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

Ke Yang
Computer Science Department, Carnegie Mellon University,
5000 Forbes Ave. Pittsburgh, PA 15213, USA;
yangke@cs.cmu.edu

December 8, 2003

Abstract

We study the problem of non-interactive correlation distillation (NICD). Suppose Alice and Bob each has a string, denoted by $A = a_0 a_1 \cdots a_{n-1}$ and $B = b_0 b_1 \cdots b_{n-1}$, respectively. Furthermore, for every $k = 0, 1, ..., n-1$, (a_k, b_k) is independently drawn from a distribution $\mathcal{N}$, known as the "noise mode". Alice and Bob wish to "distill" the correlation non-interactively, i.e., they wish to each apply a function to their strings, and output one bit, denoted by X and Y, such that $\mathsf{Prob}\left[X = Y\right]$ can be made as close to 1 as possible. The problem is, for what noise model can they succeed? This problem is related to various topics in computer science, including information reconciliation and random beacons. In fact, if NICD is indeed possible for some general class of noise models, then some of these topics would, in some sense, become straightforward corollaries.

We prove two negative results on NICD for various noise models. We prove that for these models, it is impossible to distill the correlation to be arbitrarily close to 1. We also give an example where Alice and Bob can increase their correlation with one bit of communication. This example, which may be of its own interest, demonstrates that even the smallest amount of communication is provably more powerful than no communication.

An extended abstract of this paper is to appear in the Latin American Theoretical INformatics (LATIN 2004), Buenos Aires, Argentina, 2004.

1 Introduction

1.1 Non-Interactive Correlation Distillation

Consider the following scenario. Let $\mathcal{N}$ be a distribution over $\Sigma \times \Sigma$, where Σ is an alphabet. We call $\mathcal{N}$ a "noise model." Suppose Alice and Bob each receives a string $A = a_0 a_1 \cdots a_{n-1}$ and $B = b_0 b_1 \cdots b_{n-1}$, respectively, as their local inputs. For every $k = 0, 1 ..., n-1$, (a_k, b_k) is independently drawn from $\mathcal{N}$. Now, Alice and Bob wish to engage in a protocol to "distill" their correlation. An the end of the protocol, they wish to each output a bit, denoted by X and Y, respectively, such that both X and Y are "random enough", while $\mathsf{Prob}\left[X = Y\right]$ can be made as close to 1 as possible, possibly by increasing n. We call such a protocol a *correlation distillation protocol*. Furthermore, if Alice and Bob wish to do so *non-interactively*, i.e., without communication, we call this "non-interactive correlation distillation" (NICD). Notice that in NICD, the most general thing for Alice and Bob to do is to each apply a function to their local inputs and outputs one bit. The problem of NICD is, for what noise model can Alice and Bob achieve this goal?

We note that NICD is indeed possible for many noise models. For example, if a noise model $\mathcal{N}$ is in fact "noiseless," i.e. $\mathsf{Prob}_{(a,b)\in\mathcal{N}}[a = b] = 1$, then NICD is possible. However, we are interested in the "noisy" noise models, for example, the *binary symmetric model*, where Alice and Bob each has an unbiased bit as input, which agree with probability $1 - p$, and the *binary erasure model*, where Alice's input is an unbiased bit x, and Bob's input is x with probability $1 - p$, and a special symbol $\perp$ with probability p. These models are extensively studied in the context of error correcting codes [4, 11], where Alice encodes her information before sending it through a "noisy channel". It is known that there exists efficient encoding schemes that withstand these noise models and allow Alice and Bob to achieve almost perfect correlation. However, in the case of NICD, the "raw data" are already noisy. Can the techniques in error correcting codes be used here, and is NICD possible for these noise models?

1.2 Motivations and Related Work

Besides the obvious relation to error correcting codes, the study of NICD is naturally motivated by several other topics. We review these topics and discuss some of the related work.

Information Reconciliation Information reconciliation is an extensively studied topic [5, 15, 8, 9, 10] with applications in quantum cryptography and information-theoretical cryptography. In this setting, Alice and Bob each receives a sequence of random bits drawn from a noise model, while Eve, the eavesdropper, also possesses some information about the their bits. Alice and Bob wish to "reconcile" their information via an "information reconciliation protocol", where they exchange information in a noiseless, public channel in order to agree on a random string U with very high probability. Therefore, information reconciliation protocols are somewhat like correlation distillation protocols. However, the primary concern for information reconciliation is *privacy*, i.e., that Eve gains almost no information about U. Notice that Eve can see the conversation between Alice and Bob, and thus maximum privacy would be achieved if information reconciliation can be performed without communication.

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 [17], who showed how they can be used to solve various problems in cryptography. From then on, random beacons have found many applications in security and cryptography [7, 14, 3, 12]. There are many proposals to construct a *publicly verifiable* random beacon, among them are the ones that use the signals from a cosmic source [16]. In these proposals, Alice (as the beacon owner) and Bob (as the verifier) both point a telescope to an extraterrestrial object, e.g. a pulsar, and then measure the signals from it. Presumably these signals contain enough amount of randomness. Then Alice converts her measurement results into a sequence of random bits, and publishes them as beacon bits. Bob can then verify the bits by performing his own measurement and conversion. However, it is inevitable that there would be discrepancies in the results of Alice and Bob, due to measurement errors (described by a noise model). These discrepancies may cause the beacon bits published by Alice to disagree with the ones computed by Bob. One of the major concerns in the study on random beacons is to prevent *cheating* in the presence of measurement error. In other words, one needs to design a mechanism to prevent Alice from maliciously modifying her measurement data in order to affect the beacon bits, while pretending that the modification comes from the measurement error. Notice that in general, there is no communication between Alice and Bob. We note that if NICD is possible, then the cheating problem would be solved, since NICD protocols can be used to distill almost perfectly correlated bits. Then with very high probability, the bits output by Alice and Bob should agree, and this essentially removes the measurement error.

Related Work As we have discussed, the problem of NICD lays, in some sense, at the foundations of both the studies of information reconciliation and random beacons. In fact, Researchers from both ares have, to some extent, considered the problem of NICD. In particular, a basic version of the problem concerning only the binary symmetric noise model was discovered and proven independently by several researchers since as early as 1991, including Alon, Maurer,and Wigderson [2] and Mossel and O'Donnell [16]. They proved that NICD is impossible over the binary symmetric noise model. Mossel and O'Donnell studied *multi-party* version of this problem, where $k > 2$ parties wish to agree on some random bits. They also only considered the binary symmetric noise model. In fact, we are not aware of any prior work that studies NICD beyond the binary symmetric noise model.

We stress the importance of understanding the problem of NICD for general noise models. As we have mentioned, this problem is important to both the studies of information reconciliation and random beacons. In both studies, there is no reason to assume that the binary symmetric noise model is the only reasonable one. As an example, the measurement of the signals from extraterrestrial objects is not unique, and different measurements may yield different noise models. If one of these noise models admits NICD, then the problems of information reconciliation and random beacon could, in some sense, be solved. Therefore, a better understanding of NICD over more general class of noise models would be very helpful.

1.3 Our Contribution

We study NICD beyond the binary symmetric noise model. First, we prove an impossibility result for NICD over a class of so-called "regular" noise models in Section 3. Intuitively, a noise model $\mathcal{N}$ is regular if it satisfies the following three requirements: that it is *symmetric*, i.e., $\mathcal{N}(a,b) = \mathcal{N}(b,a)$ 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 $\mathcal{N}(a,b) = \mathcal{N}(b,a) = 0$ for all $a \in \Sigma_0$ and $b \in \Sigma_1$. Notice that if a noise model is not connected, that NICD 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 admits NICD.

In section 4, we move over to the binary erasure noise model. It is the simplest noise model that is not symmetric, and thus is not regular. The binary erasure model is also 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 perfect accuracy in her measurement, but Bob's measurement is noisy. Such a situation can be described by the binary erasure noise model. We prove that NICD is impossible for this noise model as well.

The impossibility results we prove suggest that for many noise models, communication is essential for correlation distillation. Thus it is interesting to ask how much communication is essential, and in particular, if a single bit of communication helps. In Section 5, we answer this question in positive 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 communications at all. The protocol itself may also be of its own interest.

1.4 Organization of the Paper

We present some preliminaries and notations in Section 2. We prove an impossibility result for the class of regular noise models in Section 3. Another negative result for the binary erasure model is proven in Section 4. We present the one-bit protocol for distilling correlation for the binary symmetric model in Section 5. In Section **??**, we conclude the paper with open problems. Some of the proofs are postponed to the Appendix.

2 Preliminaries and Notations

We use $[n]$ to denote the set $\{0,1,...,n-1\}$. We often work with symbols from a particular *alphabet*, which is a finite set of cardinality q and is normally denoted by Σ. We often identify Σ with $[q]$.

All vectors are column vectors by default. A *string* is a sequence of symbols from an alphabet. We identify a string with a vector and 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 use $\mathbf{1}_n$ to denote the all-one vector (whose each entry is 1) of dimension n. When the dimension is clear from the context, it is often omitted.

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

We will work with tensor products. Let A and B both be vectors or both be matrices. We use $A \otimes B$ to denote the tensor product of A and B, and $A^{\otimes n}$ to denote the n-th tensor power of A, which is the tensor product of n copies of A.

Definition 1 (Noise Model) *A* noise model *over an alphabet Σ, often denoted by $\mathcal{N}$, is a probabilistic distribution over $\Sigma \times \Sigma$. The n-th tensor power of a noise model $\mathcal{N}$ is the distribution of a pair of length-n strings (A,B), where $A = a_0a_1\cdots a_{n-1}$ and $B = b_0b_1\cdots b_{n-1}$, and (a_k,b_k) is independently drawn from $\mathcal{N}$ for $k = 0,1,...,n-1$.*

In this paper we study *randomized, non-interactive* protocols. For the impossibility results in Section 3 and Section 4, we assume that Alice and Bob each outputs a single bit, since it suffices to prove a negative result on the "minimally useful" protocols. We shall consider protocols that outputs multiple bits in Section 5.

Since Alice and Bob do not communicate, the most general thing they can do is to apply a (randomized) function to their private inputs and outputs a bit.

Definition 2 (Protocols) *A* protocol $\mathcal{P}$ *over a noise model* $\mathcal{N}$ *is a family of function pairs* (ϕ_n^A, ϕ_n^B) *for* $n > 0$, *where* $\phi_n^A, \phi_n^B : \Sigma^n \mapsto [-1, 1]$ *are called the* characteristic functions. *The* output *of protocol* $\mathcal{P}$ *over noise model* $\mathcal{N}$, *denoted by* $\mathcal{P}(\mathcal{N})$, *is a sequence of distributions* $\{\mathcal{D}_1, \mathcal{D}_2, ...\}$, *where the* n*-th distribution* $\mathcal{D}_n$ *is of the bit pair* (X_n, Y_n), *defined as follows.*

$$\boxed{(a, b) \leftarrow \mathcal{N}^{\otimes n};\ x \leftarrow \phi_n^A(a), y \leftarrow \phi_n^B(b);\ X_n \leftarrow \mathsf{B}_{(1+x)/2}, Y_n \leftarrow \mathsf{B}_{(1+y)/2} : (X_n, Y_n)}$$

Where B_p *is the* Bernoulli Distribution *of parameter* p, *defined as* $\mathsf{B}_p(0) = 1 - p$ *and* $\mathsf{B}_p(1) = p$.

Definition 3 (Statistical Distance) *The* statistical distance *between two probabilistic distributions* A *and* B, *denoted as* $\mathsf{SD}(A, B)$, *is defined to be* $\mathsf{SD}(A, B) = \frac{1}{2} \sum_x |A(x) - B(x)|$ *where the summation is taken over the support of* A *and* B. *If* $\mathsf{SD}(A, B) \leq \epsilon$, *we say* A *is* ϵ-close *to* B.

Definition 4 (δ-Locally Uniform Protocols) *A protocol* $\mathcal{P}$ *is* δ-locally uniform *over a noise model* $\mathcal{N}$, *if for every* $n > 0$, *both* X_n *and* Y_n *are* δ*-close to the uniform distribution over* $\{0, 1\}$, *where* (X_n, Y_n) *is the* n*-th distribution of* $\mathcal{P}(\mathcal{N})$. *A protocol is* locally uniform *if it is* 0*-locally uniform.*

Definition 5 (Correlation of Protocols) *The* correlation *of a protocol* $\mathcal{P}$ *over a noise model* $\mathcal{N}$, *denoted by* $\mathsf{Cor}_{\mathcal{N}}[\mathcal{P}]$, *is defined to be*

$$\mathsf{Cor}_{\mathcal{N}}[\mathcal{P}] = \liminf_n \{2 \cdot \mathsf{Prob}\,[X_n = Y_n] - 1\} \tag{1}$$

where (X_n, Y_n) *is the* n*-th distribution of* $\mathcal{P}(\mathcal{N})$.

3 An Impossibility Result for Regular Noise Models

We prove a general impossibility result for NICD over the regular noise models.

Definition 6 (Distribution Matrix) *Let* $\mathcal{N}$ *be a noise model over* Σ, *where* $|\Sigma| = q$. *We say a* $q \times q$ *matrix* M *is the* distribution matrix *for* $\mathcal{N}$, *if* $M_{x,y} = \mathcal{N}(x, y)$ *for all* $x, y \in \Sigma$.[1] *We write the distribution matrix of* $\mathcal{N}$ *by* $M_{\mathcal{N}}$.

Definition 7 (Regular Noise Model) *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* M*'s largest absolute eigenvalue and the second largest. We call* $q \cdot \epsilon$ *the* scaled eigenvalue gap *of* M. *A noise model* $\mathcal{N}$ *is* regular *if its distribution matrix is regular.*

Theorem 1 *If* $\mathcal{N}$ *is a regular noise model over* Σ *with scaled eigenvalue gap* ϵ, *then the correlation of any* δ*-locally uniform protocol over the* $\mathcal{N}$ *is at most* $1 - \epsilon(1 - 4\delta^2)$.

Notice that a distribution matrix M is non-negative (that every entry is non-negative). By the Perron-Frobenius Theorem [13], 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.

[1] Here we identify Σ with $[q]$.

Proof: Consider a protocol $\mathcal{P}$ over the noise model $\mathcal{N}$. We define $q = |\Sigma|$ and identify Σ with $[q]$ for the rest of the proof. We use M to denote the distribution matrix of $\mathcal{N}$ and denote the eigenvector of M by $v_0, v_1, ..., v_{q-1}$ with corresponding eigenvalues $\lambda_0, ..., \lambda_{q-1}$. We assume that $|\lambda_0| > |\lambda_1| \geq \cdots \geq |\lambda_{q-1}|$. 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 $|\lambda_1| = (1-\epsilon)/q$.

Consider the characteristic functions ϕ_n^A and ϕ_n^B. It is easy to see that

$$\mathsf{Prob}\,[X_n = 1] = \frac{1}{2} \cdot \left[1 + \sum_{a \in \Sigma^n} \sum_{b \in \Sigma^n} \mathcal{N}^{\otimes n}(a,b) \cdot \phi^A(a)\right] \tag{2}$$

Clearly, $M^{\otimes n}$ is the distribution matrix for $\mathcal{N}^{\otimes n}$. We will be using a result about the eigenvalues and eigenvectors of $M^{\otimes}$, stated in Lemma 1.

Since $\mathcal{P}$ is δ-locally uniform, we have

$$\left|\sum_{a \in \Sigma^n} \sum_{b \in \Sigma^n} \mathcal{N}^{\otimes n}(a,b) \cdot \phi^A(a)\right| \leq 2\delta \tag{3}$$

or $|(\phi^A)^T \cdot M^{\otimes n} \cdot \mathbf{1}_{q^n}| \leq 2\delta$, as we identify ϕ^A with the q^n-dimensional vector represented by its truth table. 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 $1/q^n$ (see Lemma 1). Since M is symmetric, so is $M^{\otimes n}$. Thus we have

$$|\mathbf{1}_{q^n}^T \cdot \phi^A| \leq 2\delta \cdot q^n. \tag{4}$$

Similarly we have

$$|\mathbf{1}_{q^n}^T \cdot \phi^B| \leq 2\delta \cdot q^n. \tag{5}$$

Now, we consider the correlation of $\mathcal{P}$. Let (X_n, Y_n) be the outputs of Alice and Bob. Then we have

$$2 \cdot \mathsf{Prob}\,[X_n = Y_n] - 1 = \sum_{A \in \Sigma^n} \sum_{B \in \Sigma^n} \mathcal{N}^{\otimes n}(A,B) \cdot \phi^A(A) \cdot \phi^B(B) \tag{6}$$

In other words, we have

$$2 \cdot \mathsf{Prob}\,[X_n = Y_n] - 1 = (\phi^A)^T \cdot M^{\otimes n} \cdot \phi^B \tag{7}$$

We diagonalize the matrix $M^{\otimes n}$. First we define a natural notion of inner product: $\langle A, B \rangle = \frac{1}{q^n} \sum_{x \in \Sigma^n} A[x]B[x]$. It is obvious that under this inner product, both ϕ_n^A and ϕ_n^B have norm at most 1. 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 $|\mu_0| \geq |\mu_1| \geq \cdots \geq |\mu_{q^n-1}|$. By Lemma 1, the eigenvalues μ_t are of the form $\prod_{i=1}^n \lambda_{k_i}$, where $k_i \in [q]$. Therefore $M^{\otimes n}$ has a unique maximum eigenvalue $\mu_0 = \lambda_0^n = 1/q^n$, which corresponds to the eigenvector $\mathbf{1}_q^{\otimes n} = \mathbf{1}_{q^n}$. The second largest absolute eigenvalue of $M^{\otimes n}$ is $|\mu_1| = \lambda_0^{n-1} \cdot |\lambda_1| = (1-\epsilon)/q^n$.

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 by Parseval, we have $\sum_t \alpha_t^2 \leq 1$, $\sum_t \beta_t^2 \leq 1$. Furthermore, from (4) and (5), we have $|\alpha_0| \leq 2\delta$ and $|\beta_0| \leq 2\delta$.

Putting things together, we have

$$\begin{aligned}
\mathsf{Cor}_{\mathcal{N}^{\otimes n}}[\mathcal{P}] &= (\phi^A)^T \cdot M^{\otimes n} \cdot \phi^B \\
&= q^n \cdot \sum_{t\in[q^n]} \alpha_t \cdot \beta_t \cdot \mu_t \\
&\leq \epsilon \cdot |\alpha_0 \beta_0| + (1-\epsilon) \sum_{t\in[q^n]} |\alpha_t \cdot \beta_t| && \text{(eigenvalue gap)} \\
&\leq \epsilon \cdot 4\delta^2 + (1-\epsilon) \left(\sum_t \alpha_t^2\right)\left(\sum_t \beta_t^2\right) && \text{(Cauchy-Schwartz)} \\
&\leq 1 - \epsilon(1 - 4\delta^2).
\end{aligned}$$

■

Lemma 1 *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 $j \in [b]$.*

The proof to this lemma is postponed to Appendix A.

Definition 8 (Binary Symmetric Noise Model) *The* binary symmetric noise model *is a distribution over alphabet $\{0,1\}$, denoted by $\mathcal{S}_p$ and is defined as $\mathcal{S}(0,0) = \mathcal{S}(1,1) = (1-p)/2$ and $\mathcal{S}(0,1) = \mathcal{S}(1,0) = p/2$.*

Corollary 1 *The correlation of any locally uniform protocol over the binary symmetric noise model $\mathcal{S}_p$ is at most $1 - 2p$.*

It is easy to see that this bound is tight, since the naïve protocol where both Alice and Bob outputs their first bits is locally uniform with correlation $1 - 2p$.

Proof: Notice that $\mathcal{S}_p$ is regular with scaled eigenvalue gap $2p$. ■

This corollary was independently discovered by various researchers, including Alon, Maurer,and Wigderson [2], and Mossel and O'Donnell [16], and the latter attributing it as a "folklore".

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 9 (Binary Erasure Noise Model) *The* binary erasure noise model *is a distribution over alphabet $\{0,1,\perp\}$, denoted by $\mathcal{E}_p$ and defined as $\mathcal{E}(0,0) = \mathcal{E}(1,1) = (1-p)/2$, $\mathcal{E}(0,\perp) = \mathcal{E}(1,\perp) = p/2$.*

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 2 *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, NICD is impossible if Bob's measurement is noisy.

The proof to Theorem 2 uses a more involved Fourier analysis and is postponed to Appendix A.

5 A One-bit Communication Protocol

We present a protocol that non-trivially distills correlation over the binary symmetric noise model with one bit of communication. Recall that over 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. With one bit of communication, 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 define the correlation of a protocol to be $2 \cdot \min_{i=1,2} \{\mathsf{Prob}\ [X_i = Y_i]\} - 1$. 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 1.

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

Theorem 3 *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 does not fully utilize the one-bit communication, since she sends 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 better result, since Bob gets more information. Nevertheless, the XOR does not work, in some sense due to its "symmetry". Consider the case 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

```
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 $(b_1, b_2)$.
```

Figure 1: The AND protocol: 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 $\mathcal{S}_p^{\otimes 2}$.

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 $1/4 \cdot 1 + (3/4) \cdot (1 - 2p) = 1 - 3p/2$.

Acknowledgment

The author thanks Luis von Ahn, Manuel Blum, Ning Hu, Joe Kilian, John Langford, and Steven Rudich for helpful discussions, and for Ueli Maurer and Avi Wigderson for providing the references.

References

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

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

[3] Y. Aumann and M.O. Rabin. Information theoretically secure communication in the limited storage space model. in *Crypto 99*:65-79, 1999.

[4] R. E. Blahut, Theory and practice of error control codes. *Addison-Wesley*, 1983.

[5] 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.

[BBP+96a] 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.

[BBP+96b] 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.

[6] 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.

[7] 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).

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

[9] C. Cachin and U. Maurer. Linking information reconciliation and privacy amplification. In *Journal of Cryptology*, vol. 10, no. 2, pp. 97-110, 1997.

[10] C. Cachin and U.Maurer. Unconditional security against memory-bounded adversaries. In *Advances in Cryptology - CRYPTO '97*, LNCS 1294, pp. 292–306, 1997.

[11] T. M. Cover and J. A. Thomas. Elements of information theory. *John Wiley and Sons*, 1991.

[12] Y. Z. Ding. Oblivious transfer in the bounded storage model. In *Advances in Cryptology — CRYPTO 2001*, LNCS 2139, pages 155 – 177, 2001.

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

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

[15] 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.

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

[17] M. Rabin. Transaction Protection by Beacons. In *Journal of Computer and System Sciences*, 27(2):256-267, October 1983.

A Proofs

Lemma 1 *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: (to Lemma 1) 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 above 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 2 *The correlation of any locally uniform protocol over the noise model $\mathcal{E}_p$ is at most $\sqrt{1 - p(1 - 4\delta^2)}$.*

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

$$x[i] = \begin{cases} v[j] & \text{if } u[i] = \star \text{ and } j = |\{l \,:\, 0 \le l < i, u[j] = \star\}| \\ \\ \perp & \text{if } u[i] = \perp \end{cases} \tag{8}$$

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.

$$\phi^A(x) = \sum_s \alpha_s \oplus_s (x) \tag{9}$$

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

$$|\alpha_0| \leq 2\delta. \tag{10}$$

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

$$\psi_u(x) = \sum_s \beta_{u,s} \oplus_s (x) \tag{11}$$

Again we have $\sum_s \beta_{u,s}^2 \leq 1$ 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

$$\sum_{u \in \{\star,\perp\}^n} \lambda^{\deg(u)} = \sum_{k=0}^{n} \binom{n}{k} \lambda^k = \frac{1}{(1-p)^n} \tag{12}$$

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

$$\eta = \left(\frac{1-p}{2}\right)^n \sum_{u \in \{\star,\perp\}^n} \lambda^u \sum_x \phi^A(x)\psi_u(x|_u) \tag{13}$$

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

$$\eta = (1-p)^n \sum_{u \in \{\star,\perp\}^n} \lambda^u \sum_{s \sqsubseteq u} \alpha_s \beta_{u,s|_u}$$

$$
\begin{aligned}
&\leq && (1-p)^n \left(\sum_{u \in \{\star,\perp\}^n} \lambda^u \right)^{1/2} \cdot \left[\sum_{u \in \{\star,\perp\}^n} \lambda^u \left(\sum_{s \sqsubseteq u} \alpha_s \beta_{u,s|_u} \right)^2 \right]^{1/2} \qquad \text{(Cauchy-Schwartz)} \\
&= && (1-p)^{n/2} \cdot \left[\sum_{u \in \{\star,\perp\}^n} \lambda^u \cdot \left(\sum_{s \sqsubseteq u} \alpha_s^2 \right) \cdot \left(\sum_{s \sqsubseteq u} \beta_{u,s|_u}^2 \right) \right]^{1/2} \qquad \text{(Eq. 12)} \\
&\leq && (1-p)^{n/2} \cdot \left[\sum_{u \in \{\star,\perp\}^n} \lambda^u \cdot \left(\sum_{s \sqsubseteq u} \alpha_s^2 \right) \right]^{1/2} \qquad \text{(Parseval, } \textstyle\sum_{s \sqsubseteq u} \beta_{u,s|_u}^2 \leq 1 \text{)} \\
&= && (1-p)^{n/2} \cdot \left[\sum_{s} \alpha_s^2 \sum_{u : s \sqsubseteq u} \lambda^u \right]^{1/2} \\
&= && (1-p)^{n/2} \cdot \left[\sum_{s} \alpha_s^2 \cdot (1+\lambda)^{n-|s|} \right]^{1/2} \\
&\leq && (1-p)^{n/2} \left[(1+\lambda)^{n-1} (1 + 4\delta^2 (1+\lambda)) \right]^{1/2} \qquad \text{(Eq. 10)} \\
&= && \sqrt{1 - p(1-4\delta^2)}
\end{aligned}
$$

$\blacksquare$