\mathcal{D}_B$ over set B is a distribution $\mathcal{D}$ over set $A \times B$, such that $\mathcal{D}(a,b) = \mathcal{D}_A(a) \cdot \mathcal{D}_B(b)$. We write this as $\mathcal{D} = \mathcal{D}_A \otimes \mathcal{D}_B$. We use $\mathcal{D}^{\otimes k}$ to denote the matrix obtained by taking the tensor product of k copies of $\mathcal{D}$, and call it the n-*th tensor power *of $\mathcal{D}$.*

Definition 5.4 (Tensor Product Classical Noise Model) *A probabilistic classical noise model $\mathsf{N}^{\mathsf{cp}}_{\Sigma,n}$ is a* tensor product classical noise model*, if there exists a probabilistic distribution $\mathcal{D}$ over $\Sigma \times \Sigma$ such that $\mathsf{N}^{\mathsf{cp}}_{\Sigma,n}$ is formed by the pair $(a_0 a_1 \cdots a_{n-1}, b_0 b_1 \cdots b_{n-1})$, where (a_k, b_k) is independently drawn from $\mathcal{D}$, for $k = 0, 1, ..., n-1$. The distribution $\mathcal{D}$ is called the* base distribution *of $\mathsf{N}^{\mathsf{cp}}_{\Sigma,n}$.*

In other words, the distribution of $\mathsf{N}^{\mathsf{cp}}_{\Sigma,n}$ is simply the n-th tensor power of the distribution $\mathcal{D}$ with symbols rearranged.

5.2 The Binary Symmetric Model

We first prove the negative result to perhaps the most basic version of the problem.

Definition 5.5 (Binary Symmetric Model) *A* binary symmetric model *of parameter (n,p), denoted as $\mathcal{S}_{n,p}$, is a probabilistic noise model defined as follows*

$$\mathcal{S}_{n,p}(a,b) = \frac{1}{2^n}(1-p)^{n-|a \oplus b|} \cdot p^{|a \oplus b|} \tag{5.2}$$

where $a, b \in \{0,1\}^n$.

The binary symmetric model is indeed a tensor produce noise model, and its base distribution is defined as $\mathcal{D}(0,0) = \mathcal{D}(1,1) = (1-p)/2$ and $\mathcal{D}(0,1) = \mathcal{D}(1,0) = p/2$. This model is closely related to the so-called "Binary Symmetric Channel". Imagine that Alice generates a uniform bit A as her local input, and sends it to Bob through a noisy channel that flips each bit independently with probability p. If we denote the bit received by Bob by B, then the distribution of (A, B) is precisely $\mathcal{D}$.

Now suppose the bit strings of Alice and Bob are described by $\mathcal{S}_{n,p}$. Alice and Bob each wishes to output one bit, denoted by a and b, respectively, such that the correlation between a and b is maximized. We also require that a and b themselves be unbiased. What is the maximum possible correlation of a and b, if Alice and Bob are not allowed to communicate?

If Alice and Bob simply output the kth bit of their strings, for any $k \in [n]$, their outputs will have a correlation $1-2p$. This method is very simple, and almost appear naïve. Do there exist more sophisticated methods which will yield a higher correlation? Intuitively, it is not entirely clear that there do not. Our first negative result addresses this problem and proves that in fact the "naïve" method is optimal, and no protocol can yield a higher correlation than $1-2p$.

First, we need to define a restricted class of protocols, namely, locally uniform protocols.

Definition 5.6 (Locally Uniform Protocols) *A protocol $\mathcal{P}$ is* locally uniform *over a probabilistic noise model N^{cp}, if the distribution of its outputs are locally uniform bits, i.e., both O^A and O^B are uniform distributions over $\{0,1\}$, where $(O^A, O^B) = \mathcal{P}(\mathsf{N}^{\mathsf{cp}})$.*

Theorem 5.1 (NICD for the Binary Symmetric Model) *The correlation of any locally uniform, randomized, non-interactive protocol over the binary symmetric model of parameter (n,p) is at most $1-2p$ for $p \leq 1/2$.*

The deterministic version of Theorem 5.1 (where the protocol is restricted to deterministic) was discovered and proven independently since 1991 by many researchers, including Alon, Maurer, Wigderson, Mossel, O'Donnell, and Yang [3, 63, 96], and was attributed to "folklore" by Mossel and O'Donnell [63].

Proof: To prove the theorem, it suffices to consider protocols of yield 1, namely, protocols where Alice and Bob only output one bit each.

Consider a non-interactive protocol $\mathcal{P}$. Since there is no communication, the most general characterization of the protocol would be that both Alice and Bob apply a (randomized) boolean function to their share of bit strings, and output the result.

We define the *character functions* of Alice and Bob as follows. The character function of Alice, denoted by ϕ^A, maps strings from $\{0,1\}^n$ to real numbers within $[-1,+1]$. Over input x,

$$\phi^A(x) = 2 \cdot \mathsf{Prob}\left[\text{Alice outputs 1 over input } x\right] - 1, \tag{5.3}$$

where the probability is taken over the random bits used by Alice. Similarly the character function ϕ^B of Bob can be defined.

Since $\mathcal{P}$ is locally uniform over the binary symmetric model $\mathcal{S}_{n,p}$, we have

$$E_{x,y\in\mathcal{S}_{n,p}}[\phi^A(x)] = E_{x,y\in\mathcal{S}_{n,p}}[\phi^B(y)] = 0 \tag{5.4}$$

Notice that for any x, we have

$$\sum_y \mathcal{S}_{n,p}(x,y) = \frac{1}{2^n}\sum_x (1-p)^{n-|x\oplus y|}\cdot p^{|y\oplus y|} = \frac{1}{2^n}$$

and thus (5.4) simplifies to

$$\sum_x \phi^A(x) = \sum_x \phi^B(x) = 0 \tag{5.5}$$

It is easy to verify if Alice receives x as her input and Bob receives y, then $\phi^A(x)\cdot\phi^B(y)$ is the correlation between their outputs. Therefore, the correlation of protocol $\mathcal{P}$ over the binary symmetric model is

$$\begin{aligned}
\mathsf{Cor}_{\mathcal{S}_{n,p}}[\mathcal{P}] &= \sum_{x\in\{0,1\}^n}\sum_{y\in\{0,1\}^n} S_{n,p}(x,y)\cdot\phi^A(x)\cdot\phi^B(y) \\
&= \frac{1}{2^n}\sum_{x\in\{0,1\}^n}\sum_{y\in\{0,1\}^n}(1-p)^{n-|x\oplus y|}\cdot p^{|x\oplus y|}\cdot\phi^A(x)\cdot\phi^B(y)
\end{aligned}$$

Now we view the summation above as a quadratic form. We define a $2^n\times 2^n$ matrix S, where $S_{x,y} = (1-p)^{n-|x\oplus y|}\cdot p^{|x\oplus y|}$

We identify the character functions ϕ^A and ϕ^B with their truth tables, which are 2^n-dimensional real vectors. Then it is easy to verify that

$$\mathsf{Cor}_{\mathcal{S}_{n,p}}[\mathcal{P}] = \frac{1}{2^n}(\phi^A)^T \cdot S \cdot \phi^B \tag{5.6}$$

We can diagonalize the matrix S and it turns out it is a positive matrix with eigenvectors being parity functions. More formally, define parity functions as $\oplus_a x = (-1)^{a \cdot x}$, where $a, x \in \{0,1\}^n$, and $a \cdot x$ is the inner product of a and x. Then each $\oplus_a$ is an eigenvector with eigenvalue $\lambda_a = (1-2p)^{|a|}$. The statement and the proof are postponed to Lemma 5.1 (after this proof). An important observation is that the unique largest eigenvalue is 1, with corresponding eigenvector $\oplus_0$. Here we use 0 as a shorthand to denote the all-zero vector. All other eigenvalues are at most $1-2p$.

We now perform a Fourier Analysis to vectors ϕ^A and ϕ^B. First we define an inner product for 2^n-dimensional vectors: for vectors A and B, their inner product is defined as $\langle A, B \rangle = \frac{1}{2^n}\sum_{x\in\{0,1\}^n} A[x]B[x]$. It is then easy to verify that all the parity functions $\{\oplus_a\}_{a\in\{0,1\}^b}$ form an orthonormal basis. We can then write $\phi^A = \sum_{s\in\{0,1\}^n} \alpha_s \oplus_s$ and $\phi^B = \sum_{s\in\{0,1\}^n} \beta_s \oplus_s$. Notice that since $|\phi^A[x]| \leq 1$, we know that $||\phi^A|| \leq 1$, and thus $\sum_s \alpha_s^2 \leq 1$, by Parseval. Similarly we have $\sum_s \beta_s^2 \leq 1$.

Furthermore, since $\oplus_0$ is the constant function, we know that

$$\alpha_0 = \frac{1}{2^n}\sum_x \phi^A(x) \oplus_0 (x) = \frac{1}{2^n}\sum_x \phi^A(x) = 0$$

Similarly we have $\beta_0 = 0$.

Next, we break down the summation in (5.6):

$$\begin{aligned}
\mathsf{Cor}_{\mathcal{S}_{n,p}}[\mathcal{P}] &= \frac{1}{2^n}(\phi^A)^T \cdot S \cdot \phi^B \\
&= \frac{1}{2^n}\left(\sum_a \alpha_a \cdot \oplus_a^T\right) \cdot S \cdot \left(\sum_b \beta_b \cdot \oplus_b\right) \\
&= \frac{1}{2^n}\left(\sum_a \alpha_a \cdot \oplus_a^T\right) \cdot \left(\sum_b \beta_b \cdot \lambda_b \cdot \oplus_b\right) \\
&= \sum_a \alpha_a \cdot \beta_a \cdot \lambda_a
\end{aligned}$$

Now, since $\alpha_0 = \beta_0 = 0$, and $\lambda_a \leq 1-2p$ for all $a \neq 0$, we have

$$\mathsf{Cor}_{\mathcal{S}_{n,p}}[\mathcal{P}] \leq (1-2p) \cdot \sum_a \alpha_a \cdot \beta_a \leq (1-2p) \cdot \left(\sum_a \alpha_a^2\right)^{\frac{1}{2}} \cdot \left(\sum_a \beta_a^2\right)^{\frac{1}{2}} = 1-2p \tag{5.7}$$

The second inequality is by Cauchy-Schwartz. ■

Lemma 5.1 *Let S be a $2^n \times 2^n$ matrix defined by $S_{xy} = p^{|x \oplus y|}(1-p)^{n-|x\oplus y|}$. Let $e_a(x)$ be function defined by $e_a(x) = (-1)^{a \cdot x}$ Then we have $S \cdot e_a = (1-2p)^{|a|} \cdot e_a$.*

Proof: Notice that for any $x \in \{0,1\}^n$, we have

$$\begin{aligned}
(S \cdot e_a)[x] &= \sum_y p^{|x\oplus y|}(1-p)^{n-|x\oplus y|} \cdot (-1)^{a\cdot y} \\
&= \sum_y p^{|y|}(1-p)^{n-|y|} \cdot (-1)^{a\cdot(x\oplus y)} \\
&= (-1)^{a\cdot x} \cdot \sum_y p^{|y|}(1-p)^{n-|y|} \cdot (-1)^{a\cdot y} \\
&= e_a[x] \cdot \sum_y p^{|y|}(1-p)^{n-|y|} \cdot (-1)^{a\cdot y}
\end{aligned}$$

Now it should already be clear that e_a is an eigenvector. Next, we compute the corresponding eigenvalue. We shall prove that

$$\sum_y p^{|y|}(1-p)^{n-|y|} \cdot (-1)^{a\cdot y} = (1-2p)^{n-|a|}.$$

WLOG we assume that a contains k 1's followed by $(n-k)$ 0's. We partition each y into y_0

and y_1, where y_0 contains the first k bits, and y_1 contains the last $(n-k)$ bits. We use $y_0; y_1$ to denote the concatenation of y_0 and y_1. Then the previous formula becomes

$$\begin{aligned}
\sum_y p^{|y|}(1-p)^{n-|y|} \cdot (-1)^{a \cdot y} &= \sum_{y_0 \in \{0,1\}^k} \sum_{y_1 \in \{0,1\}^{n-k}} p^{|y_0|+|y_1|}(1-p)^{n-|y_0|-|y_1|} \cdot (-1)^{|y_0|} \\
&= \left(\sum_{y_0} (-p)^{|y_0|}(1-p)^{k-|y_0|}\right) \cdot \left(\sum_{y_1} (p)^{|y_1|}(1-p)^{n-k-|y_1|}\right) \\
&= (-p+1-p)^k \cdot (p+1-p)^{n-k} \\
&= (1-2p)^k
\end{aligned}$$

■

In fact, this proof implies more than the theorem. From the proof, we can see that the only protocols that saturate the $1-2p$ upper bound are the ones where Alice and Bob both output the k-th bit or the complement of the k-th bit, for some $k \in [n]$. To see this, we re-examine the proof. The only way to make (5.7) an equality is that for all a such that $\lambda_a < 1-2p$, we have $\alpha_a \cdot \beta_a = 0$. Also we must have $\sum_a \alpha_a^2 = \sum_a \beta_a^2 = 1$, and $\alpha_a = \beta_a$ for all by. Putting things together, we see that for all a's of Hamming weight more than 1, we have $\alpha_a = 0$. So we have $\phi^A(x) = \sum_{|a|=1} \alpha_a \oplus_a (x)$. There are n such α_a's, and we can always find an x such that $\oplus_a(x)$ has the same sign as α_a. Denote this x by $\tilde{x}$, and then we have

$$\phi^A(\tilde{x}) = \sum_{|a|=1} \alpha_a \oplus_a (\tilde{x}) = \sum_{|a|=1} |\alpha_a| \geq \sum_{|a|=1)} \alpha_a^2 = 1 \tag{5.8}$$

However, we have $\phi^A(\tilde{x}) \leq 1$, and thus the inequality in (5.8) must be an equality, which means each α_a is either 0, 1, or -1. So there exists a k such that both ϕ^A and ϕ^B are parity functions $\oplus_{\{k\}}$ or its complement. These functions correspond to the "naïve" protocols where Alice and Bob both outputs the k-th bit or the complement of the k-th bit.

We can further extend Theorem 5.1 to protocols that are not locally uniform.

Definition 5.7 (δ-Locally Uniform Protocols) *A protocol $\mathcal{P}$ is δ-*locally uniform *over a probabilistic noise model N^{cp}, if the distribution of its output are locally δ-close to uniform bits, i.e., both O^A and O^B are δ-close to uniform distributions over $\{0,1\}$, where $(O^A, O^B) = \mathcal{P}(\mathsf{N}^{\mathsf{cp}})$.*

Theorem 5.2 (NICD for the Binary Symmetric Model, extended) *The correlation of any δ-locally uniform, randomized, non-interactive protocol over the binary symmetric model of parameter (n,p) is at most $1-2p(1-4\delta^2)$ for $p \le 1/2$.*

Proof: Consider a $(\{0,1\},n,1)$-protocol $\mathcal{P}$. But definition it outputs a single bit-pair $O = (O^A, O^B)$. Since $\mathcal{P}$ is δ-locally uniform, we know that O^A is δ-close to $\mathcal{U}_{\{0,1\}}$. We define $\mathsf{Prob}\,[O^A = 0] = 1/2 - t$, then we have $\mathsf{Prob}\,[O^A = 1] = 1/2 + t$, and $\mathsf{SD}(O^A, \mathcal{U}_{\{0,1\}}) = |t|$. Therefore if we denote the character functions of Alice and Bob by ϕ^A and ϕ^B, respectively, then we have $|t| \le \delta$. On the other hand, $E_{x,y\in\mathcal{S}_{n,p}}[\phi^A(x)] = 2\cdot\mathsf{Prob}\,[O^A = 1] - 1 = 2t$, and thus we have

$$|\sum_x \phi^A(x)| \le \delta \cdot 2^{n+1} \tag{5.9}$$

Similarly we have

$$|\sum_x \phi^B(x)| \le \delta \cdot 2^{n+1} \tag{5.10}$$

As in the proof to Theorem 5.1, we perform Fourier analysis to ϕ^A and ϕ^B, and white $\phi^A = \sum_{s\in\{0,1\}^n} \alpha_s \oplus_s$ and $\phi^B = \sum_{s\in\{0,1\}^n} \beta_s \oplus_s$. Then we know from (5.9) and (5.9) that $|\alpha_0| \le 2\delta$ and $|\beta_0| \le 2\delta$.

Then we know that

$$\mathsf{Cor}_{\mathcal{S}_{n,p}}[\mathcal{P}] \;=\; \sum_a \alpha_a \cdot \beta_a \cdot \lambda_a \le 4\delta^2 + (1-4\delta^2)(1-2p) = 1 - 2p(1-4\delta^2)$$

■

Theorem 5.2 shows a trade-off between the "local uniformness" of a protocol and its correlation.

5.3 General Noise Models

Here, we extend the previous result to a general class of noise models.

Definition 5.8 (Distribution Matrix) *Let $\mathcal{D}$ be a probabilistic distribution over $\Sigma \times \Sigma$, where $|\Sigma| = q$. We say a $q \times q$ matrix M is the* distribution matrix *for $\mathcal{D}$, if $M_{x,y} = \mathcal{D}(x,y)$ for all*

$x, y \in \Sigma$.[1] *We write the distribution matrix of* $\mathcal{D}$ *by* $M_{\mathcal{D}}$.

Definition 5.9 (Regular Matrix) *A* $q \times q$ *matrix* M *is* regular *if it is symmetric and* $\mathbf{1}_q$ *is the unique eigenvector with the largest absolute eigenvalue. Let* ϵ *be the difference between the largest absolute eigenvalue and the second largest. Then* $q \cdot \epsilon$ *is called the* scaled eigenvalue gap *of* M. *A distribution* $\mathcal{D}$ *is* regular *if its distribution matrix is regular.*

Notice that a distribution matrix M is non-negative (that every entry is non-negative). By the Perron-Frobenius Theorem [59], if M is symmetric, irreducible, and has $\mathbf{1}_q$ as an eigenvector, then $\mathbf{1}_q$ is the unique eigenvector with the largest eigenvalue, and thus M is regular. Therefore, intuitively, a noise model N^{cp} is regular if it satisfies the following three requirements: that it is *symmetric*, i.e., $\mathsf{N}^{\mathsf{cp}}(a, b) = \mathsf{N}^{\mathsf{cp}}(b, a)$ for every $a, b \in \Sigma$; that it is *locally uniform*, i.e., both the distributions of the local inputs of Alice and Bob are uniform; that it is *connected*, i.e., Σ cannot be partitioned into Σ_0 and Σ_1 such that $\mathsf{N}^{\mathsf{cp}}(a, b) = \mathsf{N}^{\mathsf{cp}}(b, a) = 0$ for all $a \in \Sigma_0$ and $b \in \Sigma_1$. Notice that if a noise model is not connected, that non-interactive correlation distillation is indeed possible for such a model. Suppose Σ is partitioned into Σ_0 and Σ_1. If Alice and Bob interpret symbols in Σ_0 as a "0" and symbols in Σ_1 as a "1", then they essentially have a noiseless binary noise model which allows for non-interactive correlation distillation.

Theorem 5.3 (NICD for the General Noise Model) *If* $\mathsf{N}^{\mathsf{cp}}_{\Sigma,n}$ *is a tensor product noise model whose base distribution is regular with scaled eigenvalue gap* ϵ, *then the correlation of any* δ*-locally uniform, randomized, non-interactive* $(\Sigma, n, 1)$*-protocol over the classical probabilistic noise model* $\mathcal{D}^{\otimes n}$ *is at most* $1 - \epsilon(1 - 4\delta^2)$.

To see that Theorem 5.3 is indeed a more general result, notice that the base distribution of the binary symmetric model is indeed regular with scaled eigenvalue gap $2p$.

Proof: The strategy of this proof is the same as of that to Theorem 5.1. We convert the correlation of a protocol $\mathcal{P}$ into a quadratic form, and then we diagonalize the matrix and use Fourier analysis to upper bound the correlation.

[1] Here we identify Σ with $[q]$.

We define $q = |\Sigma|$ and identify Σ with $[q]$ for the rest of the proof. We still use ϕ^A and ϕ^B to denote the character functions of Alice and Bob (notice both Alice and Bob still only output one bit in $\mathcal{P}$). We use M to denote the distribution matrix of the distribution $\mathcal{D}$. We denote the eigenvector of M by $v_0, v_1, ..., v_{q-1}$ with corresponding eigenvalues $\lambda_0, ..., \lambda_{q-1}$. We assume that $\lambda_0 > \lambda_1 \geq \cdots \lambda_{q-1}$. Since M is regular, λ_0 is the unique largest eigenvalue that corresponds to eigenvector $\mathbf{1}_q$.

Since M is the distribution matrix, we know that the sum of all its entries is 1. Thus we have

$$1 = \mathbf{1}_q^T \cdot M \cdot \mathbf{1}_q = \lambda_0 \cdot \mathbf{1}_q^T \cdot \mathbf{1}_q = \lambda_0 \cdot q,$$

or $\lambda_0 = 1/q$. Since the scaled eigenvalue gap of M is ϵ, we know that the second largest absolute eigenvalue of M is $(1-\epsilon)/q$.

The distribution matrix of $\mathcal{D}^{\otimes n}$ is $M^{\otimes n}$. As in the proof to Theorem 5.1, we denote the character functions of Alice and Bob by ϕ^A and ϕ^B, respectively. Both ϕ^A and ϕ^B are vectors of dimension q^n. Since $\mathcal{P}$ is δ-locally uniform, we have

$$\left| \sum_{x \in \Sigma^n} \sum_{y \in \Sigma^n} \mathcal{D}^{\otimes n}(x,y) \cdot \phi^A(x) \right| \leq 2\delta$$

or $|\mathbf{1}_{q^n}^T \cdot M^{\otimes n} \cdot \phi^A| \leq 2\delta$. Since $\mathbf{1}_q$ is an eigenvector of M with eigenvalue $1/q$, $\mathbf{1}_{q^n}$ is an eigenvector of $M^{\otimes n}$ with eigenvalue $1/q^n$. Since M is symmetric, so is $M^{\otimes n}$. Thus we have $|\mathbf{1}_{q^n}^T \cdot \phi^A| \leq 2\delta \cdot q^n$. Similarly we have $|\mathbf{1}_{q^n}^T \cdot \phi^B| \leq 2\delta \cdot q^n$.

Again, as in the proof of Theorem 5.1, we can express the correlation of protocol $\mathcal{P}$ in terms of a quadratic form: $\mathsf{Cor}_{\mathcal{D}^{\otimes n}}[\mathcal{P}] = (\phi^A)^T \cdot M^{\otimes n} \cdot \phi^B$.

We diagonalize the matrix $M^{\otimes n}$. First we define a natural notion of inner product: $\langle A, B \rangle = \frac{1}{q^n} \sum_{x \in \Sigma^n} A[x]B[x]$. Since $M^{\otimes n}$ is symmetric, it has a set of eigenvectors that form an orthonormal basis. We denote the eigenvectors of $M^{\otimes n}$ by u_t with corresponding eigenvalues μ_t, where $t \in [q^n]$. We assume that $\mu_0 \geq \mu_1 \geq \cdots \mu_{q^n-1}$. By the property of the tensor product (see Lemma 5.2 after this proof), the eigenvalues μ_t are of the form $\prod_{i=1}^n \lambda_{k_i}$, where $k_i \in [q]$. Therefore $M^{\otimes n}$ also has a unique maximum eigenvalue $\lambda_0^n = 1/q^n$, which corresponds to the eigenvector $\mathbf{1}_q^{\otimes n} = \mathbf{1}_{q^n}$. The second largest value is $(1-\epsilon)/q \cdot 1/q^{n-1} = (1-\epsilon)/q^n$. In other words, $M^{\otimes n}$ has the same scaled eigenvalue gap as M.

Now we perform a Fourier analysis to vectors ϕ^A and ϕ^B. We write $\phi^A = \sum_{t\in[q^n]} \alpha_t \cdot u_t$ and $\phi^B = \sum_{t\in[q^n]} \beta_t \cdot u_t$. Then we have $\sum_t \alpha_t^2 \le 1$, $\sum_t \beta_t^2 \le 1$, and $|\alpha_0| \le 2\delta$, $|\beta_0| \le 2\delta$.

Now, putting things together, we have

$$\mathsf{Cor}_{\mathcal{D}^{\otimes n}}[\mathcal{P}] = (\phi^A)^T \cdot M^{\otimes n} \cdot \phi^B = q^n \cdot \sum_{t\in[q^n]} \alpha_t \cdot \beta_t \cdot \mu_t \le 4\delta^2 + (1-\epsilon)\cdot(1-4\delta^2) \le 1-\epsilon(1-4\delta^2)$$

■

Lemma 5.2 *Let A be an $a\times a$ matrix of eigenvectors $v_0, ..., v_{a-1}$, with corresponding eigenvalues $\lambda_0, ..., \lambda_{a-1}$. Let B be a $b \times b$ matrix of eigenvectors $u_0, ..., u_{b-1}$, with corresponding eigenvalues $\mu_0, ..., \mu_{b-1}$. Then the eigenvalues of the matrix $A\otimes B$ are $v_i \otimes u_j$ with corresponding eigenvalues $\lambda_i \cdot \mu_j$, for $i \in [a]$ and $j \in [b]$.*

Proof: We prove that for every $i \in [a]$ and $j \in [b]$, $(A\otimes B)(v_i \otimes u_j) = \lambda_i \cdot \mu_j \cdot (v_i \otimes u_j)$, which will imply that $(v_i \otimes u_j)$ is an eigenvector. Then, since $(A \otimes B)$ is an $(ab)\times(ab)$ matrix, it only has ab eigenvectors. Therefore this would imply our lemma.

Now we prove that $(A\otimes B)(v_i \otimes u_j) = \lambda_i \cdot \mu_j \cdot (v_i \otimes u_j)$.

$$\begin{aligned}
(A\otimes B)(v_i \otimes u_j)[(x,y)] &= \sum_{s\in[a],t\in[b]} (A\otimes B)_{(x,y),(s,t)} \cdot (v_i \otimes u_j)[(s,t)] \\
&= \sum_{s\in[a],t\in[b]} A_{x,s}\cdot B_{y,t} \cdot v_i[s] \cdot u_j[t] \\
&= \left(\sum_{s\in[a]} A_{x,s}\cdot v_i[s]\right) \cdot \left(\sum_{t\in[b]} B_{y,t}\cdot u_j[t]\right) \\
&= \lambda_i \cdot v_i[x] \cdot \mu_j \cdot u_j[y] \\
&= \lambda_i \cdot \mu_j \cdot (v_i \otimes u_j)[(x,y)]
\end{aligned}$$

Since the equation holds for all $x \in [a], y \in [b]$, we have $(A\otimes B)(v_i \otimes u_j) = \lambda_i \cdot \mu_j \cdot (v_i \otimes u_j)$. ■

Theorem 5.3 provides a general negative answer to the question of non-interactive correlation distillation. Notice the upper bound on the correlation is independent of n, the size of the input to the protocols. Therefore, if the noise model is regular, then Alice and Bob cannot distill the correlation any higher than what is dictated by the scaled eigenvalue gap, even if they are willing to collect many samples from the same model and then "concentrate" them them into one single symbol.

5.4 The Binary Erasure Noise Model

We prove a similar impossibility result for another noise model, namely the binary erasure noise model. Intuitively, this model describes the situation where Alice sends an unbiased bit to Bob, which is erased (and replaced by a special symbol $\perp$) with probability p.

Definition 5.10 (Binary Erasure Noise Model) *The* binary erasure noise model*, denoted by $\mathcal{E}_p$ is a tensor product noise model with base distribution $\mathcal{D}_{\mathcal{E}}$ over alphabet $\{0, 1, \perp\}$, defined as $\mathcal{D}_{\mathcal{E}}(0, 0) = \mathcal{D}_{\mathcal{E}}(1, 1) = (1 - p)/2$, $\mathcal{D}_{\mathcal{E}}(0, \perp) = \mathcal{D}_{\mathcal{E}}(1, \perp) = p/2$.*

Perhaps the binary erasure noise model is the simplest noise model that is not symmetric, and thus isn't regular. It is, however, a realistic one. Consider as example the situation where Alice and Bob receive their inputs by observing a pulsar. It is quite likely that the noise of the measurements by Alice and Bob are of the "erasure-type", i.e., the corruption of information can be detected. Furthermore, it is also possible that Alice and Bob have different measurement apparatus and different levels of accuracy. In the random beacon problem, Alice (as the beacon owner) might own a more sophisticated (and more expensive) measuring device with higher accuracy, while Bob (as the verifier) has a more noisy measurement device. An extreme case would be that Alice has near-perfect accuracy in her measurement, but Bob's measurement is noisy. Such a situation can be well approximated by the binary erasure noise model.

Notice that in this model, Alice's input is the uniform distribution over $\{0, 1\}$, and Bob's input is 0 and 1 with probability $(1 - p)/2$ each, and $\perp$ with probability p. A naïve protocol under this model only uses the first pair of the inputs. Alice outputs her bit, and Bob outputs his bit if his input is 0 or 1, and outputs a random bit if his input is $\perp$. This is a locally uniform protocol with correlation $1 - p$.

The next theorem shows that no protocol can do much better than the naïve protocol.

Theorem 5.4 (NICD for the Binary Erasure Model) *The correlation of any locally uniform protocol over the noise model $\mathcal{E}_p$ is at most $\sqrt{1-p(1-4\delta^2)}$.*

We suspect that it is not a tight bound, but it is sufficient to show that it is bounded away from 1 and is independent from n. Therefore, even with perfect accuracy in Alice's measurement, non-interactive correlation distillation is impossible if Bob's measurement is noisy.

Proof: We introduce more notations. A *binary string* is a string over alphabet $\{0,1\}$. For a binary string x, we denote its *Hamming weight* by $|x|$, which is the number of 1's in x. We call a vector v over alphabet $\{0,1,\perp\}$ an *extended bit vector*, and define its *degree*, denoted by $\deg(v)$, to be the number of $\perp$'s in it. An *error vector*, denoted by u is a vector over alphabet $\{\star,\perp\}$, and its *degree* also the number of $\perp$'s in it. Take a k-dimensional bit vector v and an n-dimensional error vector u of degree $(n-k)$, we define their *composition* to be an n-dimensional extended bit vector x defined as

$$x[i] = \begin{cases} v[j] & \text{if } u[i]=\star \text{ and } j = |\{l \;:\; 0 \le l < i, u[j]=\star\}| \\ \\ \perp & \text{if } u[i]=\perp \end{cases} \tag{5.11}$$

and we write this as $x = v \triangleright u$. As an example, we have $(1,0,1) \triangleright (\perp,\star,\star,\perp,\star) = (\perp,1,0,\perp,1)$. Notice that every extended bit vector x can be uniquely written as such a composition of a bit vector v and an error vector u. So we denote v to be the *extracted bit vector* of x, and write it as $v = [x]$; we denote u to be the *error vector* of x and write it as $u = \{x\}$.

For a bit vector x and an extended bit vector v, both of dimension n, we say x is *consistent* with v, if for every i such that $v[i] \neq \perp$, we have $x[i] = v[i]$. We denote this as $x \sqsubseteq v$.

For a bit vector x and an error vector u of degree d, we define the *restricted vector* of x with respect to u to be the unique $(n-d)$-dimensional bit vector v such that $x \sqsubseteq (v \triangleright u)$, and we write this as $v = x|_u$. The *excluded vector* of x with respect to u is the d-dimensional vector v' defined to be $v'[i] = x[k]$ where $k = |\{j \;:\; 0 \le j < i, u[j]=\star\}|$. We also write $x = v \overset{u}{\frown} v'$, and say x is *joined* by v and v' with respect to u.

We now fix a protocol $\mathcal{P}$ and consider its characteristic functions ϕ^A and ϕ^B (we omit the subscript n). Both are real functions over $\{0,1,\perp\}^n$. Both since in the erasure model, the input to Alice never contains $\perp$, we assume that ϕ^A is a function over $\{0,1\}^n$. We perform Fourier analysis to ϕ^A, using parity functions as the orthonormal basis.

$$\phi^A(x) = \sum_s \alpha_s \oplus_s (x) \tag{5.12}$$

where we have $\sum_s \alpha_s^2 \le 1$. Since $\mathcal{P}$ is δ-locally uniform, we have

$$|\alpha_0| \le 2\delta. \tag{5.13}$$

The analysis for ϕ^B is more complicated. We decompose ϕ^B into 2^n "sub-functions", according to the 2^n error vectors. For error vector u, we define a function ψ_u that maps $(n-k)$-dimensional bit vectors to $\{-1,+1\}$, where k is the degree of u. Then we perform a Fourier analysis for every sub-function, and write

$$\psi_u(x) = \sum_s \beta_{u,s} \oplus_s (x) \tag{5.14}$$

Again we have $\sum_s \beta_{u,s}^2 \le 1$ for every error vector u.

We define $\lambda = p/(1-p)$, then it is easy to see that the probability that Bob receives an extended error vector of degree d is $\lambda^d \cdot (1-p)^n$. Furthermore, it is easy to verify that

$$\sum_{u\in\{\star,\perp\}^n} \lambda^{\deg(u)} = \sum_{k=0}^{n} \binom{n}{k} \lambda^k = \frac{1}{(1-p)^n} \tag{5.15}$$

For the rest of the proof, we write λ^u as a shorthand for $\lambda^{\deg(u)}$.

Finally, we estimate the correlation between the outputs. We denote it by η and it is not hard to see that

$$\eta = \left(\frac{1-p}{2}\right)^n \sum_{u\in\{\star,\perp\}^n} \lambda^u \sum_x \phi^A(x)\psi_u(x|_u) \tag{5.16}$$

By substituting in the Fourier coefficients, we have

$$\begin{aligned}
\eta &= \left(\frac{1-p}{2}\right)^n \sum_{u\in\{\star,\perp\}^n} \lambda^u \sum_x \sum_{s\subseteq\{0,1\}^n} \sum_{t\subseteq\{0,1\}^{n-\deg(u)}} \alpha_s\beta_t \oplus_s (x) \oplus_t (x|_u) \\
&= \left(\frac{1-p}{2}\right)^n \sum_{u\in\{\star,\perp\}^n} \lambda^u \sum_{s\subseteq\{0,1\}^n} \sum_{t\subseteq\{0,1\}^{n-\deg(u)}} \alpha_s\beta_t \left(\sum_x \oplus_s(x) \oplus_t (x|_u)\right)
\end{aligned}$$

Now, we fix an error vector u of degree r, and fix sets s, t. We write $s = s_0 \cup s_1$, such that for every $i \in s_0$, we have $u[i] = \star$ and for every $i \in s_1$, we have $u[i] = \perp$. We write this as $s_0 = s|_u$. If $s_1 = \emptyset$, we say that s is *consistent* with u, and we write this as $s \sqsubseteq u$. Then we have

$$\begin{aligned}
\sum_{x \in \{0,1\}^n} \oplus_s(x) \oplus_t (x|_u) &= \sum_{v \in \{0,1\}^{n-d}} \sum_{v' \in \{0,1\}^d} \oplus_{s_0}(v) \oplus_{s_1} (v') \oplus_t (v) \\
&= \sum_{v \in \{0,1\}^{n-d}} \oplus_{s_0 \oplus t}(v) \sum_{v' \in \{0,1\}^d} \oplus_{s_1}(v')
\end{aligned}$$

So the only we we get non-zero as a result is when $s_0 = t$ and $s_1 = \emptyset$, which means $s = t$. Therefore, we have

$$\begin{aligned}
\eta &= (1-p)^n \sum_{u \in \{\star,\perp\}^n} \lambda^u \sum_{s \sqsubseteq u} \alpha_s \beta_{u,s|_u} \\
&\leq (1-p)^n \left(\sum_{u \in \{\star,\perp\}^n} \lambda^u \right)^{1/2} \cdot \left[\sum_{u \in \{\star,\perp\}^n} \lambda^u \left(\sum_{s \sqsubseteq u} \alpha_s \beta_{u,s|_u} \right)^2 \right]^{1/2} \quad \text{(Cauchy-Schwartz)} \\
&= (1-p)^{n/2} \cdot \left[\sum_{u \in \{\star,\perp\}^n} \lambda^u \cdot \left(\sum_{s \sqsubseteq u} \alpha_s^2 \right) \cdot \left(\sum_{s \sqsubseteq u} \beta_{u,s|_u}^2 \right) \right]^{1/2} \quad \text{(Eq. 5.15)} \\
&\leq (1-p)^{n/2} \cdot \left[\sum_{u \in \{\star,\perp\}^n} \lambda^u \cdot \left(\sum_{s \sqsubseteq u} \alpha_s^2 \right) \right]^{1/2} \quad \text{(Parseval, } \textstyle\sum_{s \sqsubseteq u} \beta_{u,s|_u}^2 \leq 1 \text{)} \\
&= (1-p)^{n/2} \cdot \left[\sum_s \alpha_s^2 \sum_{u : s \sqsubseteq u} \lambda^u \right]^{1/2} \\
&= (1-p)^{n/2} \cdot \left[\sum_s \alpha_s^2 \cdot (1+\lambda)^{n-|s|} \right]^{1/2} \\
&\leq (1-p)^{n/2} \left[(1+\lambda)^{n-1} (1 + 4\delta^2(1+\lambda)) \right]^{1/2} \quad \text{(Eq. 5.13)} \\
&= \sqrt{1 - p(1 - 4\delta^2)}
\end{aligned}$$

■

Chapter 6

A Positive Result on One-bit Correlation Distillation

The impossibility results from the previous chapter suggest that for many noise models, communication is essential for correlation distillation. Thus it is interesting to ask how much communication is essential. In particular, we were interested in the question "does a single bit of communication help?" We answer this question positively by presenting a protocol that non-trivially distills correlation from the binary symmetric noise model with one bit of communication. This result shows that even the minimal amount of communication is provably more powerful than no communication at all.

The result in this Chapter relative to this thesis is summarized in Figure 6.1.

Recall that over a binary symmetric noise model $\mathcal{S}_{n,p}$, no non-interactive, locally uniform protocols can have a correlation more than $1 - 2p$. Now, we consider protocols with one bit of communication. Suppose Alice sends one bit to Bob, which Bob receives with perfect accuracy. If we still only require Alice and Bob each to output a single bit, then the problem is trivial: Alice can generate an unbiased bit x and send it to Bob, and then Alice and Bob both output x. This protocol has perfect correlation. Thus, to make the problem non-trivial, we require that Alice and Bob must output two bits each. Suppose Alice outputs (X_1, X_2) and Bob outputs (Y_1, Y_2). We

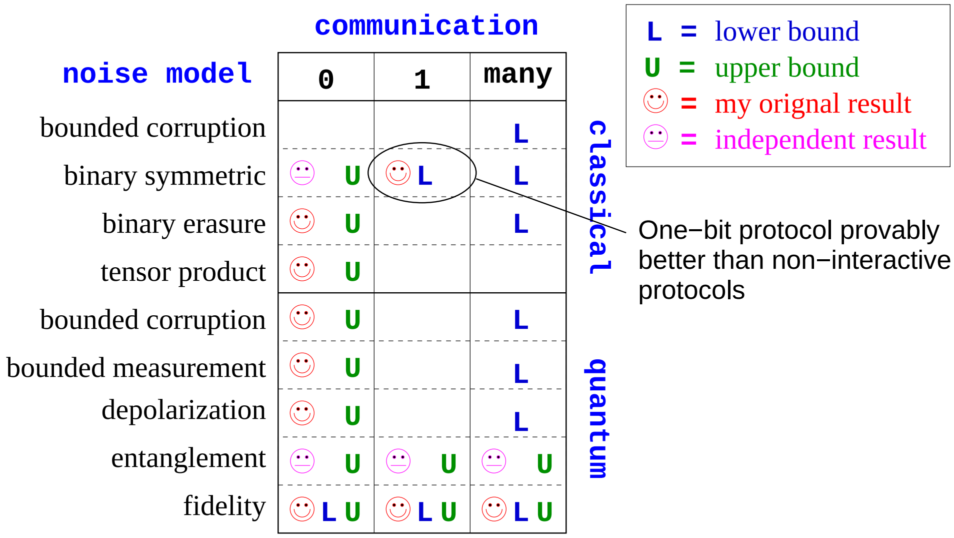

Figure 6.1: The Result in Chapter 6.

define the correlation of a protocol to be

$$2 \cdot \min_{i=1,2} \left\{ \mathsf{Prob}\left[X_i = Y_i\right] \right\} - 1$$

In this situation, we say a protocol is *locally uniform*, if both (X_1, X_2) and (Y_1, Y_2) are uniformly distributed.

Now we describe a locally uniform protocol of correlation about $1 - 3p/2$. The protocol is called the "AND" protocol. Both Alice and Bob only take the first two bits as their inputs. Alice directly output her bits, and sends the AND of her bits to Bob. Then, intuitively, Bob "guesses" Alice's bits using the Bayes rule and outputs them. A technical issue is that Bob has to "balance" his output so that the protocol is still locally uniform. The detailed description is in Figure 6.2.

We can easily verify (by a straightforward computation) the following result.

Theorem 6.1 (One-bit Protocol for the Binary Symmetric Model) *The AND protocol is a locally uniform protocol with correlation* $1 - \frac{3p}{2} + \frac{p^2}{4-2p}$. ■

This is a constant-factor improvement over the non-interactive case.

This result may seem a little surprising. It appears that Alice isn't fully utilizing the one-bit communication, since she is sending an AND of two bits, whose entropy is less than 1. It is tempting to speculate that by having Alice send the XOR of the two bits, Alice and Bob can obtain a better

STEP I Alice computes $r := a_1 \wedge a_2$, sends r to Bob, and outputs (a_1, a_2).

STEP II Bob, upon receiving r from Alice:

IF $r = 1$ THEN output $(1, 1)$.
ELSE IF $b_1 = b_2 = 1$ THEN output
- $(0, 0)$ with probability $p/(2-p)$;
- $(0, 1)$ with probability $(1-p)/(2-p)$;
- $(1, 0)$ with probability $(1-p)/(2-p)$;

ELSE output (b_1, b_2).

Alice receives input bits a_1, a_2, and Bob received input bits b_1, b_2, where (a_1a_2, b_1b_2) is drawn from $\mathcal{S}_p^{\otimes 2}$

Figure 6.2: The AND protocol

result, since Bob gets more information. Nevertheless, the XOR doesn't work, in some sense due to its "symmetry". Consider the case that Alice sends the XOR of her bits to Bob. Bob can compute the XOR of his bits, and if the two XOR's agree, Bob knows that with high probability, both his bits agrees with Alice's. However, if the two XOR's don't agree, Bob knows one of his bits is "corrupted", but he has no information about which one. Furthermore, however Bob guesses, he will be wrong with probability 1/2. On the other hand, in the AND protocol, if Bob receives a "1" as the AND of the bits from Alice, he knows for sure that Alice has $(1, 1)$ and thus he simply outputs $(1, 1)$; if $r = 0$ and $b_1 = b_2 = 1$, he knows that his input is "corrupted", and he "guesses" Alice's bit according to the Bayes rule of posterior probabilities. If Bob receives a "0" as the AND and $(b_1, b_2) \neq (1, 1)$, then the data looks "consistent" and Bob just outputs his bits. In this way, 1/4 of the time (when Bob receives a 1), Bob knows Alice's bits for sure and can achieve perfect correlation; otherwise Alice and Bob behave almost like in the non-interactive case, which gives $1 - 2p$ correlation. So the overall correlation is about $1/4 \cdot 1 + (3/4) \cdot (1 - 2p) = 1 - 3p/2$.

Chapter 7

Non-Interactive Entanglement Distillation

We study *non-interactive entanglement distillation* (NIED) protocols. As in the case of non-interactive classical correlation distillation, non-interactive entanglement distillation also serves as the most basic problem in the study of communication complexity of EDPs. Notice that a priori, it is not necessarily obvious that non-interactive protocols would be useless. In fact, Bennett et. al. [21] constructed a non-interactive entanglement distillation protocol for a specific noise model where Alice and Bob share a large number of identical copies of some pure state.[1] However, as we shall soon see, non-interactive entanglement distillation is impossible for a number of less "benign" noise models.

In this section, we only study protocols that only output one qubit pair, since these are the minimally "useful" protocols, and a lower bound on their fidelities suffices as a general lower bound. In particular, we consider three noise models, namely the bounded decoherence model, the bounded corruption model, and the depolarization model, and prove corresponding bounds on the fidelity of non-interactive EDPs over them. These bounds are tight or almost tight.

The results in this Chapter relative to this thesis are summarized in Figure 7.1.

[1] They call their scheme "entanglement concentration".

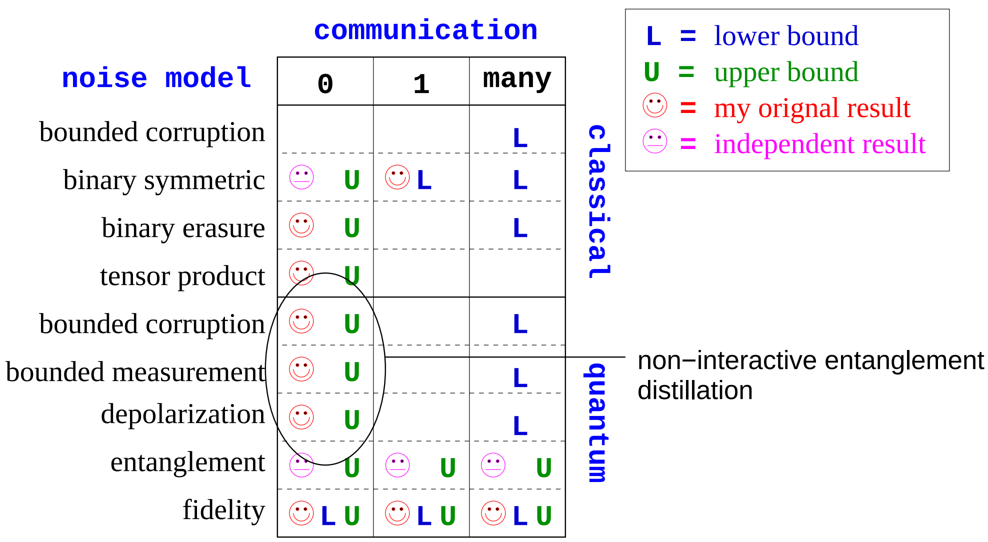

Figure 7.1: Results in Chapter 7.

7.1 The Bounded Measurement Model

We define the bounded measurement noise model, and prove a tight lower bound on the fidelity of non-interactive protocols over such a model. We first need some more notation. An *error indicator vector* is a n-dimensional vector from an alphabet $\mathbf{v} \in \{0, 1, *\}$. The *degree* of a vector $\mathbf{v}$, denoted by $\deg(\mathbf{v})$, is the number of entries in $\mathbf{v}$ that are not "$*$". Each $\mathbf{v}$ corresponds to a *measurement error state* $|\,\phi_{\mathbf{v}}\rangle = \bigotimes_{j=0}^{n-1} |\,\phi_j\rangle$, where

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

The *degree* of a measurement error state $|\,\phi_{\mathbf{v}}\rangle$ is the degree of $\mathbf{v}$.

Definition 7.1 (Bounded Measurement Model) *A* bounded measurement model *of parameter (n, t), denoted by $\mathcal{M}_{n,t}$, is an adversarial quantum noise model consisting of all measurement error states of degree at most t. In other words,*

$$\mathcal{M}_{n,t} = \{|\,\phi_{\mathbf{v}}\rangle \mid \deg(\mathbf{v}) \leq t\} \tag{7.1}$$

Intuitively, the bounded measurement model describes the situation where up to t (unknown) EPR pairs are measured in the computational basis (thus each pair results in either $|0\rangle|0\rangle$ or $|1\rangle|1\rangle$). Therefore, this model is in some sense more "benign" than the quantum bounded corruption model, where the corruptions on an EPR pair can be more general. However, this simpler model is already interesting enough to ensure a non-trivial result.

Theorem 7.1 (NIED for the Bounded Measurement Model) *The fidelity of any non-interactive, randomized public-coin entanglement distillation protocols over a bounded measurement model $\mathcal{M}_{n,r}$ is at most $1 - \frac{r}{2n}$.*

Notice that there exists a very simple non-interactive, randomized public-coin protocol that achieves a fidelity of $1 - \frac{r}{2n}$. Alice and Bob use their shared randomness to select a random input qubit pair to output. If this pair is not measured, it has fidelity 1; if the pair is measured, it has fidelity $\frac{1}{2}$. Clearly, a random pair is measured with probability at most $\frac{r}{n}$. Therefore, the overall fidelity is at least $1 - \frac{r}{2n}$, and the upper bound in Theorem 7.1 is tight.

Despite the fact that the matching upper bound is almost trivial, the proof to this lower bound does not appear so.

Proof: (of Theorem 7.1) We consider a slightly different noise model, where r *random* EPR pairs are measured. This corresponds to the density matrix

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

We shall prove that no *deterministic* non-interactive protocol can have a fidelity higher than $1 - \frac{r}{2n}$ if ρ is the input. Then, we conclude that no share-randomized protocol can have a fidelity higher than $1 - \frac{r}{2n}$, too, since fidelity is a linear function.

Consider a deterministic non-interactive protocol $\mathcal{P}$. Notice $\mathcal{P}$ doesn't involve any communication, we can model it as Alice and Bob both applying a unitary operation to their share of qubits, output the first qubits 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

$$\begin{aligned} U_A|\,x\rangle &\longrightarrow |\,\phi_x\rangle \\ U_B|\,x\rangle &\longrightarrow |\,\psi_x\rangle \end{aligned}$$

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

$$\frac{1}{2^r\binom{n}{r}} \sum_{\deg(\mathbf{v})=r} \left[[\mathsf{F}^{\mathsf{b}}((U_A \otimes U_B)|\phi_{\mathbf{v}}\rangle\langle\phi_{\mathbf{v}}|(U_A \otimes U_B)^\dagger)\right] \le 1 - \frac{r}{2n}, \tag{7.2}$$

which implies the theorem.

By Lemma 2.3, Eq.(7.2) is equivalent to

$$\frac{1}{2^r\binom{n}{r}} \sum_{\deg \mathbf{v}=r} \left[\sum_{U\in\{I,X,Y,Z\}} \langle\phi_{\mathbf{v}}\,|(U_A \otimes U_B)^\dagger (U \otimes U^*)(U_A \otimes U_B)|\,\phi_{\mathbf{v}}\rangle \right] \le 4(1 - \frac{r}{2n}) \tag{7.3}$$

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

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

for any unitary operation U. So, Eq.(7.3) is equivalent to

$$\frac{1}{2^n\binom{n}{r}} \sum_{\deg \mathbf{v}=r} \sum_{x\sqsubseteq\mathbf{v}} \sum_{y\sqsubseteq\mathbf{v}} \sum_{U\in\{I,X,Y,Z\}} \langle\phi_x\,|U|\,\phi_y\rangle \cdot \langle\psi_x\,|U^*|\,\psi_y\rangle \le 4(1 - \frac{r}{2n}) \tag{7.4}$$

However, by Cauchy-Schwartz, we have

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

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

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

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

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

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 $x \neq y$,

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

Also notice that by Lemma 2.2, we have $\sum_{U \in \{I,X,Y,Z\}} |\langle \phi_x | U | \phi_x \rangle|^2 \leq 2$ for any x.

Putting things together, we have

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

Similarly, we have

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

too.

Thus we have

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

which proves (7.4). ■

7.2 The Bounded Corruption Model

We prove a similar upper bound on the fidelity of non-interactive protocols over a bounded corruption model.

Theorem 7.2 (NIED for the Bounded Corruption Model) *The fidelity of any non-interactive, randomized public-coin entanglement distillation protocols over a quantum bounded corruption model $\mathcal{B}^{\mathsf{q}}_{n,r}$ is at most $1 - \frac{r}{2n}$.*

Notice that if Alice and Bob use their shared random bits to select an input pair and output them, they will achieve a fidelity of $1-\frac{r}{n}$. So this upper bound is almost tight (up to a constant factor).

Proof: (of Theorem 7.2) As in the proof for Theorem 7.1, we consider a different "random corruption" noise model, where r EPR pairs are randomly chosen and each is independently replaced by a random Bell state. We shall prove that the fidelity of any deterministic, non-interactive protocol over such a noise model is at most $1-\frac{r}{2n}$, which will imply our theorem.

It is easy to verify that

$$\frac{1}{4}\left(\Phi^{+}+\Phi^{-}+\Psi^{+}+\Psi^{-}\right)=\frac{1}{4}\left(|00\rangle\langle 00|+|01\rangle\langle 01|+|10\rangle\langle 10|+|11\rangle\langle 11|\right)=\frac{I}{4} \tag{7.5}$$

So we can interpret the random corruption noise model as randomly choosing r EPR pairs and replace each of them by the completely mixed state $I/4$.

We present more notations and definitions. As *corruption indicator vector*, often denoted by $\mathbf{u}$, is an n-dimensional vector, whose each entry is an element from alphabet $\{00,01,10,11,*\}$. Its *degree* is the number of entries that are not $*$. There are $4^r\binom{n}{r}$ corruption indicator vectors of degree r, where each $\mathbf{u}$ corresponds to a unique bipartite state $|\,\psi_{\mathbf{u}}\rangle$ in the following way:

$$|\,\psi_{\mathbf{u}}\rangle=\bigotimes_{j=0}^{n-1}|\,\phi_j\rangle,\quad \text{where }\ |\,\phi_j\rangle=\begin{cases}|\,0\rangle^A|\,0\rangle^B & \text{if } \mathbf{u}[j]=00\\ |\,0\rangle^A|\,1\rangle^B & \text{if } \mathbf{u}[j]=11\\ |\,1\rangle^A|\,0\rangle^B & \text{if } \mathbf{u}[j]=10\\ |\,1\rangle^A|\,1\rangle^B & \text{if } \mathbf{u}[j]=11\\ \Phi^+ & \text{if } \mathbf{u}[j]=*\end{cases} \tag{7.6}$$

We call such an $|\,\psi_{\mathbf{u}}\rangle$ an *corruption error state.*

An $2n$-bit string x is *consistent* with a corruption indicator vector $\mathbf{u}$, if $x[j];x[n+j]=\mathbf{u}[j]$ for all j such that $\mathbf{v}[j]\neq *$, and $x[j]=x[n+j]$ for all j such that $\mathbf{v}[j]=*$. We write this as $x\sqsubseteq\mathbf{u}$. There are 2^{n-r} bit-strings consistent with a corruption indicator vector of degree r. We often view x as the concatenation of 2 n-bit string: $x=l;r$, and we write them as $l=\mathsf{LT}(x)$ and $r=\mathsf{RT}(x)$.

With the notations, we can write the corruption error states as

$$\psi_{\mathbf{u}} = \frac{1}{2^{(n-r)/2}} \sum_{x \sqsubseteq \mathbf{u}} |\,\mathsf{LT}(x)\rangle^A |\,\mathsf{RT}(x)\rangle^B \tag{7.7}$$

We define the *discrepancy* of a $2n$-bit string x to be $\mathsf{DIS}(x) = \mathsf{LT}(s) \oplus \mathsf{RT}(s)$, where "$\oplus$' stands for bit-wise XOR. The *degree of discrepancy* of x is $|\mathsf{DIS}(x)|$, the Hamming weight of $\mathsf{DIS}(x)$. Clearly, there are $\binom{n}{d}2^n$ 0-1 vectors of dimension $2n$ having degree of discrepancy d. Furthermore, if x has degree of discrepancy d, then the number of degree-r corruption indicator vectors $\mathbf{u}$ such that $x \sqsubseteq \mathbf{u}$ is $\binom{n-d}{r-d}$. This is because for every j such that $x[j] \neq x[n+j]$, we must have $\mathbf{u}[j] = x[j]; x[n+j]$ in order to have $x \sqsubseteq \mathbf{u}$. So the only freedom for $\mathbf{u}$ is to put $(n-r)$ $*$'s in the $n-d$ places where $x[j] = x[n+j]$.

Consider a deterministic non-interactive protocol $\mathcal{P}$. Again, since $\mathcal{P}$ is non-interactive, we can model it by a pair of unitary operators (U_A, U_B), such that $\mathcal{P}$ consists of Alice and Bob each applying their operators, outputs the first qubits, and discarding the rest. We write the unitary operators as

$$\begin{aligned} U_A|\,x\rangle &\longrightarrow |\,\phi_x\rangle \\ U_B|\,x\rangle &\longrightarrow |\,\psi_x\rangle \end{aligned}$$

Then as in the proof to Theorem 7.1, we shall prove that

$$\frac{1}{4^r\binom{n}{r}} \sum_{\deg \mathbf{u}=r} \left[\sum_{U \in \{I,X,Y,Z\}} \langle \psi_{\mathbf{u}}\,|(U_A \otimes U_B)^\dagger (U \otimes U^*)(U_A \otimes U_B)|\,\psi_{\mathbf{u}}\rangle \right] \leq 4(1 - \frac{r}{2n}) \tag{7.8}$$

which implies our theorem.

Notice that

$$(U_A \otimes U_B)|\,\psi_{\mathbf{u}}\rangle = \frac{1}{2^{(n-r)/2}} \sum_{x \sqsubseteq \mathbf{u}} |\,\phi_{\mathsf{LT}(x)}\rangle |\,\psi_{\mathsf{RT}(x)}\rangle$$

and so we have

$$\langle\psi_{\mathbf{u}}|(U_A\otimes U_B)^\dagger(U\otimes U^*)(U_A\otimes U_B)|\psi_{\mathbf{u}}\rangle = \frac{1}{2^{n-r}}\sum_{x\sqsubseteq\mathbf{u}}\sum_{y\sqsubseteq\mathbf{u}}\langle\phi_{\mathsf{LT}(x)}|U|\phi_{\mathsf{LT}(y)}\rangle\cdot\langle\psi_{\mathsf{RT}(x)}|U^*|\psi_{\mathsf{RT}(y)}\rangle$$

So we only need to prove that

$$\frac{1}{2^{n+r}\binom{n}{r}}\sum_{\deg\mathbf{u}=r}\sum_{x\sqsubseteq\mathbf{u}}\sum_{y\sqsubseteq\mathbf{u}}\sum_{U\in\{I,X,Y,Z\}}\langle\phi_{\mathsf{LT}(x)}|U|\phi_{\mathsf{LT}(y)}\rangle\cdot\langle\psi_{\mathsf{RT}(x)}|U^*|\psi_{\mathsf{RT}(y)}\rangle\leq 4(1-\frac{r}{2n}) \quad (7.9)$$

By Cauchy-Schwartz, we have

$$\begin{aligned}&\sum_{\deg\mathbf{u}=r}\sum_{x\sqsubseteq\mathbf{u}}\sum_{y\sqsubseteq\mathbf{u}}\sum_{U\in\{I,X,Y,Z\}}\langle\phi_{\mathsf{LT}(x)}|U|\phi_{\mathsf{LT}(y)}\rangle\cdot\langle\psi_{\mathsf{RT}(x)}|U^*|\psi_{\mathsf{RT}(y)}\rangle\\ \leq\ &\left(\sum_{\deg\mathbf{u}=r}\sum_{x\sqsubseteq\mathbf{u}}\sum_{y\sqsubseteq\mathbf{u}}\sum_{U\in\{I,X,Y,Z\}}|\langle\phi_{\mathsf{LT}(x)}|U|\phi_{\mathsf{LT}(y)}\rangle|^2\right)^{\frac{1}{2}}\cdot\\ &\left(\sum_{\deg\mathbf{u}=r}\sum_{x\sqsubseteq\mathbf{u}}\sum_{y\sqsubseteq\mathbf{u}}\sum_{U\in\{I,X,Y,Z\}}|\langle\psi_{\mathsf{RT}(x)}|U^*|\psi_{\mathsf{RT}(y)}\rangle|^2\right)^{\frac{1}{2}}\end{aligned}$$

Now we estimate

$$\sum_{\deg\mathbf{u}=r}\sum_{x\sqsubseteq\mathbf{u}}\sum_{y\sqsubseteq\mathbf{u}}\sum_{U\in\{I,X,Y,Z\}}|\langle\phi_{\mathsf{LT}(x)}|U|\phi_{\mathsf{LT}(y)}\rangle|^2$$

Notice we can write x as $x=\mathsf{LT}(x);(\mathsf{LT}(x)\oplus\mathsf{DIS}(x))$ and y as $y=\mathsf{LT}(y);(\mathsf{LT}(y)\oplus\mathsf{DIS}(y))$. If there exists an extended indicator vector $\mathbf{u}$ such that $x\sqsubseteq\mathbf{u}$ and $y\sqsubseteq\mathbf{u}$, we must have $\mathsf{DIS}(x)=\mathsf{DIS}(y)$. This is because that for every j such that $\mathsf{DIS}(x)[j]=1$, $x[j]$ and $x[n+j]$ differ. Thus we must have $\mathbf{v}[j]=x[j];x[n+j]$, which implies that $\mathbf{v}[j]=y[j];y[n+j]$, and $\mathsf{DIS}(y)[j]=1$. In fact, for every j such that $\mathsf{DIS}(x)[j]=1$, we have $x[j]=y[j]$ and $x[n+j]=y[n+j]$.

So we have

$$\begin{aligned}&\sum_{\deg \mathbf{u}=r}\sum_{x\sqsubseteq \mathbf{u}}\sum_{y\sqsubseteq \mathbf{u}}\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_{\mathsf{LT}(x)}|U|\,\phi_{\mathsf{LT}(y)}\rangle|^2\\ =\ &\sum_{a\in\{0,1\}^n}\sum_{b\in\{0,1\}^n}\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_a|U|\,\phi_b\rangle|^2 \sum_{c\in\{0,1\}^n}\ \sum_{\deg \mathbf{u}=r:\ [(a;(a\oplus c))\sqsubseteq \mathbf{u}]\wedge[(b;(b\oplus c))\sqsubseteq \mathbf{u}]} 1\end{aligned}$$

by a substituting a for $\mathsf{LT}(x)$, b for $\mathsf{LT}(y)$, and c for $\mathsf{DIS}(x)$.

Now we fix a and b, and compute

$$\sum_{c\in\{0,1\}^n}\ \sum_{\deg \mathbf{u}=r:\ [(a;(a\oplus c))\sqsubseteq \mathbf{u}]\wedge[(b;(b\oplus c))\sqsubseteq \mathbf{u}]} 1$$

We define $k=|a\oplus b|$. For every j where $a[j]\neq b[j]$, we must have $c[j]=0$ and $\mathbf{u}[j]=*$. For every j where $a[j]=b[j]$, if we have $c[j]=1$, then we must have $\mathbf{u}[j]=a[j];(a[j]\oplus 1)$; if we have $c[j]=0$, then $\mathbf{u}$ can be either $a[j];a[j]$ or $*$. Therefore, of $n-k$ positions where $a[j]=b[j]$, r would be chosen where $\mathbf{u}$ has a non-$*$ entry. Of these r places, one has the freedom to choose $c[j]=0$ or $c[j]=1$. For all other places, $c[j]=0$ and $\mathbf{u}=*$. So we have

$$\sum_{c\in\{0,1\}^n}\ \sum_{\deg \mathbf{u}=r:\ [(a;(a\oplus c))\sqsubseteq \mathbf{u}]\wedge[(b;(b\oplus c))\sqsubseteq \mathbf{u}]} 1 = 2^r\cdot\binom{n-k}{r}$$

In other words,

$$\sum_{\deg \mathbf{u}=r}\sum_{x\sqsubseteq \mathbf{u}}\sum_{y\sqsubseteq \mathbf{u}}\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_{\mathsf{LT}(x)}|U|\,\phi_{\mathsf{LT}(y)}\rangle|^2 = \sum_{a\in\{0,1\}^n}\sum_{b\in\{0,1\}^n}\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_a|U|\,\phi_b\rangle|^2\cdot 2^r\cdot\binom{n-|a\oplus b|}{r} \tag{7.10}$$

Since $|\,\phi_a\rangle$'s are orthogonal, we have

$$\sum_a\sum_b\sum_{U\in\{I,X,Y,Z\}} |\langle\phi_a|U|\,\phi_b\rangle|^2 \le 2^{n+2}$$

Also by Lemma 2.2, we have

$$\sum_a |\langle\phi_a|U|\,\phi_a\rangle|^2 \le 2^{n+1}$$

Therefore

$$
\begin{aligned}
& \sum_{a\in\{0,1\}^n} \sum_{b\in\{0,1\}^n} \sum_{U\in\{I,X,Y,Z\}} |\langle \phi_a |U| \phi_b\rangle|^2 \cdot 2^r \cdot \binom{n-|a\oplus b|}{r} \\
\leq \; & \sum_a |\langle \phi_a |U| \phi_a\rangle|^2 \cdot 2^r \left[\binom{n}{r} - \binom{n-1}{r}\right] + 2^r \binom{n-1}{r} \sum_{a\in\{0,1\}^n} \sum_{b\in\{0,1\}^n} \sum_{U\in\{I,X,Y,Z\}} |\langle \phi_a |U| \phi_b\rangle|^2 \\
\leq \; & 2^{n+r+1} \left[\binom{n}{r} - \binom{n-1}{r}\right] + 2^{n+r+2} \binom{n-1}{r} \\
= \; & 2^{n+r+2} \binom{n}{r} (1 - \frac{r}{2n})
\end{aligned}
$$

which implies (7.9), which implies the theorem. ■

7.3 The Depolarization Model

Depolarization Model We define the depolarization noise model, which is a commonly used model for quantum noises [92, 69]. Intuitively, a depolarization model of parameter p describes the situation where each of Bob's qubits is replaced by a completely mixed state independently with probability p. In particular, if Alice and Bob initially share the Bell state Φ^+, then the "depolarization" noise moves it to

$$\rho_p = (1 - \frac{3p}{4})|\Phi^+\rangle\langle\Phi^+| + \frac{p}{4}(|\Phi^-\rangle\langle\Phi^-| + |\Psi^+\rangle\langle\Psi^+| + |\Psi^-\rangle\langle\Psi^-|) \tag{7.11}$$

which is also known as the "Werner state" [92].

Definition 7.2 (Depolarization Model) *A* depolarization model *of parameter* (n, p) *is a probabilistic quantum noise model defined as* $\mathcal{D}_{n,p} = \rho_p^{\otimes n}$.

Theorem 7.3 (NIED for the Depolarization Model) *The fidelity of any non-interactive, randomized public-coin entanglement distillation protocols over a depolarization model* $\mathcal{D}_{n,p}$ *is at most* $1 - \frac{p}{2}$.

Notice that there exists a very simple non-interactive protocol of fidelity $1 - \frac{3p}{4}$. If Alice and Bob simply outputs the first qubit of their shares, the fidelity of the output is $1 - \frac{3p}{4}$. Notice that this protocol is deterministic. Therefore the upper bound in Theorem 7.3 is almost tight (up to a constant factor).

Proof: **(of Theorem 7.3)** Notice that in the depolarization model, the probability that r EPR pairs are corrupted is $\binom{n}{r}p^r(1-p)^{n-r}$. Conditioned on that r pairs are corrupted, each of these r pairs are replaced by a completely mixed state, and this it is exactly the "random corruption" model in the proof of Theorem 7.2. So in this case, the fidelity of any non-interactive protocol is at most $1 - r/2n$. Thus, the overall fidelity of any non-interactive protocol is at most $\sum_r \binom{n}{r}p^r(1-p)^{n-r} \cdot (1 - r/2n) = 1 - p/2$. ■

Chapter 8

The Fidelity Noise Model

We introduce the fidelity noise model and study the communication complexity of entanglement distillation protocols over this model. We start by discussing the motivation for this noise model, namely, the problem of General Entanglement Extraction. Then we introduce the model and present our results.

The results in this Chapter relative to this thesis are summarized in Figure 8.1.

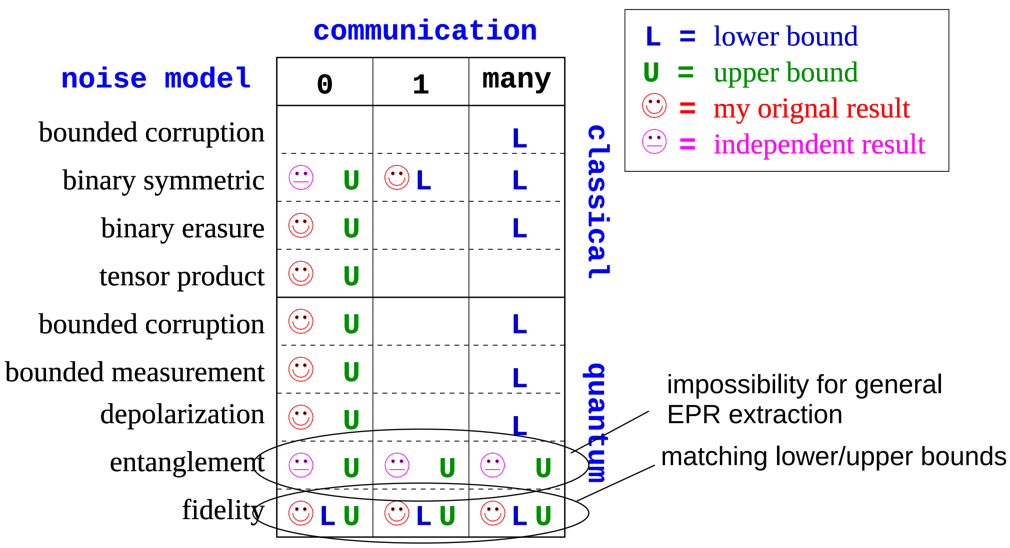

Figure 8.1: Results in Chapter 8.

8.1 Motivation: General Entanglement Extraction

The problem of general entanglement extraction is formulated (informally) as follows. Given an arbitrarily state of certain entanglement (say k), is it possible for Alice and Bob to extractor "high-quality" entanglement, namely EPR pairs? This problem is naturally motivated by an analogy between classical randomness extraction and quantum entanglement distillation.

8.1.1 Classical Randomness Extraction

Classical randomness extraction is a fascinating topic in theoretical computer science by itself. The motivation for study randomness extraction is that randomness plays an important role in classical computation (see Motwani and Raghavan [64] for a comprehensive explanation), but it can be very expensive, if not impossible, to have a perfect random source that produces unbiased, uncorrelated random bits. Therefore, it is very natural to ask if it is possible to perform randomized computation using less-than-perfect random sources. In particular, is it possible to have an automatic process to convert any randomized computation that was designed to have a perfect random source as input into one that works with imperfect random sources?

A series of results established by various researchers answered positive to this question, and the notion of randomness extractors was developed along this line of research. Intuitively, a randomness extractor is a procedure that converts input from an imperfect random source to almost-perfect random bits as its output. Technically, an extractor also takes a small number of perfect random bits from an auxiliary input. But the size of auxiliary input is normally logarithmically small as compare to the size of its main input. See Figure 8.2.

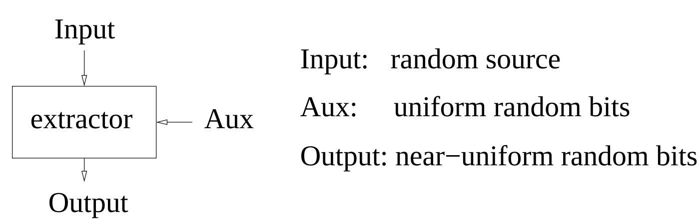

Figure 8.2: Classical randomness extractor

We briefly review some of the work on extractors and refer the readers to Nisan and Ta-Shma [70] and Shaltiel [81] for a more comprehensive and up-to-date survey. In the early stages of research on extractors, people have considered various specific models of "imperfect random bits". Von Neumann [66] showed that a linear number of perfect random bits can be extracted from independent tosses of a biased coin with unknown bias. Blum [11] extended the model of a biased coin to a Markov chain. Santha and Vazirani [80] considered extractors with many independent, yet adversarial random sources, as input. This contrasts with the modern stage, started by Nisan and Zuckerman [71], where researcher started to study extractors over *arbitrary* input. Today, the state-of-art extractors can extract near-perfect random bits from *random source* [89, 79]. We also have a quite good understanding about the limit of extractors. For example, we know that the yield (size of the output) of an extractor is determined by the *min entropy* of the input, and that the size of the auxiliary input needs to be logarithmic in the size of input. On the other hand, there exist constructions of extractors that match these limits [89, 79].

8.1.2 Similarity Between Extractors and EDPs

We discuss the similarity between classical extractors and quantum entanglement distillation protocols. Entanglement plays a central role in quantum information theory and quantum computation. It was argued that entanglement is the essential physical phenomenon that gives quantum computation its power of exponential speed-up over classical computation. Although it is still under heated debate and relentless research whether entanglement is essential for quantum computation [24, 48, 20], it is widely believed that that entanglement plays a crucial part for quantum information theory. However, somewhat like in the case of classical randomness, it is very hard to have a perfect source of entanglement. EPR pairs, as with currently technology, are notoriously hard to maintain. They decohere very easily and become "less entangled". As randomness extractors convert less-than-perfect random bits into near-perfect ones, entanglement distillation protocols convert less entangled quantum states into almost perfect EPR pairs.

There exist even deeper similarities. An extractor, being a deterministic procedure, cannot create randomness by itself. It needs to "distill" the randomness from the input bits into randomness of the purest form, namely unbiased, uncorrelated random bits. An entanglement distillation protocol, being an LOCC protocol, cannot create entanglement by itself. Therefore an EDP also needs to distill the entanglement from the input into EPR pairs, which are the entanglement of the purest form — each pair is maximally entangled and separable from the rest.

Moreover, the early stage of searches on EDPs greatly resembles that on the randomness extractors, in that people have considered various specific models of "imperfect EPR pairs" and constructed protocols over these specific models. As an example, the first work we are aware of on EDPs is by Bennett, Bernstein, Popescu, and Schumacher [21], which used the model where many identical copies of the pure state $|\phi\rangle = (\cos\theta|01\rangle + \sin\theta|10\rangle)$ is given as the input. The resemblance of this model, as well as the solution, to the the biased coin model used by von Neumann [66] is striking. More complicated models were proposed later, as Bennett, Brassard, Popescu, Schumacher, Smolin, and Wootters [22] studied the case where the input is identical copies of a mixed state. Horodecki, Horodecki, and Horodecki [42, 45] and Rains [75, 76, 77] studied the case where the input is many identical copies of a known pure state. Notice that the classical counterpart of this state would be an input with known distribution, for which case the problem of randomness extraction was long solved by Shannon [82]. This sharp contrast somewhat demonstrates the difficulty of quantum information theory, as very simple problems in classical information theory can become highly non-trivial in the quantum case.

However, despite the similarities and the correspondence between the early stages in research on randomness extractors and entanglement distillation protocols, there has not been a counterpart of the modern stage of extractors in the study of EDPs. In other words, there hasn't much work on EDPs over arbitrary entangled states. This observation naturally motivates the entanglement noise model and the study on EDPs over such a model.

8.1.3 The Entanglement Noise Model and the Impossibility Result

We describe the entanglement noise model, which contains all pure states of a certain amount of entanglement.

Definition 8.1 (Entanglement Noise Model) *A* entanglement noise model *of parameter (n, k), denoted by $\mathcal{E}_{n,k}$, is an adversarial quantum noise model consisting of all $2n$-qubit pure states of entanglement at least k. In other words,*

$$\mathcal{E}_{n,k} = \{|\phi\rangle \in \mathcal{H}_{2^{2n}} \mid \mathsf{E}(\phi) \geq k\} \tag{8.1}$$

Unfortunately, there don't exist entanglement distillation protocols over the entanglement noise model. This is true even if we restrict ourselves to starting states with the maximum possible entanglement and only requires the protocol to output a single EPR pair Φ^+.

Theorem 8.1 (Entanglement Model) *There do not exist perfect* $(n,1)$*-protocols over the entanglement noise model* $\mathcal{E}_{n,n}$*.*

Proof: Consider a quantum system of $2n$ qubits. The maximum possible entanglement of such a system is n. Unlike in the classical world where there is just one probability distribution over 2^n elements with entropy n (the uniform distribution), there are infinitely many quantum states with entanglement n. Namely, any quantum state of the form

$$|\,\phi\rangle = \sum_{i=0}^{N-1} \alpha_i |\,i\rangle |\,i\rangle \tag{8.2}$$

with $|\alpha_i|^2 = 1/N$ for all $i \in \{0, \dots, N-1\}$ has entanglement $\log N$, where we denote $N = 2^n$. In particular, this includes

$$|\,\phi_a\rangle = \sum_{b=0}^{N-1} \frac{1}{\sqrt{N}} e^{2i\, ab\pi/N} |\,b\rangle |\,b\rangle$$

for $a \in \{1, \dots, N\}$. Assume that we have a protocol that extracts Φ^+ from any $|\,\phi_a\rangle$. This means that, given $|\,\phi_a\rangle$, the protocol ends with the final state of the form $\Phi^+ \otimes |\,\phi'_a\rangle$. We consider running this protocol on the mixed state ρ that is $|\,\phi_0\rangle$ with probability $1/N$, $|\,\phi_1\rangle$ with probability $1/N$, ..., $|\,\phi_{N-1}\rangle$ with probability $1/N$. Then, the final state is of the form $\Phi^+ \otimes \rho'$ where ρ' is some mixed state.

The problem is that ρ is equivalent to the mixed state that is $|\,0\rangle|\,0\rangle$ with probability $1/N$, $|\,1\rangle|\,1\rangle$ with probability $1/N$, ..., $|\,N-1\rangle|\,N-1\rangle$ with probability $1/N$. (This equivalence can be verified by writing out the density matrices of both states.) None of the states $|\,i\rangle|\,i\rangle$ is entangled, so the mixed state obtained by combining them is also not entangled. Yet, since this mixed state is equivalent to ρ, it gets transformed into $\Phi^+ \otimes \rho'$, which is entangled.

We have constructed a protocol that transforms a disentangled starting state into entangled end state without quantum communication. Since this is impossible [21], our assumption is wrong and there is no protocol that extracts any Φ^+ from an arbitrary $|\,\phi_a\rangle$. ■

The argument described above is still valid if we relax the requirement to extracting a state close to Φ^+.

This is a clear distinction between the situation of classical randomness extraction and quantum entanglement extraction. In the classical case, all the probabilities are non-negative real numbers, and the min entropy of a random distribution already characterizes the distribution well. In the quantum case, the magnitudes are complex numbers, and the entanglement alone isn't good enough to describe the state. Even more interestingly, since one has the freedom to switch bases in quantum, we can build a mixed state which is a mixture of maximally entangled states, yet the mixed state itself is completely disentangled. This phenomenon doesn't seem to have a counterpart in classical probability.

8.2 The Fidelity Noise Model

With the motivation of studying EDPs for a general class of noise models and the impossibility result for the (too general) entanglement noise model, we consider the fidelity noise model as one that is still quite general, but also useful. Intuitively, the entanglement noise model fails because there exists many maximally entangled states that are orthogonal to each other, and no protocol can work with all of them. Therefore, some "closeness" condition is needed, i.e., we need some guarantee that the input state is close to a fixed maximally entangled state. This intuition naturally leads to the fidelity noise model, which, informally speaking, describes the situation where the input state has a reasonably high fidelity with the perfect EPR pairs.

We give the definition of the fidelity noise model.

Definition 8.2 (Fidelity Noise Model) *A* fidelity noise model *of parameter* (n, a)*, denoted by* $\mathcal{F}_{n,a}$*, is an adversarial quantum noise model consisting of all* $2n$*-qubit mixed states of fidelity at least* a*. In other words,*

$$\mathcal{F}_{n,a} = \{\rho \in \mathcal{H}_{2^{2n}} \mid \mathsf{F}(\rho) \geq a\} \tag{8.3}$$

This model was also independently considered by Lo and Chau [57] and Shor and Preskill [84] in proving the security of the BB84 quantum key distribution protocol [16], and by Barnum et. al. [23] in studying the so-called "purity-testing protocols".

8.3 Our Results

We present our results here, where are arranged in three parts. The first part is concerned with absolute protocols, whereas we prove both lower and upper bounds for the quality of the optimal protocols; the second part relates conditional protocols with so-called "purity-testing protocols" and we construct a protocol called "random hashing" that works with the fidelity model; in the third part, we prove an almost tight bound (up to an additive constant) on the communication complexity of EDPs over the fidelity model. Our result implies that the "random hashing" protocol is optimal.

8.3.1 Part I: Absolute Protocols

We prove that no absolute protocol would work well over a fidelity noise model. In fact, we can prove an even stronger result, which extends to protocols that accept perfect EPR pairs as auxiliary inputs.

Protocols with Auxiliary Input We consider protocols with auxiliary inputs as a slight extension to "standard" entanglement distillation protocols. In addition to the input states, Alice and Bob also receive k EPR pairs (each pair is shared between Alice and Bob) as auxiliary inputs. Obviously a protocol with auxiliary input would be more powerful than one without. An immediate example is that a deterministic protocol with auxiliary inputs can simulate a randomized public-coin protocol, since Alice and Bob can use the shared EPR pairs to simulate shared random bits.

Theorem 8.2 (Absolute Protocols for the Fidelity Model) *The fidelity of any (n, m)-protocol with $k < m$ EPR pairs as auxiliary inputs over a fidelity model $\mathcal{F}_{n,1-\epsilon}$ is at most $1 - \frac{2^m - 2^k}{2^m} \frac{2^n}{2^n - 1}\epsilon$. Moreover, this upper bound is tight, in that for every n, m, n, there exists an (n, m)-protocol using k EPR pairs as auxiliary inputs of fidelity $1 - \frac{2^m - 2^k}{2^m} \frac{2^n}{2^n - 1}\epsilon$.*

Typically, the size of the auxiliary input, k is very small compared to the size of the input and the output. Since a protocol with k EPR pairs of auxiliary input can trivially output k perfect EPR pairs, we require that m, the size of the output of such a protocol to be greater than k. In particular, even in the "minimal case", where $k = 0$ and $m = 1$, the maximum possible fidelity of any protocol is bounded by $1 - \frac{2^{n-1}}{2^n - 1}\epsilon \leq 1 - \epsilon/2$. So it is impossible to arbitrarily increase the fidelity to be close to 1, even with unlimited communication.

To prove this theorem, we need the following lemma (we define $N = 2^n$, $K = 2^k$ and $M = 2^m$).

Lemma 8.1 *Let $|\phi\rangle = (|\mathbf{0}_N \otimes \mathbf{0}_N\rangle) \otimes \Phi_k$ be a state in a bipartite system $\mathcal{H}^A_{NK} \otimes \mathcal{H}^B_{NK}$ shared between Alice and Bob. Let σ be the state Alice and Bob output after performing (arbitrary) LOCC operations. Suppose that σ is in the subspace $\mathcal{H}^A_M \otimes \mathcal{H}^B_M$. We have $\mathsf{F}(\sigma) \leq \frac{K}{M}$.*

This lemma is a a direct corollary of a result by Vidal, Jonathan, and Nielsen [91]. For the completeness of the paper, we give a somewhat simpler proof here.

For a self-adjoint matrix M, we define its *spectrum* written as $\mathcal{S}(M)$, to be a vector formed by the eigenvalues of M, and whose entries are sorted in a decreasing order. In other words, if the eigenvalues of M are $\lambda_1, \lambda_2, ..., \lambda_d$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$, then $\mathcal{S}(M) = (\lambda_1, \lambda_2, ..., \lambda_d)$.

For a mixed state ρ, if we write ρ as

$$\rho = \sum_{i=1}^{d} p_i \cdot |\phi_i\rangle\langle\phi_i|$$

where $p_1 \geq p_2 \geq \cdots \geq p_d$, and $\{|\phi_i\rangle\}$ is an orthonormal basis, then

$$\mathcal{S}(\rho) = (p_1, p_2, ..., p_d)$$

A useful fact about the spectrum of a tensor product of two matrices is the following:

Lemma 8.2 *Let A and B be square matrices such that the eigenvalues for A are $\{\lambda_1, \lambda_2, ..., \lambda_m\}$ and the eigenvalues for B are $\{\mu_1, \mu_2, ..., \mu_m\}$. Then the eigenvalues for the matrix $A \otimes B$ are $\{\lambda_i \cdot \mu_j\}_{i=1,2,...,m,\, j=1,2,..,n}$.*

Proof: If $A \cdot \vec{v} = \lambda \cdot \vec{v}$ and $B \cdot \vec{u} = \mu \cdot \vec{u}$, then $(A \otimes B) \cdot (v \otimes u) = (\lambda \cdot \mu)(v \otimes u)$ ■

A corollary the above fact is as follows.

Corollary 8.1 *Let ρ^A, ρ^B be the density matrices for quantum systems $\mathcal{H}^A$ and $\mathcal{H}^B$. Then we have*

$$\mathsf{rank}\,(\rho^A \otimes \rho^B) \geq \mathsf{rank}\,(\rho^A) \tag{8.4}$$

Proof: The rank of a matrix equals the number of non-zero eigenvalues of this matrix. Since ρ^B is a density matrix, it has trace 1, and thus it has at least one non-zero eigenvalue — assume it is μ_1. We denote the eigenvalues of ρ^A by $\lambda_1, \lambda_2, ..., \lambda_m$, then by Lemma 8.2, $\lambda_1 \cdot \mu_1, \lambda_2 \cdot \mu_1, ..., \lambda_m \cdot \mu_1$ are all eigenvalues of $\rho^A \cdot \rho^B$. Therefore, they contain at many non-zero numbers as the eigenvalues of ρ^A. ■

Proof: (of Lemma 8.1)

We consider an arbitrary protocol $\mathcal{P}$ between Alice and Bob involving only LOCC. We assume that $\mathcal{P}$ consists of *steps*, where each step could be one of the following operations [1]:

1. **Unitary Operation:**

   Alice (or Bob) applies a unitary operation to her (or his) subsystem.

2. **Measurement:**

   Alice (or Bob) performs a measurement to her (or his) subsystem.

3. **Tracing Out:**

   Alice (or Bob) discards part of her (or his) subsystem, or equivalently, traces out part of the subsystem.

4. **Classical Operation:**

   Alice (or Bob) sends a (classical) message to the other party.

We first convert this protocol $\mathcal{P}$ into another protocol $\mathcal{P}'$ in the following way: for each tracing-out operation Alice (or Bob) performs, we insert a measurement operation right before the tracing-out, and the measurement is a full measurement of the subsystem to be traced out. Notice that $\mathcal{P}'$ will have exactly the same output as $\mathcal{P}$, since the subsystem that was traced out isn't part of the output. However, $\mathcal{P}'$ has the property that for each subsystem traced out in the protocol, that subsystem is disentangled from the rest, since it is already completely measured.

Now we analyze the new protocol $\mathcal{P}'$. We denote the partial density matrix of Alice for the state $|\phi\rangle$ by ρ^A:

$$\rho^A = \mathrm{Tr}_B(|\phi\rangle\langle\phi|) \tag{8.5}$$

[1] We assume that Alice have enough ancillary qubit at the beginning of the protocol and not more new ancillary qubits need to be introduced during the protocol.

Since we know $|\phi\rangle$ precisely, we can compute ρ^A precisely, and in particular, its spectrum. It is easy to verify that the spectrum of ρ^A is

$$\mathcal{S}(\rho^A) = (\underbrace{1/K, 1/K, ...1/K}_{K}, \underbrace{0, 0, ..., 0}_{(N-1)K})$$

So the rank of ρ^A (which is also the Schmidt Number of $|\phi\rangle$) is K.

We focus on how ρ^A changes with the local operations Alice performs (apparently it doesn't change with Bob's local operations): we shall prove that the rank of ρ^A never increases. There are 3 types of operations Alice can perform: unitary operations, local measurements, and tracing out a subsystem, we analyze them one by one:

- **Unitary Operations**

  This operation changes a mixed state ρ^A to $U\rho^A U^\dagger$, where U is a unitary operation. Obviously the rank doesn't change.

- **Local Measurements**

  Suppose measurement operator is $\{M_m\}$ satisfying $\sum_m M_m^\dagger M_m = I$, and the measurement yields result m. Then Alice ends in state

  $$\rho_m = \frac{M_m \rho^A M_m^\dagger}{\mathrm{Tr}(M_m^\dagger M_m \rho^A)}$$

  Again, we have $\mathsf{rank}\,(\rho_m) \leq \mathsf{rank}\,(\rho^A)$.

- **Tracing Out a Subsystem**

  We write $\mathcal{H}^A = \mathcal{H}^{A_0} \otimes \mathcal{H}^{A_1}$, and we suppose that the subsystem $\mathcal{H}^{A_1}$ is traced out. We write the partial density matrix for $\mathcal{H}^{A_0}$ as ρ^{A_0}, and we have $\rho^{A_0} = \mathrm{Tr}_{A_1}(\rho^A)$.

  We know that in protocol $\mathcal{P}'$, the subsystem $\mathcal{H}^{A_0}$ is disentangled from the subsystem $\mathcal{H}^{A_1}$. Thus we have

  $$\rho^A = \rho^{A_0} \otimes \rho^{A_1}$$

  for some density matrix ρ^{A_1}. and by Corollary 8.1, we have $\mathsf{rank}\,(\rho^{A_0}) \leq \mathsf{rank}\,(\rho^A)$.

So, as Alice and Bob perform local operations, the rank of the partial density matrix for Alice never increases. This fact remains true even if Alice and Bob perform classical communications (this just means that Alice has the ability to perform different local operations according to Bob's measurement result, but no local operation Alice performs can increase the rank).

We denote the density matrix for the final state after the protocol $\mathcal{P}$ to be ρ_E, and we define $\rho_E^A = \mathrm{Tr}_B(\rho_E)$ to be the partial density matrix for Alice. Then we have $\mathsf{rank}\,(\rho_E^A) \leq K$. Notice ρ_E^A should be an $M \times M$ matrix since Alice and Bob are supposed to arrive at a state in $\mathcal{H}_M^A \otimes \mathcal{H}_M^B$. We use ρ_0^A to denote the partial density matrix for Alice if we trace out the system $\mathcal{H}_M^B$ from the target state Ψ_M. It is easy to verify that $\rho_0^A = \frac{1}{M} I$, where I is the identity matrix.

By monotonicity of fidelity, we have

$$F(\rho_E, |\Psi_M\rangle\langle\Psi_M|) \leq F(\rho_E^A, \rho_0^A)$$

However, we have

$$\begin{aligned} F(\rho_E^A, \rho_0^A) &= \mathrm{Tr}\sqrt{(\rho_E^A)^{1/2}\rho_0^A(\rho_E^A)^{1/2}} \\ &= \sqrt{\frac{1}{M}} \cdot \mathrm{Tr}\sqrt{\rho_E^A} \end{aligned}$$

We write the spectrum of ρ_E^A as

$$\mathcal{S}(\rho_E^A) = (\lambda_1, \lambda_2, ..., \lambda_M)$$

and we know that $\lambda_{K+1} = \lambda_{K+2} = \cdots = \lambda_M = 0$ since $\mathsf{rank}\,(\rho_E^A) \leq K$. Therefore, we have

$$\mathrm{Tr}\sqrt{\rho_E^A} \;=\; \sum_{l=1}^{M} \sqrt{\lambda_l} = \sum_{l=1}^{K} \sqrt{\lambda_l} \leq \sqrt{K} \cdot \left(\sum_{l=1}^{K} \lambda_l \right) = \sqrt{K}$$

and thus

$$F(\rho_E^A, \rho_0^A) = \sqrt{\frac{1}{M}} \cdot \mathrm{Tr}\sqrt{\rho_E^A} \leq \sqrt{\frac{K}{M}}$$

Therefore we have

$$F(\rho_E) = F(\rho_E, |\Psi_M\rangle\langle\Psi_M|) \leq F(\rho_E^A, \rho_0^A) \leq \frac{K}{M}$$

∎

Proof: (of Theorem 8.2)

We prove the theorem by demonstrating a particular mixed state ρ of fidelity $1-\epsilon$, such that no LOCC protocol can increase its fidelity to more than $1-\frac{M-K}{M}\frac{N}{N-1}\epsilon$.

Let $\epsilon' = \frac{N}{N-1}\epsilon$. We define the state ρ to be

$$\rho = (1-\epsilon')\cdot\Phi_n + \epsilon'\cdot|\mathbf{0}_N\otimes\mathbf{0}_N\rangle\langle\mathbf{0}_N\otimes\mathbf{0}_N| \tag{8.6}$$

In other words, ρ is the maximally entangled state Φ_n with probability $(1-\epsilon')$ and the completely disentangled state $\mathbf{0}_N\otimes\mathbf{0}_N$ with probability ϵ'.

It is easy to verify that $\mathsf{F}(\rho) = 1-\epsilon$, since $\langle\Phi_n\,|\,\mathbf{0}_N\otimes\mathbf{0}_N\rangle = \frac{1}{\sqrt{N}}$ and, therefore,

$$\mathsf{F}(\rho) = (1-\epsilon')\mathsf{F}(\Phi_n) + \epsilon'\mathsf{F}(|\mathbf{0}_N\otimes\mathbf{0}_N\rangle\langle\mathbf{0}_N\otimes\mathbf{0}_N|) = (1-\epsilon') + \frac{1}{N}\epsilon' = 1-(1-\frac{1}{N})\epsilon' = 1-\epsilon. \tag{8.7}$$

For an arbitrary LOCC protocol $\mathcal{P}$, we define $f_1 = \mathsf{F}(\mathcal{P}(\Phi_n))$ and $f_2 = \mathsf{F}(\mathcal{P}(|\mathbf{0}_N\otimes\mathbf{0}_N\rangle\langle\mathbf{0}_N\otimes\mathbf{0}_N|))$

Then we have $f_1 \le 1$ and by Lemma 8.1, $f_2 \le K/M$.

By the linearity of fidelity of quantum operations, we know that

$$\mathsf{F}(\mathcal{P}(\rho)) = (1-\epsilon')f_1 + \epsilon' f_2 \le 1-\frac{M-K}{M}\epsilon' = 1-\frac{M-K}{M}\frac{N}{N-1}\epsilon. \tag{8.8}$$

Now, we prove that this lower bound is tight by demonstrating an (n,m)-protocol that saturates the bound in (8.8). The protocol is called the “random permutation protocol”. In the simplest version, it doesn't use any auxiliary input (i.e. $k=0$). Again, we define $N=2^n$, $M=2^m$, and $K=2^k$.

Construction 8.1 (Random Permutation Protocol)

1. *Alice generates a uniformly random permutation π on $\{0,1\}^n$ using classical randomness and transmits the permutation to Bob.*

2. *Alice applies permutation π on $\mathcal{H}_N^A$, mapping $|\,i\rangle$ to $|\,\pi(i)\rangle$, Bob does the same on $\mathcal{H}_N^B$.*

3. *Alice and Bob decompose $\mathcal{H}_N$ as $\mathcal{H}_M\otimes\mathcal{H}_L$, $L=N/M$ and measure the $\mathcal{H}_L$ part.*

4. *Alice sends the result of her measurement to Bob, Bob sends his result to Alice.*

5. *They compare the results. If the results are the same, they output the state that they have in $\mathcal{H}_M^A \otimes \mathcal{H}_M^B$. If the results are different, they output $|\, Z_M\rangle \otimes |\, Z_M\rangle$.*

We compute the fidelity of the random permutation protocol. We need one more notation. For a symmetric, bipartite system $\mathcal{H} = \mathcal{H}_N^A \otimes \mathcal{H}_N^B$, we denote by $\mathcal{H}^{\mathcal{D}}$ the N-dimensional subspace spanned by

$$\left\{ \sum_{i=0}^{N-1} \alpha_i \cdot |\, i\rangle^A |\, i\rangle^B \right\}$$

and we call it the *diagonal subspace* of $\mathcal{H}_N^A \otimes \mathcal{H}_N^B$. A mixed state ρ is in the diagonal subspace, if there exists a decomposition of ρ:

$$\rho = \sum_i p_i \cdot |\phi_i\rangle\langle\phi_i|$$

such that all pure states $|\, \phi_i\rangle$ are in the diagonal subspace.

We start with the case when the state of Alice and Bob is in the diagonal subspace.

Lemma 8.3 *If the input state to the random permutation protocol is in the diagonal subspace, then the fidelity of the output is $\frac{M-1}{M}\frac{N}{N-1}\epsilon$.*

Proof: Without loss of generality, we assume that the input state is pure. Let $|\, \phi\rangle = \sum_{i=1}^N \alpha_i |\, i\rangle^A |\, i\rangle^B$ be the starting state. For a permutation π, let U_π be the unitary transformation defined by $U_\pi \left(|\, i\rangle^A \otimes |\, j\rangle^B\right) = |\, \pi(i)\rangle^A |\, \pi(j)\rangle^B$. Then, if Alice and Bob use a permutation π, the resulting state is

$$|\, \phi_\pi\rangle = U_\pi |\, \phi\rangle = \sum_{i=1}^N \alpha_i |\, \pi(i)\rangle^A |\, \pi(i)\rangle^B = \sum_{i=1}^N \alpha_{\pi^{-1}(i)} |\, i\rangle^A |\, i\rangle^B.$$

There are $N!$ permutations π on a set of N elements. Therefore, each of them gets applied with probability $1/N!$. This means that the final state is a mixed state of $|\, \phi_\pi\rangle$ with probabilities $1/N!$ each. We calculate the density matrix ρ of this state. It is equal to

$$\sum_\pi \frac{1}{N!} |\, \phi_\pi\rangle\langle\phi_\pi\,| = \sum_\pi \frac{1}{N!} \begin{pmatrix} \alpha_{\pi^{-1}(1)}\alpha^*_{\pi^{-1}(1)} & \alpha_{\pi^{-1}(1)}\alpha^*_{\pi^{-1}(2)} & \dots & \alpha_{\pi^{-1}(1)}\alpha^*_{\pi^{-1}(N)} \\ \alpha_{\pi^{-1}(2)}\alpha^*_{\pi^{-1}(1)} & \alpha_{\pi^{-1}(2)}\alpha^*_{\pi^{-1}(2)} & \dots & \alpha_{\pi^{-1}(2)}\alpha^*_{\pi^{-1}(N)} \\ \dots & \dots & \dots & \dots \\ \alpha_{\pi^{-1}(N)}\alpha^*_{\pi^{-1}(1)} & \alpha_{\pi^{-1}(N)}\alpha^*_{\pi^{-1}(2)} & \dots & \alpha_{\pi^{-1}(N)}\alpha^*_{\pi^{-1}(N)} \end{pmatrix}.$$

We claim that all diagonal entries ρ_{ii} are equal to $1/N$ and all off-diagonal entries ρ_{ij}, $i \neq j$ are equal to some value a which is real. This follows from the symmetries created by summing over all permutations.

Consider a diagonal entry ρ_{ii}. For each $j \in \{1, \dots, N\}$, there are $(N-1)!$ permutations that map j to i. Therefore,

$$\rho_{ii} = \sum_{j=1}^{N} (N-1)! \frac{1}{N!} \alpha_j \alpha_j^* = \frac{1}{N} \sum_{j=1}^{N} |\alpha_j|^2.$$

$\sum_{j=1}^{N} |\alpha_j|^2$ is the same as $\|\phi\|^2$ which is equal to 1. Therefore, $\rho_{ii} = \frac{1}{N}$.

Next, consider an off-diagonal entry ρ_{ij}. For each k, l, $k \neq l$, there are $(N-2)!$ permutations that map k to i and l to j. Therefore,

$$\rho_{ij} = \sum_{k=1}^{N} \sum_{l=1, l\neq k}^{N} (N-2)! \frac{1}{N!} \alpha_k \alpha_l^* = \sum_{k=1}^{N} \sum_{l=1, l\neq k}^{N} \frac{1}{N(N-1)} \alpha_k \alpha_l^*.$$

This immediately implies that ρ_{ij} is the same for all $i \neq j$. Also, notice that $(\alpha_k \alpha_l^*)^* = \alpha_k^* \alpha_l$. Therefore, $\alpha_k \alpha_l^* + \alpha_l \alpha_k^*$ is real and ρ_{ij} (which is a sum of terms of this form) is real as well. Let $a = \rho_{ij}$. We have shown that

$$\rho = \begin{pmatrix} \frac{1}{N} & a & \dots & a \\ a & \frac{1}{N} & \dots & a \\ \dots & \dots & \dots & \dots \\ a & a & \dots & \frac{1}{N} \end{pmatrix}.$$

Notice that the density matrix ρ can be also obtained from a mixed state that is Φ_n with probability Na and each of basis states $|\, i\rangle^A |\, i\rangle^B$ with probability $\frac{1}{N} - a$.

We now consider applying steps 3-5 to those states. Measuring $\mathcal{H}_L^A \otimes \mathcal{H}_L^B$ for Φ_n always gives the same results and leaves Alice and Bob with the state Φ_m in $\mathcal{H}_M^A \otimes \mathcal{H}_M^B$. The fidelity of this state with Φ_m is, of course, 1. Measuring $\mathcal{H}_L^A$ and $\mathcal{H}_L^B$ for $|\, i\rangle^A |\, i\rangle^B$ also gives the same results and leaves Alice and Bob with some basis state $|\, i'\rangle^A |\, i'\rangle^B$ in the diagonal subspace of $\mathcal{H}_M^A \otimes \mathcal{H}_M^B$. The fidelity of this state and $|\, \Psi_M\rangle$ is $\frac{1}{M}$. By the linearity of fidelity, if we apply those steps to the state ρ, we get that the fidelity is

$$Na + N\left(\frac{1}{N} - a\right)\frac{1}{M} = \frac{1}{M} + Na\left(1 - \frac{1}{M}\right). \tag{8.9}$$

We now lower-bound a. Again by the linearity of fidelity, we have $\mathsf{F}(\rho) = \frac{1}{N!}\sum_\pi \mathsf{F}(|\,\phi_\pi\rangle)$. Since permuting the basis states $|\,i\rangle^A|\,i\rangle^B$ preserves the maximally entangled state $\Phi_n = \frac{1}{\sqrt{N}}\sum_{i=1}^N |\,i\rangle^A|\,i\rangle^B$, the fidelity of any $|\,\phi_\pi\rangle$ is the same as the fidelity of $|\,\phi\rangle$. Therefore, $\mathsf{F}(\rho) = \mathsf{F}(|\,\phi\rangle) \geq 1-\epsilon$. By applying the definition of fidelity,

$$\begin{aligned}
\mathsf{F}(\rho) &= \begin{pmatrix} \frac{1}{\sqrt{N}} \\ \frac{1}{\sqrt{N}} \\ \cdots \\ \frac{1}{\sqrt{N}} \end{pmatrix} \begin{pmatrix} \frac{1}{N} & a & \dots & a \\ a & \frac{1}{N} & \dots & a \\ \dots & \dots & \dots & \dots \\ a & a & \dots & \frac{1}{N} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{N}} & \frac{1}{\sqrt{N}} & \cdots & \frac{1}{\sqrt{N}} \end{pmatrix} \\
&= N\frac{1}{N^2} + N(N-1)\frac{1}{N}a \\
&= \frac{1}{N} + (N-1)a.
\end{aligned}$$

Since $F(\rho) \geq 1-\epsilon$, it must be the case that $a \geq \frac{1}{N} - \frac{\epsilon}{N-1}$. By substituting that into (8.9), the fidelity of the final state with Φ_m is at least

$$\frac{1}{M} + N\left(\frac{1}{N} - \frac{\epsilon}{N-1}\right)\left(1-\frac{1}{M}\right) = 1 - \frac{N}{N-1}\left(1-\frac{1}{M}\right)\epsilon.$$

■

It remains to show that the protocol also succeeds for states not in the diagonal subspace. Let $|\,\phi\rangle$ be a state such that $\mathsf{F}(|\,\phi\rangle) \geq 1-\epsilon$. We decompose

$$|\,\phi\rangle = \sqrt{1-\delta}|\,\phi_1\rangle + \sqrt{\delta}|\,\phi_2\rangle,$$

with $|\,\phi_1\rangle \in \mathcal{H}^D$ and $|\,\phi_2\rangle \in (\mathcal{H}^D)^\perp$. Let $\mathsf{F}(|\,\phi_1\rangle) = 1-\delta'$. Since Φ_n is in $\mathcal{H}^D$ and $|\,\phi_2\rangle$ is orthogonal to $\mathcal{H}^D$, we have $\mathsf{F}(|\,\phi_2\rangle) = 0$ and $\mathsf{F}(|\,\phi\rangle) = (1-\delta)(1-\delta')$. Notice that $(1-\delta)(1-\delta') \geq 1-\epsilon$ because $\mathsf{F}(|\,\psi\rangle) \geq 1-\epsilon$.

Applying U_π maps $|\,\phi\rangle$ to $|\,\phi_\pi\rangle = \sqrt{1-\delta}|\,\phi_{\pi,1}\rangle + \sqrt{\delta}|\,\phi_{\pi,2}\rangle$ where $|\,\phi_{\pi,1}\rangle = U_\pi|\,\phi_1\rangle$, $|\,\phi_{\pi,2}\rangle = U_\pi|\,\phi_2\rangle$. Since U_π preserves the diagonal subspace, $|\,\phi_{\pi,1}\rangle \in \mathcal{H}^{\mathcal{D}}$ and $|\,\phi_{\pi,2}\rangle \in (\mathcal{H}^{\mathcal{D}})^\perp$. Measuring $\mathcal{H}_L^A$ and $\mathcal{H}_L^B$ for a state in $\mathcal{H}^{\mathcal{D}}$ always gives the same results and produces a state in the diagonal subspace of $\mathcal{H}_M^A \otimes \mathcal{H}_M^B$. Measuring $\mathcal{H}_L^A$ and $\mathcal{H}_L^B$ for a state in $(\mathcal{H}^{\mathcal{D}})^\perp$ either gives the different results for Alice and Bob or gives the same results but produces a state orthogonal to the diagonal subspace of $\mathcal{H}_M^A \otimes \mathcal{H}_M^B$.

The fidelity of the final state consists of two parts: the fidelity of the final state if Alice's and Bob's measurements of $\mathcal{H}_L$ give the same answer and the fidelity if measurements give the different answer. The first part is just $(1-\delta)$ times the fidelity of the final state if the starting state was $|\,\phi_1\rangle$ (instead of $|\,\phi\rangle$). Since $|\,\phi_1\rangle$ is in the diagonal subspace, Lemma 8.3 implies that the final state of the protocol $|\,\phi_1\rangle$ has the fidelity at least $1-D\delta'$ where $D=\frac{M-1}{M}\frac{N}{N-1}$. Therefore, the first part is at least

$$(1-\delta)(1-D\delta')=(1-\delta)(1-\delta')+(1-D)\delta'(1-\delta) \tag{8.10}$$

The second part is the probability of measurements giving different answers times the fidelity of the state $|\,0\rangle\otimes|\,0\rangle$ which Alice and Bob output in this case. The fidelity of this state is $\frac{1}{M}$ and the probability of this case is given by the following lemma.

Lemma 8.4 *The probability that Alice's and Bob's measurements give different answers is* $\frac{N-M}{N-1}\delta$.

Proof: First, we look at the state $|\,\phi_2\rangle$. Since this state is in $(\mathcal{H}^{\mathcal{D}})^{\perp}$, it is of the form

$$|\,\phi_2\rangle=\sum_{i,j=1,i\neq j}^{N}\alpha_{i,j}|\,i\rangle^A|\,j\rangle^B.$$

Applying U_π maps it to

$$|\,\phi_{\pi,2}\rangle=\sum_{i\neq j}\alpha_{i,j}|\,\pi(i)\rangle^A|\,\pi(j)\rangle^B=\sum_{i\neq j}\alpha_{\pi^{-1}(i),\pi^{-1}(j)}|\,i\rangle^A|\,j\rangle^B.$$

The probability of Alice and Bob getting different results is equal to the sum of $|\alpha_{\pi^{-1}(i),\pi^{-1}(j)}|^2$ over all basis $|\,i\rangle^A$, $|\,j\rangle^B$ that differ in the $\mathcal{H}_L$ part. If this sum is averaged over all permutations π, it becomes the same for all i,j, $i\neq j$. Therefore, the probability of Alice and Bob getting different results is just the fraction of pairs (i,j) that differ in the $\mathcal{H}_L$ part. It is $\frac{N-M}{N-1}$ because for each i, there are $(N-1)$ $j\in\{1,\ldots,N\}$, $j\neq i$ and $M-1$ of them differ only in the $\mathcal{H}_K$ but the remaining $N-M$ differ in the $\mathcal{H}_L$ part.

If the starting state is $|\,\phi\rangle$, the probability of Alice and Bob getting different results is δ times the probability for $|\,\phi_2\rangle$ because $|\,\phi\rangle=\sqrt{1-\delta}|\,\phi_1\rangle+\sqrt{\delta}|\,\phi_2\rangle$ and the measurements always give the same answer on $|\,\phi_1\rangle$. ■

Therefore, the second part of the fidelity is $\frac{1}{M}\frac{N-M}{N-1}\delta$. Notice that $1 - D = 1 - \frac{(M-1)N}{M(N-1)} = \frac{(M-1)N - M(N-1)}{M(N-1)} = \frac{N-M}{M(N-1)}$. Thus, the second part is $(1-D)\delta$ and the overall fidelity is at least

$$(1-\delta)(1-\delta') + (1-D)(1-\delta)\delta' + (1-D)\delta = 1 - D(\delta(1-\delta') + \delta').$$

Since $(1-\delta)(1-\delta') \geq 1-\epsilon$, $\delta(1-\delta') + \delta' \leq \epsilon$. Therefore, the overall fidelity is at least $1 - D\epsilon$. This completes the proof of the second part of Theorem 8.2 for $K = 1$.

For $K > 1$, we can just produce an entangled state of dimension $M' = M/K$ without the use of $|\,\Psi_K\rangle$ by the protocol above and then output this state and the original $|\,\Psi_K\rangle$. This achieves the fidelity of at least $1 - D\epsilon$ for $D = \frac{M'-1}{M'}\frac{N}{N-1} = \frac{M/K-1}{M/K}\frac{N}{N-1} = \frac{M-K}{M}\frac{N}{N-1}$, proving that the bound of Theorem 8.2 is tight for $k > 0$.

■

Interestingly, we can show that communication almost does not help for entanglement distillation over the fidelity model. The next theorem states that the random permutation protocol can be modified into an non-interactive one with only with a small loss of fidelity.

Theorem 8.3 (Non-interactive Absolute Protocols for the Fidelity Model) *There exists a non-interactive, randomized public-coin entanglement distillation $(n,1)$-protocol of fidelity $1 - \frac{3}{4}\frac{2^n - \frac{2}{3}}{2^n - 1}\epsilon$ over a fidelity noise model $\mathcal{F}_{n,1-\epsilon}$. Furthermore, this it is almost the best possible, in that the fidelity of any non-interactive, randomized public-coin entanglement distillation $(n,1)$-protocol over the model $\mathcal{F}_{n,1-\epsilon}$ is $1 - \frac{3}{4}\frac{2^{2n}}{2^{2n}-1}\epsilon$, for $\epsilon \leq \frac{2^{2n}-1}{2^{2n+1}}$.*

It is interesting to compare this result to a special case of Theorem 8.2, where $k = 0$ and $m = 1$. We see that with communication, the maximum fidelity of a protocol is about $1 - \epsilon/2$, and there exists a protocol that matches this bound exactly. Without communication, the maximum fidelity is about $1 - 3\epsilon/4$, and it is tight, too. Therefore, communication does help in this case, but not much.

Proof: (of Theorem 8.3) We first show that the random permutation protocol in Construction 8.1 cane modified into a non-interactive one.

Construction 8.2 (Non-interactive Random Permutation Protocol)

1. *Using the shared random string, Alice and Bob generate a uniformly random permutation $\pi \in S_{2^n}$ and $x_1 \in \{-1,1\}$, $x_2 \in \{-1,1\}$, ..., $x_{2^n} \in \{-1,1\}$.*

2. *Alice and Bob apply the transformation U mapping $U|\,i\rangle = (-1)^{x_i}|\,\pi(i)\rangle$ to their qubits.*

3. *They each output the first qubit 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.

Lemma 8.5 *Let ρ be the mixed state obtained after the first two steps. Then,*

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

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

$$p_0 + \frac{1}{2^n}p_1 \geq 1 - \epsilon \tag{8.11}$$

because each of states $U \otimes U|\,\psi\rangle$ has the same fidelity as $|\,\psi\rangle$ and fidelity is convex. We can rewrite (8.11) as $\frac{2^n-1}{2^n}p_1 + p_2 + p_3 \leq \epsilon$.

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 $1-\delta$,

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

Next, we prove the second part of the theorem, that this is almost the best a non-interactive protocol can do.

Let ρ be the mixture of Φ_n with probability $1-\frac{2^{2n}}{2^{2n}-1}\epsilon$ and the completely mixed state in $2^n\times 2^n$ dimensions with probability $\frac{2^{2n}}{2^{2n}-1}\epsilon$. Since the perfect state has fidelity 1 and the completely mixed state has fidelity $\frac{1}{2^{2n}}$, this state has fidelity $1-\epsilon$.

W.l.o.g., a non-interactive 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 $\lambda_2=\frac{1}{2}-\delta_A$, $\delta_A\ge 0$. 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 $\delta=\max(\delta_A,\delta_B)$.

Lemma 8.6 *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,

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

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

■

Lemma 8.7 *If the starting state is the completely mixed state in 2^{2n} dimensions, the fidelity of the final state is at most $\frac{1}{4}+\epsilon$.*

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

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

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

■

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

$$(1-\frac{2^{2n}}{2^{2n}-1}\epsilon)(1-\delta^2) + \frac{2^{2n}}{2^{2n}-1}\epsilon(\frac{1}{4}+\delta^2) \le 1 - \frac{3}{4}\frac{2^{2n}}{2^{2n}-1}\epsilon. \tag{8.12}$$

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 Eq (8.12) applies for any particular r, Therefore, it also applies on the average over all r. ■

8.3.2 Part II: Purity Testing Protocols and Conditional Protocols

Theorem 8.2 spells a negative result for absolute protocols over the fidelity noise model by demonstrating a state ρ such that no LOCC protocol can increase its fidelity significantly. However, the situation is vastly different for the case of conditional protocols. We shall prove that very efficient entanglement distillation protocols exist that can increase the conditional fidelity to as close to 1 as possible. As we shall see, one construction of such protocols is closely related to the notion of purity testing protocols.

Theorem 8.4 (Conditional Protocols for the Fidelity Model) *For all integers $n > s$, there exists an conditional, randomized, $(2ns + s)$-bit one-way, $(n, n - s)$ protocol over the fidelity noise model $\mathcal{F}_{n,1-\epsilon}$ with success probability at least $1 - \epsilon$ and conditional fidelity $1 - \frac{2^{-s}}{1+2^{-s}-\epsilon}$.*

We prove this theorem by first demonstrating a closely related notion, namely the purity testing protocols, and then showing how these protocols are in fact entanglement distillation protocols, followed by an explicit construction.

Purity Testing Protocols

A purity testing protocol is an LOCC protocol where the input is joint state shared by Alice and Bob which they think might be the EPR state Φ_n. Alice and Bob want to test if their shared state is indeed Φ_n, while sacrificing the least number of EPR pairs. The concept of purity testing protocols were studied implicitly by Lo and Chau [57] and Shor and Preskill [84] in the context of proving the security of the BB84 quantum key distribution protocol [16], and later explicitly by Barnum, Crépeau, Gottesman, Smith, and Tapp [23].

Definition 8.3 (Purity Testing Protocol, adapted from [23]) *A* purity testing protocol *with parameters (n, m, α) is a LOCC super-operator $\mathcal{T}_{n,m,\alpha}$ which maps $2n$ qubits (half held by Alice and half held by Bob) to $2m + 1$ qubits (m of which are held by Bob) and satisfies the following two conditions:*

- Completeness: $\mathcal{T}(\Phi_n) = \Phi_m \otimes |\,\mathsf{SUCC}\rangle$

- Soundness: *Let P be the projection on the subspace spanned by $\Phi_m \otimes |\,\mathsf{SUCC}\rangle$ and $|\,\psi\rangle \otimes |\,\mathsf{FAIL}\rangle$ for all $|\,\psi\rangle$. Then $\mathcal{T}$ is* sound with error α *if for all ρ,*

$$\mathrm{Tr}\left(P\mathcal{T}(\rho)\right) \geq 1 - \alpha.$$

It's convenient to think of purity testing as approximating the measurement given by the projector onto Φ_m and its orthogonal complement.

Purity Testing Protocols are Entanglement Distillation Protocols

We prove that every purity testing protocol is in fact an entanglement distillation protocol.

Lemma 8.8 *Every purity testing protocol $\mathcal{T}_{n,m,\alpha}$ corresponds to an conditional entanglement distillation (n, m)-protocol over the fidelity noise model $\mathcal{F}_{n,1-\epsilon}$ with success probability at least $1 - \epsilon$ and conditional fidelity at least $1 - \frac{\alpha}{1-\epsilon+\alpha}$.*

Proof: We show that the purity testing protocol $\mathcal{T}_{n,m,\alpha}$ is in fact an entanglement distillation protocol with the slightest modification. Alice and Bob simply run the purity-testing protocol, with Alice outputting FAIL when the purity testing rejects the input. Now we estimate the success probability and the conditional fidelity of this protocol.

Suppose at the end of the protocol Alice and Bob trace out everything except the $2m$ output qubits and the qubit indicating accept/reject. Consider the three projectors:

$$\begin{aligned}
P_1 &= \Phi_m \otimes |\,\mathsf{SUCC}\rangle\langle\mathsf{SUCC}\,| \\
P_2 &= (I_{\mathcal{M}} - \Phi_m) \otimes |\,\mathsf{SUCC}\rangle\langle\mathsf{SUCC}\,| \\
P_3 &= I_{\mathcal{M}} \otimes |\,\mathsf{FAIL}\rangle\langle\mathsf{FAIL}\,|
\end{aligned}$$

And define $\gamma_i = \mathrm{Tr}[P_i \rho']$ where ρ' is the final state.

If the input to the system had fidelity $1-\epsilon$, then the completeness of the purity-testing protocol implies that the fidelity of the output to $\Phi_m|\,\mathsf{SUCC}\rangle$ must be $1 - \epsilon$, and so $\gamma_1 \geq 1 - \epsilon$. Therefore the success probability is at least $1-\epsilon$. If the purity-testing protocol has soundness error α, then the soundness condition implies $\gamma_2 \le \alpha$.

Now the output fidelity conditioned on acceptance is

$$\frac{\gamma_1}{\gamma_1+\gamma_2} = 1 - \frac{\gamma_2}{\gamma_1+\gamma_2} \ge 1 - \frac{\alpha}{1-\epsilon+\alpha}.$$

This finishes the proof. ∎

Constructing Purity Testing Protocols

Purity testing protocols are in fact easy to construct and are very efficient. A particularly simple purity-testing protocol consists of picking a random stabilizer code of dimension 2^{n-s}, having Alice and Bob both measure the syndrome of the code, and then extracting the encoded state if both measurement results are the same.

Lemma 8.9 (Random hashing) *For all integers $n > s$, there exist purity testing protocols of parameters (n, m, α) such that such that $m = n - s$, $\alpha \le 2^{-s}$ and which use $ns + s + 1$ bits of classical communication.*

This lemma actually follows from the observation that the set of *all* stabilizer codes [36] of dimension 2^{n-s} is a purity-testing code family with error $\alpha \le 2^{-s}$. However, we give a direct proof with an explicit protocol description below.

Without loss of generality, we describe the protocol in terms of purifying the state $|\Psi^-\rangle$[2]. We describe a protocol with $m = n-1$ and error $\alpha = \frac{1}{2}$. Repeating the protocol s times yields $m = n-s$ and $\alpha = 2^{-s}$.

Construction 8.3 (Simple Random Hashing Protocol)

1. *Alice picks $2n$ random bits $x_1, ..., x_n, z_1, ..., z_n$ such that not all the bits are 0.*

2. *Alice will measure the operator given by*
   $X^{x_1}Z^{z_1} \otimes \cdots \otimes X^{x_n}Z^{z_n}$. *To do this Alice:*

[2] For example, Bob can perform a "phase-shift" (Z) followed by a "bit-flip" (X) to every qubit he possesses. This will transform $|\Phi^+\rangle$ to $|\Psi^-\rangle$.

(a) *Considers only qubits where* $(x_i, z_i) \neq (0,0)$. *Say there* ℓ *qubits left.*

(b) *On qubit* j*, applies either*

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \text{ if } (x_j, z_j) = (0,1),$$

$$B = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \text{ if } (x_j, z_j) = (1,1),$$

the identity if $(x_j, z_j) = (1,0)$.

(c) *Applies* CNOT *from each of the first* $\ell - 1$ *qubits onto the last.*

(d) *Measures the last in the computational basis.*

(e) *Applies the inverse transformation to the remaining qubits.*

3. *Alice sends* $x_1, ..., x_n, z_1, ..., z_n$ *and her measurement result to Bob.*

4. *Bob performs the same measurement and sends back the result.*

5. *Alice and Bob accept if the two results are different and reject otherwise.*

Proof: (of Lemma 8.9)

It is sufficient to consider the performance of the protocol on states of the form $X^{\vec{a}} Z^{\vec{b}} |\Psi^-\rangle^{\otimes n}$, where $X^{\vec{a}}$ denotes $X^{a_1} \otimes \cdots \otimes X^{a_n}$ when $\vec{a} = (a_1, ..., a_n) \in \{0,1\}^n$. Without the loss of generality, we assume all the error operators are applied to Alice's share of the EPR pairs.

The reduction to these Bell states is via a "quantum-to-classical reduction", as used in [57] for key distribution. The reduction works because ultimately, the accept/reject decision is diagonal in the Bell basis, and moreover if the input to the protocol can be described as $X^{\vec{a}} Z^{\vec{b}} |\Psi^-\rangle^{\otimes n}$, the the output can be written $X^{\vec{a}'} Z^{\vec{b}'} |\Psi^-\rangle^{\otimes m}$.

The idea is that measuring the operator $X^{x_1} Z^{z_1} \otimes \cdots \otimes X^{x_n} Z^{z_n}$ on both Alice and Bob's shares and comparing the results is equivalent to measuring the bit $\vec{a} \odot \vec{x} + \vec{b} \odot \vec{z}$, i.e. a random linear function of the vector $(\vec{x}, \vec{z})$. To see this, first observe that $HX^a Z^b = (-1)^{ab} X^b Z^b H$ and $BX^a Z^b = i^b X^{a+b} Z^b B$. Moreover, both $B \otimes B$ and $H \otimes H$ have $|\Psi^-\rangle$ as an eigenvector. Thus, in each position we will end up with a state proportional to $X^{x_j a_j + z_j b_j} Z^c |\Psi^-\rangle$ after Alice and Bob have applied their transformations and before they measure, where c is a bit. Measuring both halves in the computational basis and comparing results allows one to compute $x_j a_j + z_j b_j$. Similarly, the protocol computes $\vec{x} \odot \vec{a} + \vec{b} \odot \vec{z}$.

A random linear function will detect a non-zero vector with probability $\frac{1}{2}$. Thus, the overall error probability of the one-step protocol is bounded by $\frac{1}{2}$. Repeating the protocol s times lowers this error to 2^{-s}. ■

Proof: **(of Theorem 8.4)** It directly follows Lemma 8.8 and Lemma 8.9. ■

In fact, a closer look at the Construction 8.3 reveals that of the $(2n+1)$ bits of communication in this protocol, $2n$ of them are used for selecting a random string, which can be spared if Alice and Bob initially share a random string. This observation leads to the following corollary to Theorem 8.4.

Corollary 8.2 *For all integers $n > s$, there exists an conditional, randomized public-coin, s-bit one-way, $(n, n-s)$ protocol over the fidelity noise model $\mathcal{F}_{n,1-\epsilon}$ with success probability at least $1-\epsilon$ and conditional fidelity at least $1 - \frac{2^{-s}}{1+2^{-s}-\epsilon}$.* ■

Here, we see an exponential trade-off between the conditional fidelity and the amount of communication: each additional bit communicated will reduce the gap between the conditional fidelity and 1 by almost half. This contrasts sharply with the relation between fidelity and communication, where communication does help a little, but by only at most a constant factor.

8.3.3 Part III: The Communication Complexity

We study the communication complexity of entanglement distillation protocols over the fidelity noise model. We prove a lower bound that matches the result from Corollary 8.2 up to an additive constant. This effectively shows that the construction of Corollary 8.2 is optimal.

Definition 8.4 (Ideal Success Probability) *The* ideal success probability *of a conditional quantum entanglement distillation (n, m)-protocol is the probability that it succeeds over the input Φ_n. A protocol is* ideal *if its ideal success probability is 1.*

Theorem 8.5 (Communication Complexity of Protocols for the Fidelity Model) *The conditional fidelity of any randomized public-coin s-bit (n, m)-protocol of ideal success probability p is at most $1 - \frac{\epsilon p}{2^{s+1}}$ over a fidelity noise model $\mathcal{F}_{n,1-\epsilon}$*

An immediate corollary of this theorem is that the conditional fidelity of an s-bit ideal protocol is at most $1-\epsilon/2^{s+1}$. Therefore, to achieve a fidelity or $1-\delta$ on the output, $\log(1/\delta)+\log(\epsilon\cdot p)-1$ bits of classical communication is needed. On the other hand, Corollary 8.2 yields a communication complexity of $\log(1/\delta)+\log(1-\epsilon)$. In the case where both ϵ and p are constants, these two results match up to an additive constant. It is a rather interesting observation, besides the fact that it implies the optimal of Corollary 8.2 and the tightness of Theorem 8.5. Notice that Theorem 8.5 is proven for protocols that only output a single qubit pair — a minimal possible yield, while the construction from the random hash protocol used by Corollary 8.2 outputs $(n-s)$ qubits — an asymptotically maximum possible yield.[3] Despite the two extreme cases on the yield of the protocols, this two result match nicely.

Proof: (of Theorem 8.5) WLOG we assume the protocol only outputs one qubit pair, i.e., $m=1$, by the monotonicity of fidelity. Consider a particular input state

$$\rho_0=(1-\epsilon')\Psi_n+\epsilon'\cdot\frac{I}{2^{2n}} \tag{8.13}$$

It is a mixture of the perfect EPR pairs Φ_n (with probability $1-\epsilon'$) 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 $\epsilon'=\frac{2^{2n}}{2^{2n}-1}\epsilon$, then we have $F(\rho)=1-\epsilon$. We shall prove that no deterministic, s-bit protocol has fidelity more than $1-2^{-(s+1)}\epsilon p$ over state ρ_0, which implies the theorem.

We fix a 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: one bit is sent 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

$$\mathsf{F}^c(\mathcal{P}(\rho_0))=\frac{(1-\epsilon')p\cdot\mathsf{F}^{\mathsf{c}}(\mathcal{P}(\Phi_n))+\epsilon' q\cdot\mathsf{F}^{\mathsf{c}}(\mathcal{P}(\frac{I}{2^{2n}}))}{(1-\epsilon')p+\epsilon' q} \tag{8.14}$$

[3] Notice that because of the exponential trade-off, it is normally sufficient to have $s=o(n)$, and in that case the random hash protocol outputs almost all the input qubit pairs.

Notice that we always have $\mathsf{F}^{\mathsf{c}}(\mathcal{P}(\Phi_n)) \leq 1$. Since $\frac{I}{2^{2n}}$ is a disentangled state, $\mathcal{P}(\frac{I}{2^{2n}})$ is also disentangled. Therefore we have $\mathsf{F}^{\mathsf{c}}(\mathcal{P}(\frac{I}{2^{2n}})) \leq 1/2$ by Lemma 2.1. We shall prove that

$$q \geq p^2/2^s, \tag{8.15}$$

which will imply that

$$\mathsf{F}(\mathcal{P}(\rho_0)) \leq \frac{(1-\epsilon') + \epsilon' p/2^{s+1}}{(1-\epsilon') + \epsilon' p/2^s} = 1 - \frac{\epsilon' p}{2^{s+1}(1 - \frac{2^s}{2^s-1}\epsilon' p)} \leq 1 - \epsilon p/2^{s+1} \tag{8.16}$$

Now we prove that $q \geq p^2/2^s$. 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 local 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 local 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 $\tau_0^{\mathrm{I},A}$, $\tau_0^{\mathrm{I},A}$, $\tau_0^{\mathrm{II},A}$, and $\tau_0^{\mathrm{II},A}$ 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 local 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 $t[0], t[1], ..., t[k-1]$. We use p_t^{I} to denote the probability that this happens (that the first k bits are $t[0], t[1], ..., t[k-1]$). 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

$$\sum_{t\in\{0,1\}^k} p_t^{\mathrm{I}} = 1 \tag{8.17}$$

$$\sum_{t\in\{0,1\}^k} p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A} = \tau_k^{\mathrm{I},A} \tag{8.18}$$

$$\sum_{t\in\{0,1\}^k} p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},B} = \tau_k^{\mathrm{I},B} \tag{8.19}$$

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 local density matrices for Alice and Bob are identical in both cases:

$$\sigma_\xi^{\mathrm{I},A} = \sigma_\xi^{\mathrm{I},B} = \sigma_\xi^{\mathrm{II},A} = \sigma_\xi^{\mathrm{II},B} = \frac{I}{2^n} \tag{8.20}$$

When the protocol proceeds, the local 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:

Lemma 8.10 *For all $k = 0, 1, ..., s-1$ and all $t \in \{0,1\}^k$, we have $p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A} \preceq \sigma_t^{\mathrm{II},A}$ and $p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},B} \preceq \sigma_t^{\mathrm{II},B}$.*

Proof: 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 $t[0], t[1], ..., t[k-1]$. 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

$$\mathcal{E}_0(\sigma_t^{\mathrm{I},A}) = \frac{p_{t;0}^{\mathrm{I}}}{p_t^{\mathrm{I}}} \cdot \sigma_{t;0}^{\mathrm{I},A} \tag{8.21}$$

$$\mathcal{E}_1(\sigma_t^{\mathrm{I},A}) = \frac{p_{t;1}^{\mathrm{I}}}{p_t^{\mathrm{I}}} \cdot \sigma_{t;1}^{\mathrm{I},A} \tag{8.22}$$

$$\mathcal{E}_0(\sigma_t^{\mathrm{II},A}) = \frac{p_{t;0}^{\mathrm{II}}}{p_t^{\mathrm{II}}} \cdot \sigma_{t;0}^{\mathrm{II},A} \tag{8.23}$$

$$\mathcal{E}_1(\sigma_t^{\mathrm{II},A}) = \frac{p_{t;1}^{\mathrm{II}}}{p_t^{\mathrm{II}}} \cdot \sigma_{t;1}^{\mathrm{II},A} \tag{8.24}$$

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

$$p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A} \preceq \sigma_t^{\mathrm{II},A} \tag{8.25}$$

Combining (8.25), (8.21) and (8.23) with Lemma 2.4 yields that

$$p_{t;0}^{\mathrm{I}} \cdot \sigma_{t;0}^{\mathrm{I},A} = \mathcal{E}_0(p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A}) \preceq \mathcal{E}_0(\sigma_t^{\mathrm{II},A}) = \frac{p_{t;0}^{\mathrm{II}}}{p_t^{\mathrm{II}}} \cdot \sigma_{t;0}^{\mathrm{II},A} \preceq \sigma_{t;0}^{\mathrm{II},A} \tag{8.26}$$

Combining (8.25), (8.22) and (8.24) with Lemma 2.4 yields that

$$p_{t;1}^{\mathrm{I}} \cdot \sigma_{t;1}^{\mathrm{I},A} = \mathcal{E}_1(p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A}) \preceq \mathcal{E}_1(\sigma_t^{\mathrm{II},A}) = \frac{p_{t;1}^{\mathrm{II}}}{p_t^{\mathrm{II}}} \cdot \sigma_{t;1}^{\mathrm{II},A} \preceq \sigma_{t;1}^{\mathrm{II},A} \tag{8.27}$$

Now we consider the local 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 local 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

$$\sigma_t^{\mathrm{I},B} = \frac{p_{t;0}^{\mathrm{I}}}{p_t^{\mathrm{I}}} \cdot \sigma_{t;0}^{\mathrm{I},B} + \frac{p_{t;1}^{\mathrm{I}}}{p_t^{\mathrm{I}}} \cdot \sigma_{t;1}^{\mathrm{I},B} \tag{8.28}$$

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 local density matrix remains unchanged. Thus we have

$$\sigma_t^{\mathrm{II},B} = \sigma_{t;0}^{\mathrm{II},B} = \sigma_{t;1}^{\mathrm{II},B} \tag{8.29}$$

By inductive hypothesis, we have

$$p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},B} \preceq \sigma_t^{\mathrm{II},B} \tag{8.30}$$

Combining (8.28), (8.29), and (8.30), we have

$$p_{t;0}^{\mathrm{I}} \cdot \sigma_{t;0}^{\mathrm{I},B} \preceq p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},B} \preceq \sigma_t^{\mathrm{II},B} = \sigma_{t;0}^{\mathrm{II},B} \tag{8.31}$$
$$p_{t;1}^{\mathrm{I}} \cdot \sigma_{t;1}^{\mathrm{I},B} \preceq p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},B} \preceq \sigma_t^{\mathrm{II},B} = \sigma_{t;1}^{\mathrm{II},B} \tag{8.32}$$

So the inductive case is proved. ■

Now we are ready to prove (8.15). 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 $t[0], t[1], ..., t[s-1]$. Notice we have $p_t^{\mathrm{I}} \cdot \sigma_t^{\mathrm{I},A} \preceq \sigma_t^{\mathrm{II},A}$, and thus by Lemma 8.10, we know that in case II, the success probability is at least $p_t^{\mathrm{I}} \cdot r_t$.

Therefore, we have

$$
\begin{aligned}
p &= \sum_{t \in \{0,1\}^s} r_t \cdot p_t^{\mathrm{I}} && (8.33) \\
q &\geq \sum_{t \in \{0,1\}^s} r_t \cdot p_t^{\mathrm{I}} \cdot p_t^{\mathrm{I}} && (8.34)
\end{aligned}
$$

which implies that

$$
\begin{aligned}
q &\geq \sum_{t \in \{0,1\}^s} r_t \cdot \left(p_t^{\mathrm{I}}\right)^2 && (8.35) \\
&\geq \frac{1}{2^s} \left(\sum_{t \in \{0,1\}^s} r_t \right) \cdot \left[\sum_{t \in \{0,1\}^s} r_t \cdot \left(p_t^{\mathrm{I}}\right)^2 \right] && (8.36) \\
&\geq \frac{1}{2^s} \left(\sum_{t \in \{0,1\}^s} r_t \cdot p_t^{\mathrm{I}} \right)^2 && (8.37) \\
&= \frac{p^2}{2^t} && (8.38)
\end{aligned}
$$

This proves the theorem. ■

Bibliography

[1] D. Aharonov, A. Ta-Shma, U. Vazirani, and A. Yao. Quantum bit escrow. In *STOC 2002*, pp. 705–714.

[2] L. von Ahn, M. Blum, N. Hopper, J. Langford, and K. Yang. Beacon Bits. *manuscript, in preparation*, 2002.

[3] N. Alon, U. Maurer, and A. Wigderson. private communication.

[4] A. Ambainis. A new protocol and lower bounds for quantum coin flipping. In *STOC 2001*.

[5] A. Ambainis. private communication.

[6] Y. Aumann and M.O. Rabin. Information Theoretically Secure Communication in the Limited Storage Space Model. in *Crypto 99*:65-79, 1999.

[7] A. Ambainis, A. Smith, and K. Yang. Extracting Quantum Entanglement (General Entanglement Purification Protocols). in the *IEEE Conference of Computational Complexity (CCC 2002)*. pp. 103-112, 2002.

[8] A. Ambainis and K. Yang. Towards the Classical Communication Complexity of Entanglement Distillation Protocols with Incomplete Information. To appear in the *19th Annual IEEE Conference of Computational Complexity (CCC 2004)*, Amherst, MA. Also available at *LANL eprint* `quant-ph/0207090`.

[9] B. Barak. Constant-Round Coin-Tossing With a Man in the Middle or Realizing the Shared Random String Model. In *FOCS* 2002, pp. 345–355, 2002.

[10] M. Blum, Coin flipping by phone. In *IEEE Spring COMPCOM*, pp. 133–137, February 1982.

[11] M. Blum. Independent unbiased coin flips from a correlated biased source: a finite state Markov chain. In *FOCS 1984*, pp. 425–433, 1984.

[12] R. E. Blahut, Theory and Practice of Error Control Codes. *Addison-Wesley*, 1983.

[13] C. Bennett. Quantum cryptography using any two nonorthogonal states. In *Phys. Rec. Lett.*, 68(21):3121–3124, 1992.

[14] D. Bohm. A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. In *Phys. Rev.* 85, 166–179 (1952).

[15] G. Brassard, Quantum Communication Complexity (a survey). Available at *LANL eprint* `quant-ph/0101005`.

[16] C. H. Bennett and G. Brassard. Quantum cryptography: public key distribution and coin tossing. In *Proceeding of IEEE International Conference on Computers, Systems and Signal Processing*, pp. 175–179, Bangalore, India, December 1984.

[17] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin. Experimental Quantum Cryptography. In *Journal of Cryptology*, vol. 5, no. 1, pp. 3–28, 1992.

[18] 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.*, **70**, 1895 (1993).

[19] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum Lower Bounds by Polynomials. In *39th IEEE Symposium on Foundations of Computer Science (FOCS'98)*, pp.352-361, also available at *LANL eprint* `quant-ph/9802049`. Journal version in *Journal of the ACM*, 48(4):778-797, 2001.

[20] E. Biham, G. Brassard, D. Kenigsberg, and T. Mor. Quantum computing without entanglement. *manuscript*, 2002.

[21] 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.

[22] C. H. Bennett, H. J. Bernstein, 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, page 722, 1996.

[23] H. Barnum, C. Crépeau, D. Gottesman, A. Smith and A. Tapp. Authentication of Quantum Messages. To appear in *FOCS 2002*, also available at *LANL eprint* `quant-ph/0205128`.

[24] S. L. Braunstein, C. M. Caves, R. Jozsa, N. Linden, S. Popescu, and R. Schack. Separability of Very Noisy Mixed States and Implications for NMR Quantum Computing. In *Phys. Rev. Lett.*, **83**, 1054 (1999).

[25] 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, November 1996.

[26] C. H. Bennett, D. P. DiVincenzo, and R. Linsker. Digital recording system with time-bracketed authentication by on-line challenges and method for authenticating recordings. *US patent* 5764769 (1998).

[27] G. Brassard and L. Salvail. Secret-key reconciliation by public discussion. In *Advances in Cryptology — EUROCRYPT '93*, LNCS 765, pp. 410–423, 1994.

[28] C. H. Bennett and J. A. Smolin. Trust enhancement by multiple random beacons. In *LANL eprint* `http://xxx.lanl.gov/abs/cs.CR/0201003` (2002).

[29] 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).

[30] C. Cachin and U. Maurer. Linking Information Reconciliation and Privacy Amplification. In *Journal of Cryptology*, vol. 10, no. 2, pp. 97-110, 1997.

[31] C. Cachin and U.Maurer. Unconditional Security Against Memory-Bounded Adversaries. In *Advances in Cryptology - CRYPTO '97*, LNCS 1294, pp. 292–306, 1997.

[32] T. M. Cover and J. A. Thomas. Elements of Information Theory. *John Wiley and Sons*, 1991.

[33] Y. Z. Ding. Oblivious Transfer in the Bounded Storage Model. In *Advances in Cryptology — CRYPTO 2001*, LNCS 2139, pages 155 – 177, 2001.

[34] S. Dziembowski and U. Maurer. Tight Security Proofs for the Bounded-Storage Model. In *STOC* 2002.

[35] A. Einstein, B. Podolsky, and N. Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? In *Phys. Rev.* **47**, 777 (1935), also reprinted in *Quantum Theory and Measurement*, edited by J. A. Wheeler and W. Z. Zurek, Princeton University Press, 1983.

[36] D. Gottesman. Stabilizer codes and quantum error correction. *Ph.D. thesis, Caltech*, also available at *LANL eprint* `quant-ph/9705052`.

[37] D. Gottesman. *Private Communication*, 2001.

[38] V. Guruswami. List decoding of error-correcting codes. *Ph.D. thesis*, MIT, 2001.

[39] V. Guruswami. List Decoding with Side Information. In *IEEE Conference on Computational Complexity*, 2003.

[40] O. Goldreich. The Foundations of Cryptography - Volume 1, *Cambridge University Press*, 2001.

[41] 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 *LNAL erint* `quant-ph/9903001`.

[42] M. Horodecki, P. Horodecki, and R. Horodecki. Distillability of Inseparable Quantum Systems. In `quant-ph/9607009`.

[43] M. Horodecki, P. Horodecki. Reduction criterion of separability and limits for a class of protocols of entanglement distillation. In *LANL eprint* `quant-ph/9708015`.

[44] M. Horodecki, P. Horodecki, and R. Horodecki. Mixed-state entanglement and distillation: is there a "bound" entanglement in nature? in *LANL eprint* `quant-ph/9801069`.

[45] 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`.

[46] A. Harrow and H. K. Lo. A tight lower bound on the classical communication cost of entanglement dilution. In *LANL eprint* `quant-ph/0204096`.

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

[48] R. Jozsa and N. Linden. On the role of entanglement in quantum computational speed-up. Available at *LANL e-print* `quant-ph/0201143`.

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

[50] H. Klauck. One-way communication complexity and the Nečiporuk lower bound on formula size. available at *LANL eprint* `http://xxx.lanl.gov/abs/cs.CC/0111062`, conference versions at ISAAC '97, Complexity '98, STOC '00.

[51] H. Klauck. Lower Bounds for quantum communication complexity. In the *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS'01)*, 2001. Also available at *LANL eprint* `quant-ph/0106160`.

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

[53] B. Leslau. Attacks on symmetric quantum coin-tossing protocols. In *LANL eprint* `quant-ph/0104075`.

[54] Y. Lindell. Parallel Coin-Tossing and Constant-Round Secure Two-Party Computation. In *Advances in Cryptology — CRYPTO 2001*, LNCS 2139, pages 171 – 189, 2000.

[55] H. K. Lo. Classical communication cost in distributed quantum information processing — a generalization of quantum communication complexity. In *LANL eprint* `quant-ph/9912009`.

[56] H. K. Lo and H. F. Chau. Why quantum bit commitment and ideal quantum coin tossing are impossible. In *Physica D*, 120:177–187, 1998.

[57] 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 eprint* `quant-ph/9803006`.

[58] 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 eprint* `quant-ph/9902045`.

[59] P. Lancaster and M. Tismenetsky. The theory of matrices, second edition, with applications. Academic Press, 1985.

[60] U. M. Maurer. Conditionally-perfect secrecy and a provably secure randomized cipher. In *Journal of Cryptology,* 5:53-66, 1992.

[61] U. M. Maurer. Secret key agreement by public discussion from common information. In *IEEE Transactions on Information Theory*, vol 39, pp. 733–742, May 1993.

[62] D. Mayers, L. Salvail, and Y. Chiba-Kohno. Unconditionally secure quantum coin-tossing. In *LANL eprint* `9904078`.

[63] E. Mossel and R. O'Donnell. Coin Flipping from a Cosmic Source: On Error Correction of Truly Random Bits. *manuscript.*

[64] R. Motwani and P. Raghavan. Randomized algorithms. *Cambridge University Press*, 1995.

[65] M. Naor, A. Orlitsky, and P. Shor. Three results on interactive communication. In *IEEE Transactions on Information Theory*, 39:1608–1615, 1993.

[66] J. von Neumann, Various techniques used in connection with random digits. In *Notes by G. E. Forsythe, National Bureau of Standards*, 1952, vol. 12, pages 36-38.

[67] M. Nielsen. Conditions for a Class of Entanglement Transformations. In *Phys. Rev. Lett.*, **83** (2), pp 436–439 (1999), also available at *LANL eprint* `quant/ph/9811053`.

[68] M. Nielsen. Probability Distributions Consistent with a Mixed State. Available at *LANL eprint* `quant/ph/9909020`.

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

[70] N. Nisan and A. Ta-Shma. Extracting randomness: a survey and new constructions. In *JCSS* 58(1): pp. 148–173, 1999.

[71] N. Nisan and D. Zuckerman. Randomness is linear in space. In *JCSS* 52(1): pp. 43–52, 1996. A preliminary version under the name "More deterministic simulation in logspace" appeared in *STOC 1993*, pp. 235–244, 1993.

[72] A. Orlitsky. Worst-case interactive communication I: Two messages are almost optimal. In *IEEE Transactions on Information Theory*, 36(5):1111–1126, 1990.

[73] A. Peres. Quantum theory: concepts and methods. *Kluwer Academic*, 1993.

[74] M. Rabin. Transaction Protection by Beacons. In *Journal of Computer and System Sciences*, 27(2):256-267, October 1983.

[75] E. M. Rains. A rigorous treatment of distillable entanglement. In *LANL eprint* `quant-ph/9809078`.

[76] E. M. Rains. An improved bound on distillable entanglement. In *LANL eprint* `quant-ph/9809082`.

[77] E. M. Rains. A semidefinite program for distillable entanglement. In *LANL eprint* `quant-ph/0008047`.

[78] A. Razborov. Quantum Communication Complexity of Symmetric Predicates. (Russian). To appear in *Izvestia of the Russian Academy of Science, mathematics*, No 6, 2002. English version available at *LANL eprint* `quant-ph/0204025`.

[79] O. Reingold, R. Shaltiel and A. Wigderson. Extracting randomness via repeated condensing. In *FOCS 2000*, pp. 22–31, 2000.

[80] M. Santha and U. V. Vazirani. Generating quasi-random sequences from slightly-random sources. In *FOCS 1984*, pp. 434–440, 1984.

[81] R. Shaltiel. Recent developments in extractors. Available at `http://www.wisdom.weizmann.ac.il/~ronens/papers/survey.ps`.

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

[83] P. Shor. Scheme for Reducing Deconherence in Quantum Memory. In *Phys. Rev. A* **52**, 2493 (1995).

[84] 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 *LANL eptin* `quant-ph/0003004`.

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

[86] M. Sudan. Coding theory: Tutorial & Survey. In *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, pages 36-53, 2001.

[87] M. Sudan. Algorithmic Introduction to Coding Theory. Course note available at `http://theory.lcs.mit.edu/ madhu/FT01/`.

[88] R. Spekkens and T. Rudolph. Degrees of concealment and bindingness in quantum bit commitment protocols. In *Phys. Rev. A*, 65:012310, 2002.

[89] A. Ta-Shma, C. Umans, D. Zuckerman. Loss-less condensers, unbalanced expanders and extractors. In *Proceedings of STOC'01*, pp. 143-152.

[90] G. Vidal. Entanglement of pure states for a single copy. In *Phys. Rev. Lett.* 83 (1999) 1046-1049, also available at *LNAL eprint* `quant-ph/9902033`.

[91] G. Vidal, D. Jonathan, and M. Nielsen. Approximation transformations and tobust manipulation of bipartite pure state entanglement. In *Phys. Rev. A* **62**, 012304 (2000), also available at *LANL epint* `quant-ph/9910099`.

[92] R. F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model In *Phys. Rev. A* (40), 4277 (1989).

[93] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. In *Nature*, **299**, 802 (1982).

[94] A. Yao. Some Complexity Questions Related to Cistributed Computing. In *Proceedings of the 11th ACM Symposium on Theory of Computing (STOC'79)*, pp. 209–213, 1979.

[95] A. Yao. Quantum Circuit Complexity. In *Proceedings of the 34th IEEE Symposium on Foundations of Computer Science, (FOCS'93)*, pp. 351–361, 1993.

[96] K. Yang. On the (im)possibility of non-interactive correlation distillation. Appeared in the *Latin American Theoretical INformatics (LATIN 2004)*, Buenos Aires, Argentina, 2004.

Appendix A

Private Communication with Ambainis and Gottesman

I attach the results from Ambainis and Gottesman on entanglement distillation protocols that beat quantum error correcting codes.

A.1 Quoted communication from Daniel Gottesman

```
The most interesting one is when you have 9 EPR pairs and
at most 2 errors.  The smallest QECC to correct 2 errors
encodes 1 qubit is 11.

Using two-way communications, you can use the following
procedure: divide the 9 EPR pairs up into a group of 5 and
a group of 4.  On the group of 5, measure the 4 generators
of 5-qubit code (which has distance 3, and can therefore
correct 1 general error, or detect 2 errors).  On the group
of 4, measure the 2 generators of the [[4,2,2]] code (that
is, X X X X and Z Z Z Z, parity checks in the X and Z bases).
The 4-qubit code has distance 2, which means it cannot correct
```

a general error, but it can detect any single error. We initially use the information to detect errors on the two sets. We divide the results up into 3 cases:

1) error detected on group of 5, no error on group of 4

In this case, there is at least one error in the group of 5, so there could only have been at most one error in the group of 4, which we would have detected. Therefore, there were no errors in the group of 4, and we can use the 2 remaining pairs from that group.

Result: 2 EPR pairs.

2) no error on group of 5 (there may or may not be an error detected on the group of 4).

In this case, we know there cannot be any errors on the group of 5, or we would have detected them. Therefore, we can use the one remaining EPR pair from the group of 5 safely.

Result: 1 EPR pair.

3) error detected in both groups

In this case, we know there is exactly one error in each group. The group of 4 is hopeless -- we cannot correct errors, but the group of 5 is also a code to correct one general error, and we know there is only one error there.

Therefore, we can correct that error, and extract a single
good EPR pair

Result: 1 EPR pair.

In all cases, we get at least 1 good EPR pair out.

A.2 Quoted communication from Andris Ambainis

here is a very simple particular case of what you wrote to me a while ago. Take 4 EPR pairs (8 bits). Measure XOR of all odd bits, destroying the 4th pair.

1) If it is 1, take the 2nd and the 3rd pairs, measure the XOR of their odd bits, destroying the 3rd pair. If this XOR is 0, we know that the 2nd pair does not have an error. If it is 1, the 1st pair doesn't have an error.

2) If it is 0, measure the XOR of the even bits of the 2nd and the 3rd pair. The rest is similar to 1).

In contrast, the smallest quantum error correcting code for correcting one error uses 5 qubits. So, we have another case where our protocols beat QECCs for small number of qubits/errors.

Appendix B

List of Symbols

B.1 Mathematical Notations

X, Y, Z : Pauli matrices

Φ^+, Φ^-, Ψ^+, Ψ^- : Bell states

Φ_n : n EPR pairs

$\mathsf{E}(|\,\phi\rangle)$: the entanglement of the pure state $|\,\phi\rangle$

$\mathcal{E}(\rho)$: a superoperator over the mixed state ρ

$\mathcal{H}_N$: a Hilbert space of dimension N.

$S(\rho)$: the von Neumann entropy of the mixed state ρ

(n, k, d)**-code** : a classical error correcting code

$[n, k, d]$**-code** : a linear classical error correcting code

$[\![n, k, d]\!]$**-code** : a quantum error correcting code

B.2 Protocols

(Σ, n, m)**-protocol** : a classical correlation distillation protocol over alphabet Σ with inputs from $\Sigma^n \times \Sigma^n$ and outputs in $\Sigma^m \times \Sigma^m$

(n, m)**-protocol** : a quantum entanglement distillation protocol with inputs from $\mathcal{H}_{2^n} \otimes \mathcal{H}_{2^n}$ and outputs in$\mathcal{H}_{2^m} \otimes \mathcal{H}_{2^m}$

B.3 Noise Models

$\mathcal{B}^{\mathsf{c}}_{n,r}$: the classical bounded corruption model

$\mathcal{BE}_{n,r}$: the bounded erasure model

$\mathcal{B}^{\mathsf{q}}_{n,r}$: the quantum bounded corruption model

$\mathcal{M}_{n,t}$: the bounded measurement model

$\mathcal{D}_{n,p}$: the depolarization model

$\mathcal{E}_{n,k}$: the entanglement noise model

$\mathcal{F}_{n,a}$: the fidelity noise model

$\mathcal{T}_{n,m,\alpha}$: the purity testing protocol

Index