3-2014

On Extracting Common Random Bits From Correlated Sources on Large Alphabets

Siu On Chan

Elchanan Mossel
University of Pennsylvania

Joe Neeman

Recommended Citation

On Chan, S., Mossel, E., & Neeman, J. (2014). On Extracting Common Random Bits From Correlated Sources on Large Alphabets. *IEEE Transactions on Information Theory, 60* (3), 1630-1637. http://dx.doi.org/10.1109/TIT.2014.2301155

On Extracting Common Random Bits From Correlated Sources on Large Alphabets

Abstract

Suppose Alice and Bob receive strings $X=(X_1,...,X_n)$ and $Y=(Y_1,...,Y_n)$ each uniformly random in $[s]^n$, but so that X and Y are correlated. For each symbol i, we have that $Y_i=X_i$ with probability $1-\varepsilon$ and otherwise Y_i is chosen independently and uniformly from [s]. Alice and Bob wish to use their respective strings to extract a uniformly chosen common sequence from $[s]^k$, but without communicating. How well can they do? The trivial strategy of outputting the first k symbols yields an agreement probability of $(1-\varepsilon+\varepsilon/s)^k$. In a recent work by Bogdanov and Mossel, it was shown that in the binary case where s=2 and $k=k(\varepsilon)$ is large enough then it is possible to extract k bits with a better agreement probability rate. In particular, it is possible to achieve agreement probability $(k\varepsilon)^{-1/2}\cdot 2^{-k\varepsilon/(2(1-\varepsilon/2))}$ using a random construction based on Hamming balls, and this is optimal up to lower order terms. In this paper, we consider the same problem over larger alphabet sizes s and we show that the agreement probability rate changes dramatically as the alphabet grows. In particular, we show no strategy can achieve agreement probability better than $(1-\varepsilon)^k(1+\delta(s))^k$ where $\delta(s)\to 0$ as $s\to\infty$. We also show that Hamming ball-based constructions have much lower agreement probability rate than the trivial algorithm as $s\to\infty$. Our proofs and results are intimately related to subtle properties of hypercontractive inequalities.

Keywords

hamming codes, probability, random processes, 1-ε probability, hamming ball-based constructions, agreement probability rate, common random bit extraction, correlated sources, hypercontractive inequalities, large alphabet size, lower order terms, random construction, trivial algorithm, correlation, information theory, joints, noise, noise measurement, protocols, upper bound, randomness extraction, hypercontractivity, symmetric channels

Disciplines

Computer Sciences | Statistics and Probability

On extracting common random bits from correlated sources on large alphabets

Siu On Chan*[2], Elchanan Mossel†[1,2], and Joe Neeman‡[1]

[1]Department of Statistics, UC Berkeley
[2]Department of Computer Science, UC Berkeley

January 24, 2014

Abstract

Suppose Alice and Bob receive strings $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$ each uniformly random in $[s]^n$ but so that X and Y are correlated . For each symbol i, we have that $Y_i = X_i$ with probability $1-\epsilon$ and otherwise Y_i is chosen independently and uniformly from $[s]$.

Alice and Bob wish to use their respective strings to extract a uniformly chosen common sequence from $[s]^k$ but without communicating. How well can they do? The trivial strategy of outputting the first k symbols yields an agreement probability of $(1-\epsilon+\epsilon/s)^k$. In a recent work by Bogdanov and Mossel it was shown that in the binary case where $s = 2$ and $k = k(\epsilon)$ is large enough then it is possible to extract k bits with a better agreement probability rate. In particular, it is possible to achieve agreement probability $(k\epsilon)^{-1/2} \cdot 2^{-k\epsilon/(2(1-\epsilon/2))}$ using a random construction based on Hamming balls, and this is optimal up to lower order terms.

In the current paper we consider the same problem over larger alphabet sizes s and we show that the agreement probability rate changes dramatically as the alphabet grows. In particular we show no strategy can achieve agreement probability better than $(1-\epsilon)^k(1+\delta(s))^k$ where $\delta(s) \to 0$ as $s \to \infty$. We also show that Hamming ball based constructions have *much lower* agreement probability rate than the trivial algorithm as $s \to \infty$. Our proofs and results are intimately related to subtle properties of hypercontractive inequalities.

*Supported by NSF grant DMS-1106999 and DOD ONR grant N000141110140
†Supported by NSF grant DMS-1106999 and DOD ONR grant N000141110140
‡Supported by NSF grant DMS-1106999 and DOD ONR grant N000141110140

1 Introduction

For an integer $s \geq 2$, consider two $[s]^n$-valued random variables X, Y (where $[s] = \{0, 1, \ldots, s-1\}$) which are sampled by first choosing X uniformly and then, independently for every coordinate i, taking Y_i to be a copy of X_i with probability $1 - \epsilon$ and an independent sample from $[s]$ otherwise. We will write $\mathbb{P}_\epsilon$ for this joint distribution on X and Y. Note that X and Y are both uniformly distributed in $[s]^n$.

The non-interactive correlation distillation (NICD) is defined as follows: suppose that one party (Alice) receives X and another (Bob) receives Y. Without any communication, each party chooses a string that is uniformly distributed in $[s]^k$ with the goal of maximizing the probability that the two strings chosen by Alice and Bob are identical.

1.1 Motivation and Related Work

This problem was studied in [1] in the case $s = 2$, with motivation from various areas. One major motivation comes from the goal of extracting a unique identification string from process variations [3, 12], particularly in a noisy setup [9].

The case where the goal of the two parties is to extract a single bit was studied independently a number of times; in this case the optimal protocol is for the two parties to use the first bit. See [11] for references and for studying the problem of extracting one bit from two correlated sequences with different correlation structures.

In [4, 5] a related question is studied: if m parties receive noisy versions of a common random string, where the noise of each party is independent, what is the strategy for the m parties that maximizes the probability that the parties agree on a *single* random bit of output without communicating? [4] shows that for large m using the majority functions on all bits is superior to using a single bit and [5] uses hypercontractive inequalities to show that for large m, majority is close to being optimal. Both results were recently extended to general string spaces in [6].

For any $k \in \mathbb{N}$, one protocol – which we will call the "trivial protocol" – is for both parties to take the first k symbols of their strings. The success probability of this protocol is $(1 - (1 - \frac{1}{s})\epsilon)^k \approx \exp(-k\epsilon(1 - \frac{1}{s}))$. When $s = 2$ and the protocol outputs a single bit (ie. $k = 1$), it is known (see e.g. [4]) that the optimal protocol is for both parties to choose the first bit. For larger k, this is no longer true. Bogdanov and Mossel [1] studied the case $s = 2$, and showed that any protocol which outputs a uniformly random length-k string

has a success probability of at most $\exp(-k\epsilon(\ln 2)/2)$. In other words, if p is the success probability of the trivial algorithm for choosing a k-bit string, then every protocol with success probability at least p emits at most $k/\ln 2$ bits.

Bogdanov and Mossel showed that their bound was sharp by providing an example (for a restricted range of ϵ and k) with success probability which, for any $\delta > 0$, is at least $\exp(-k\epsilon(1+\delta)/2)$ for small ϵ and large k. In other words, if p is the success probability of the trivial algorithm for choosing a k-bit string, then they gave a protocol that succeeds with probability p and produces a string of length $k/((1+\delta)\ln 2)$. Their construction was built by taking random translations of Hamming balls; we will return to it in more detail later.

1.2 Our results

We study an extension of the upper bound of [1] to a larger alphabet. In our main result we show that in the case of large alphabets, the constant-factor gap between the upper bound and the performance of the trivial algorithm vanishes; hence, the trivial algorithm is almost optimal for large alphabets. In particular we show no strategy can achieve agreement probability better than $(1-\epsilon)^k(1+\delta(s))^k$ where $\delta(s) \to 0$ as $s \to \infty$.

We then turn to analyze generalizations of the Hamming ball based construction of [1]. Interestingly we show that these have *much lower* agreement probability rate than the trivial algorithm as $s \to \infty$.

In this respect it is interesting to compare the case of a large number of parties that extract a single symbol to the case of two parties who extract a longer string. In the first case, the results of [6] generalize those of [4,5] to show that Hamming ball based protocols are almost optimal for all values of s when the number of parties m is large. In the case presented here, Hamming ball type constructions quickly deteriorate as s increases and the trivial protocol becomes almost optimal.

The difference between the two phenomena may be explained by the fact that the problem studied in [4,5] is closely related to reverse-hypercontractive inequalities which hold uniformly in s [6], while the problem studied here is closely related to hypercontractive inequalities which deteriorate as s increases.

Our results show that the trivial algorithm is optimal up to a factor of $(1+\delta(s))^k$ where $\delta(s) \to 0$ as $s \to \infty$. An interesting open problem is to find an almost optimal algorithm for large s, i.e., an algorithm whose agreement probability is provably optimal up to a factor of $2^{-o(k)}$. It is quite possible

that the trivial protocol is optimal for some large fixed values of s and all large enough k.

2 Definitions and results

A *protocol* for NICD is defined by two functions $f, g : [s]^n \to [s]^*$. Upon receiving their strings $X, Y \in [s]^n$, the two parties compute $f(X)$ and $g(Y)$ respectively. The protocol is successful if both parties agree on the same output; that is, if $f(X) = g(Y)$. Therefore, finding an optimal NICD algorithm is equivalent to finding functions $f, g : [s]^n \to [s]^*$ which maximize $\mathbb{P}_\epsilon(f(X) = g(Y))$.

In the introduction, we mentioned the requirement that f and g are uniformly distributed on $[s]^k$. In fact, we will require less for our negative results and guarantee more in our positive results. In particular, for our negative results, we will only assume that f and g have min-entropy at most k, meaning that $\mathbb{P}(f(X) = z) \le s^{-k}$ for all $z \in [s]^*$ and similarly for g. Of course, if $f : [s]^n \to [s]^k$ is uniformly distributed then it has min-entropy k.

2.1 Reduction to a question about sets

Using an observation of [1], we can reduce the NICD problem to the problem of finding a sets $A \subset [s]^n$ which maximize $\mathbb{P}_\epsilon(Y \in A | X \in A)$. On the one hand, if we are given good functions f and g then we can find a set A such that $\mathbb{P}(Y \in A | X \in A)$ is large:

Theorem 2.1. *For any functions $f, g : [s]^n \to [s]^*$ having min-entropy k there is a set $A \subset [s]^n$ with $|A| \le s^{n-k}$ such that for every $0 \le \epsilon \le 1$,*

$$\mathbb{P}_\epsilon(Y \in A | X \in A) \ge \mathbb{P}_\epsilon(f(X) = g(Y)).$$

On the other hand, if we have a good set A then we can construct a function f by taking certain translates of A.

Theorem 2.2. *If $A \subset [s]^n$ with $\frac{1}{8}s^{n-k} \le |A| \le \frac{1}{4}s^{n-k}$ then there is a function $f : [s]^n \to [s]^k$ such that*

1. *$f(X)$ is uniformly distributed on $[s]^k$*
2. *$f(X)$ is uniformly distributed on $[s]^k$ conditioned on $f(X) = f(Y)$*
3. *for every $0 \le \epsilon \le 1$,*

$$\mathbb{P}_\epsilon(f(X) = f(Y)) \ge \frac{1}{16}\mathbb{P}_\epsilon(Y \in A | X \in A).$$

Note that the f that we produce in Theorem 2.2 satisfies stronger requirement than the one that we require in Theorem 2.1. Indeed, the f from Theorem 2.2 is uniformly distributed instead of only having a small minimum entropy. Moreover, $f(X)$ is uniformly distributed given $f(X) = f(Y)$, which means that a successful execution of the protocol will result in the two parties having uniformly random strings.

2.2 Negative results on the performance of NICD

In view of Theorems 2.1 and 2.2, the NICD problem reduces to the study of $\mathbb{P}_\epsilon(Y \in A | X \in A)$ over sets $A \subset [s]^n$ with a given cardinality. Actually, it turns out to be more convenient to normalize the cardinality instead of restricting it:

Definition 2.3. *For $A \subset [s]^n$, define*

$$M_\epsilon(A) = \frac{\ln \mathbb{P}_\epsilon(Y \in A | X \in A)}{\ln \mathbb{P}(A)}.$$

To illustrate the definition, consider the set $A = \{x : x_1 = \cdots = x_k = 0\}$, which corresponds to the trivial algorithm that selects the first k symbols. In this case, $\mathbb{P}_\epsilon(Y \in A | X \in A) = (1 - (1 - s^{-1})\epsilon))^k$. Since $\mathbb{P}(A) = s^{-k}$, it follows that

$$M_\epsilon(A) = \frac{1}{\ln s} \ln \left(\frac{1}{1 - (1 - s^{-1})\epsilon} \right). \tag{1}$$

Our main result is that the above example is optimal as $s \to \infty$.

Theorem 2.4. *For every $\delta, \epsilon > 0$ there exists $S < \infty$ such that for all $n \in \mathbb{N}$ and all $s \geq S$, any set $A \subset [s]^n$ satisfies*

$$M_\epsilon(A) \geq \frac{1}{\ln s} \Big(\ln \frac{1}{1 - \epsilon} - \delta \Big)$$

Note that since $\ln \mathbb{P}(A)$ is negative, Theorem 2.4 provides an upper bound on $\mathbb{P}_\epsilon(Y \in A \mid X \in A)$ for all sets A of a fixed probability, and therefore an upper bound on the agreement probability of any NICD protocol. We remark that our proof extends to the case where the X_i are chosen independently from some distributions whose smallest atoms are at most α. In this case, the theorem holds with s replaced by $1/\alpha$.

As a corollary of Theorems 2.1 and 2.4, we obtain a bound on the performance of any NICD protocol.

Corollary 2.5. *For any $\delta, \epsilon > 0$, there exists $S < \infty$ such that for all $n, k \in \mathbb{N}$, for any $s \geq S$, and for any NICD protocol f, g on $[s]$ with min entropy at most k, the probability that the protocol succeeds with noise ϵ is at most $(1-\epsilon)^k e^{\delta k}$.*

Since the success rate of the trivial protocol with min-entropy k is bigger than $(1-\epsilon)^k$, this shows that for large s, no protocol can be succeed with much higher probability than the trivial protocol.

Proof. Fix a protocol f, g and let A be a set such that $|A| \leq s^{n-k}$ and $\mathbb{P}_\epsilon(Y \in A | X \in A) \geq \mathbb{P}_\epsilon(f(X) = g(Y))$ (such an A exists by Theorem 2.1). Then Theorem 2.4 implies (recalling that $\ln \mathbb{P}(A)$ is negative)

$$\ln \mathbb{P}_\epsilon(Y \in A | X \in A) \leq \frac{\ln \mathbb{P}(A)}{\ln s}\Big(\log \frac{1}{1-\epsilon} - \delta\Big) \leq -k\Big(\log \frac{1}{1-\epsilon} - \delta\Big)$$

Taking the exponential of both sides yields the corollary. □

Of course, we can also restate Corollary 2.5 for a fixed probability of success and a varying k:

Corollary 2.6. *For any $\delta, \epsilon > 0$, there exists $S < \infty$ such that for all $n \in \mathbb{N}$, for all $0 < p < 1$, for any $s \geq S$, and for any NICD protocol f, g that succeeds with probability at least p, if k is the min-entropy of the protocol then the trivial protocol on $\lfloor k \frac{\log(1-\epsilon)}{\log(1-\epsilon)+\delta} \rfloor$ symbols also succeeds with probability at least p.*

In other words, for a fixed probability of failure, a trivial protocol can recover almost as many symbols as any other protocol (when s is large).

The dependence of S on δ and ϵ is not made explicit in our proof. However, our proof does provide a way to approximate $S(\delta, \epsilon)$ on a computer; therefore, we produced a plot (Figure 1) showing the approximate value of S for various values of δ and ϵ.

2.3 An example: the Hamming ball

As we have already mentioned, [1] showed that when $s = 2$, the trivial algorithm is optimal up to a constant factor; As we have just seen, this constant factor converges to 1 as $s \to \infty$. However, [1] also gave a positive result: they gave an example that achieves optimal performance (at least, up to lower order terms and for a particular range of k and ϵ). Since their example can be generalized to $s > 2$, we can examine its performance as $s \to \infty$, and compare it to the trivial algorithm.

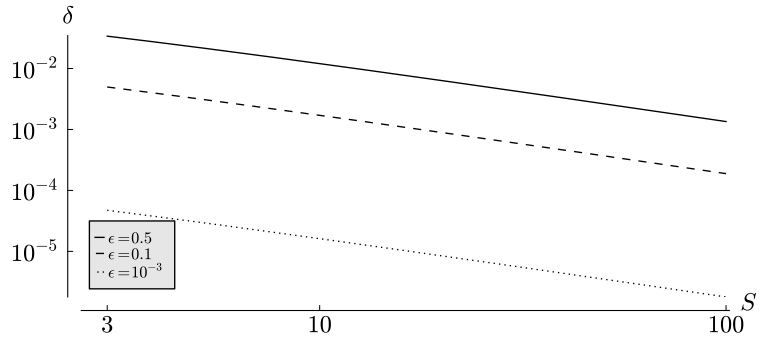

Figure 1: The relationship, in log-log scale, between S and δ in Theorem 2.4 for various values of ϵ: 0.5 (solid), 0.1 (dashed), and 10^{-3} (dotted). For each of these values of ϵ, every point (s, δ) that is above the corresponding line, and every $n \in \mathbb{N}$, all sets $A \subset [s]^n$ satisfy $M_\epsilon(A) \geq \frac{1}{\ln s}(\ln \frac{1}{1-\epsilon} - \delta)$.

Define the set

$$A_{s,\alpha,n} = \left\{ x \in [s]^n : \#\{i : x_i \neq 0\} \leq n\frac{s-1}{s} - \alpha\sqrt{n} \right\}.$$

In other words, $A_{s,\alpha,n}$ is a Hamming ball around zero of radius $n\frac{s-1}{s} - \alpha\sqrt{n}$. When $s = 2$, [1] showed that $M_\epsilon(A_{2,\alpha,n}) \approx \epsilon/2$ as $n, t \to \infty$ and $\epsilon \to 0$ (note that this does not contradict Theorem 2.4, which only holds for sufficiently large s). Since the trivial algorithm has $M_\epsilon(A) \approx \epsilon/(2 \ln 2)$ for small ϵ, this shows that the Hamming ball NICD protocol is better than the trivial one for $s = 2$. The situation reverses, however, as s grows:

Proposition 2.7. *There exists a constant c such that for any s, α and ϵ,*

$$\lim_{n\to\infty} M_\epsilon(A_{s,\alpha,n}) \geq c\epsilon.$$

Since the trivial algorithm has $M_\epsilon(A) \sim \epsilon/\ln s$, it is better than the Hamming ball protocol when s is large. In terms of the agreement probability, an argument like the proof of Corollary 2.5 shows that the agreement probability of the Hamming ball protocol is at most $(1-\epsilon)^{ck \ln s}$. In terms of the number of recovered symbols, the Hamming ball protocol with the same agreement probability as the k-symbol trivial protocol can only recover $ck/\ln s$ symbols.

3 Reduction to a single set

In this section, we will prove Theorems 2.1 and 2.2, which reduce the NICD problem to a question about optimal subsets of $[s]^n$. The proof of Theorem 2.1 is straightforward, and essentially follows directly from the Cauchy-Schwarz inequality.

Proof of Theorem 2.1. Suppose that $f, g : [s]^n \to [s]^k$ have min-entropy k. For $z \in [s]^k$, let $f_z : [s]^n \to \{0, 1\}$ be the function

$$f_z(x) = \begin{cases} 1, & \text{if } f(x) = z, \\ 0, & \text{otherwise.} \end{cases}$$

Define $g_z(x)$ similarly. Then

$$\begin{aligned} \mathbb{P}_\epsilon(f(X) = g(Y)) &= \sum_{z \in [s]^k} \mathbb{P}_\epsilon(f(X) = g(Y) = z) \\ &= \sum_{z \in [s]^k} \mathbb{E} f_z(X) g_z(Y) \\ &\le \sum_{z \in [s]^k} \sqrt{\mathbb{E} f_z(X) f_z(Y)} \sqrt{\mathbb{E} g_z(X) g_z(Y)} \\ &\le \sqrt{\sum_{z \in [s]^k} \mathbb{E} f_z(X) f_z(Y)} \sqrt{\sum_{z \in [s]^k} \mathbb{E} g_z(X) g_z(Y)}, \end{aligned}$$

where both inequalities are Cauchy-Schwarz.

For each $z \in [s]^k$, let A_z be the set $f^{-1}(z)$. Since f has min-entropy k, $|A_z| \le s^{n-k}$ for all z. Let A be the A_z which maximizes $\mathbb{P}_\epsilon[Y \in A_z \mid X \in A_z]$. Then

$$\begin{aligned} \sum_{z \in [s]^k} \mathbb{E} f_z(X) f_z(Y) &= \sum_{z \in [s]^k} \mathbb{P}_\epsilon(f(X) = f(Y) = z) \\ &= \sum_{z \in [s]^k} \mathbb{P}_\epsilon(f(X) = z) \mathbb{P}_\epsilon(Y \in A_z \mid X \in A_z) \\ &\le \mathbb{P}_\epsilon(Y \in A \mid X \in A). \end{aligned}$$

□

The idea behind Theorem 2.2 is, given a set $A \subset [s]^n$ with $\frac{1}{8} s^{n-k} \le |A| \le \frac{1}{4} s^{n-k}$, to construct a partition of $[s]^n$ out of randomly translated copies of A. Let $C \subset [s]^n$, $|C| = s^k$ be the set of "centers." We will choose C randomly; we will say how to choose it later. Let $f_C : [s]^n \to C$ to be some

function with the property that if $x \in A+c$ for a unique $c \in C$ then $f_C(x) = c$. Clearly, then,

$$\mathbb{P}_\epsilon(f_C(X) = f_C(Y)) \geq \mathbb{P}_\epsilon(\exists! c \in C \text{ such that } X, Y \in A + c). \tag{2}$$

The goal is to find a C which makes the right-hand side large; this will allow us to prove property 3 in the second part of Theorem 2.1.

Note, by the way, that it is sufficient to prove Theorem 2.2 with $[s]^k$ replaced by an arbitrary set C satisfying $|C| = s^k$. Since such a C is in bijection with $[s]^k$, the theorem as stated will follow.

Lemma 3.1. *Suppose that C is chosen (randomly) such that for any $a, b \in [s]^n$, $\mathbb{P}(a, b \in C) = s^{2(t-n)}$. Then*

$$\mathbb{E}_C \mathbb{P}_\epsilon(f_C(X) = f_C(Y)) \geq \frac{1}{16}\mathbb{P}(\epsilon(Y \in A \mid X \in A).$$

In particular, there exists a fixed C such that f_C satisfies property 3 of Theorem 2.2.

Proof. We begin from the right-hand side of (2):

$$\begin{aligned}
&\mathbb{P}_\epsilon(\exists! c \in C \text{ such that } X, Y \in A + c) &(3)\\
&\geq \mathbb{E}_C \sum_{c \in C} \mathbb{P}_\epsilon(X, Y \in A_c)\left(1 - \sum_{c' \neq c} \mathbb{P}_\epsilon(X \text{ or } Y \in A_{c'} \mid X, Y \in A_c)\right)\\
&= s^k \mathbb{E}_c \mathbb{P}_\epsilon(X, Y \in A_c)\left(1 - (s^k - 1)\mathbb{E}_{c'} \mathbb{P}_\epsilon(X \text{ or } Y \in A_{c'} \mid X, Y \in A_c)\right). &(4)
\end{aligned}$$

By our assumption on the distribution of C, $c' \neq c$ is uniformly random given c. Thus

$$\begin{aligned}
\mathbb{E}_{c'} \mathbb{P}_\epsilon(X \text{ or } Y \in A_{c'} \mid X, Y \in A_c) &\leq 2\mathbb{E}_{c'} \mathbb{P}_\epsilon(X \in A_{c'} \mid X, Y \in A_c)\\
&\leq 2\mathbb{P}_\epsilon(X \in A) \leq s^{-k}/2,
\end{aligned}$$

where the last line follows because $|A| \leq s^{n-k}/4$.

Plugging this into (4),

$$\mathbb{E}_C \mathbb{P}_\epsilon(f(X) = f(Y)) \geq \frac{s^k}{2}\mathbb{P}_\epsilon(X, Y \in A) = \frac{\mathbb{P}_\epsilon(Y \in A \mid X \in A)}{16}. \qquad \square$$

To check properties 2 and 3, we need to be a little more specific about our choice of f_C. So far, we have only assumed that $f_C(x) = c$ if c is the only member of C with $x \in A + c$. Now, take $\prec$ to be some total order

on $[s]^n$ with the property that $x \prec y$ whenever $x \in A, y \notin A$. Then define $f_C(x) = \arg\min_{c \in C}(x - c)$ (where the arg min is taken with respect to the ordering $\prec$). This defines f_C on all of $[s]^n$, and it has the property that we required before: if $f_C(x) \in A + c$ for a unique c, then $f_C(x) - c \in A$ and $f_C(x) - c' \notin A$ for every $c' \neq c$. By our requirement on $\prec$, $f_C(x) - c \prec f_C(x) - c'$ for every $c' \neq c$ and so $f_C(x) = c$.

Lemma 3.2. *If there is a subgroup $G \subset ([s]^n, +)$ and some $a \in [s]^n$ such that $C = G + a$, then f_C satisfies properties 1 and 2 of Theorem 2.2.*

Proof. For any $g \in G$,

$$f_C(x + g) = \arg\min_{c \in C}(x - (c - g)) = g + \arg\min_{c \in C - g}(x - c) = f_C(x) + g,$$

since $C - g = C$. Moreover, note that the distribution of (X, Y) is invariant under translation, in the sense that for any fixed $g \in [s]^n$, $(X, Y) + g \stackrel{d}{=} (X, Y)$. Hence,

$$\mathbb{P}(f(X) = c) = \mathbb{P}(f(X + g) = c) = \mathbb{P}(f(X) = c + g)$$

for any $c \in C, g \in G$. Since G acts transitively on C, this implies that $\mathbb{P}(f(X) = c) = 1/|C| = s^{-k}$; in other words, $f(X)$ is distributed uniformly on C.

Similarly,

$$\mathbb{P}(f(X) = f(Y) = c) = \mathbb{P}(f(X) = f(Y) = c + g)$$

for any $c \in C, g \in G$ and so $\mathbb{P}(f(X) = f(Y) = c) = s^{-k}\mathbb{P}(f(X) = f(Y))$; in other words, $f(X)$ is uniformly distributed on C conditioned on $f(X) = f(Y)$. □

Proof of Theorem 2.2. To prove Theorem 2.2, we need to find a set C which satisfies the hypotheses of Lemmas 3.1 and 3.2. In [1], they chose C to be a uniformly random k-dimensional affine subspace of $[2]^n$, but since $[s]^n$ is not a vector space for every s, we will need something slightly more complicated.

Let $s = \prod_{i=1}^m p_i^{j_i}$ be the prime factorization of s. By the Chinese remainder theorem, the group $([s]^n, +)$ is isomorphic to $\bigoplus_{i=1}^m ([p_i]^{nj_i}, +)$; let $\phi : \bigoplus_i ([p_i]^{nj_i}, +) \to [s]^n$ be an isomorphism. Independently for each $i = 1, \ldots, m$ and $j = 1, \ldots, k_i$, let $G_{i,j}$ be a uniformly random k-dimensional subspace of $[p_i]^n$ (which is a vector space), and let $a_{i,j}$ be a uniformly random element of $[p_i]^n$. Finally, define

$$C = \phi\Big(\bigoplus_{i,j}(a_{i,j} + G_{i,j})\Big) = \phi\Big(\bigoplus_{i,j} a_{i,j}\Big) + \phi\Big(\bigoplus_{i,j} G_{i,j}\Big).$$

Since $\phi(\bigoplus_{i,j} G_{i,j})$ is a subgroup of $[s]^n$, the condition of Lemma 3.2 is satisfied with probability 1.

To check the condition of Lemma 3.1, note that for any $b = \bigoplus_{i,j} b_{i,j}$ and $c = \bigoplus_{i,j} c_{i,j}$ in $\bigoplus_{i=1}^{m} [p_i]^{nk_i}$,

$$\mathbb{P}(b_{i,j}, c_{i,j} \in a_{i,j} + G_{i,j}) = p_i^{2(n-k)}$$

because $G_{i,j}$ is a uniformly random k-dimensional subspace of $[p_i]^n$. Since the $a_{i,j}$ and $G_{i,j}$ are independent, it follows that

$$\mathbb{P}(\phi(b), \phi(c) \in C) = \prod_{i,j} P(a_{i,j}, b_{i,j} \in C_{i,j}) = s^{2(n-k)}.$$

That is, the distribution of C satisfies the condition of Lemma 3.1. In particular, there exists a non-random C' that belongs to the support of C, and which also satisfies condition 3 of Theorem 2.2. By the previous paragraph, the fact that it belongs to the support of C implies that it also satisfies conditions 1 and 2. □

4 An upper bound on agreement

The proof of Theorem 2.4 uses a hypercontractive inequality in much the same way as it was used in [1]. The difference here is that [1] used only the hypercontractive inequality over the two-point space with the uniform measure, while we need one that applies to spaces with more than two points. Before stating this hypercontractive inequality, we need to define the appropriate Bonami-Beckner-type operator: for a function $g : [s] \to \mathbb{R}$, and some $0 < \tau < 1$, define $S_\tau g = \tau g + (1-\tau)\mathbb{E}g$. Thus, for any $0 < \tau < 1$, and any $1 \le p, q \le \infty$, S is an operator $L_p([s]) \to L_q([s])$. We define $T_\tau : L_p([s]^n) \to L_q([s])^n$ by $T_\tau = S_\tau^{\otimes n}$. The operator T_τ can also be written in terms of the Fourier expansion of f; see [10] for details. For us, the crucial property of T_τ is that

$$\mathbb{E}_\epsilon f(X) f(Y) = \mathbb{E}(T_\tau f)^2 \tag{5}$$

when $\tau = \sqrt{1-\epsilon}$. This fact was used in [1] for $s = 2$ to establish Theorem 2.4 in that case.

The following hypercontractive inequality is due to Oleszkiewicz [8]:

Theorem 4.1. *Fix $s \in \mathbb{N}$ and set $\alpha = \frac{1}{s}$, $\beta = 1 - \alpha$. Define*

$$\sigma(\alpha, p) = \left(\frac{\beta^{2-2/p} - \alpha^{2-2/p}}{\alpha^{1-2/p}\beta - \beta^{1-2/p}\alpha} \right)^{1/2}.$$

Then for any $f : [s]^n \to \mathbb{R}$, if $\tau \le \sigma(\alpha, p)$ then

$$\|T_\tau f\|_2 \le \|f\|_p.$$

We remark that the reason for not having an explicit $S(\delta)$ in Theorem 2.4 and its corollaries is that we do not know how to solve for p in terms of $\sigma(\alpha, p)$. However, an approximate solution can easily be found on a computer, and we used such an approximation to produce Figure 1. To obtain Theorem 2.4, it suffices to study the limit of $\sigma(\alpha, p)$ as $\alpha \to 0$. Essentially, $\sigma^2(\alpha, p) \approx \alpha^{1-2/p}$ for small α, and so if we take p to be slightly larger than what is needed to solve $\alpha^{1-2/p} = 1 - \epsilon$, then we will have $\sigma(\alpha, p) \ge 1 - \epsilon$. This will allow us to apply Theorem 4.1 with $\tau = \sqrt{1-\epsilon}$.

Lemma 4.2. *Let $p = p(\alpha, \delta, \epsilon)$ solve*

$$\alpha^{(2/p-1)-\delta/\ln\alpha} = 1 - \epsilon.$$

Then for any $\delta > 0$ and $\epsilon^* \in (0,1)$, there is an $A(\delta, \epsilon^*) > 0$ such that $\alpha < A(\delta, \epsilon^*)$ implies that for all $\epsilon \in (0, \epsilon^*)$,

$$\sigma^2(\alpha, p(\alpha, \delta, \epsilon)) \ge 1 - \epsilon.$$

Proof. Note that the definition of p ensures that $p < 2$ for all α, δ, ϵ. By the definition of σ,

$$\sigma^2(\alpha, p)\alpha^{1-2/p} = \frac{\beta^{2-2/p} - \alpha^{2-2/p}}{\beta - \alpha^{2/p}\beta^{1-2/p}} \ge \beta^{2-2/p} - \alpha^{2-2/p}. \tag{6}$$

Fix ϵ^* and δ, and note that as $\alpha \to 0$, $2 - 2/p \to 1$ uniformly for all $\epsilon \in (0, \epsilon^*)$. Hence, the right-hand side of (6) converges to 1 (uniformly in ϵ) as $\alpha \to 0$. Plugging in the definition of p,

$$\frac{\sigma^2(\alpha, p)}{1-\epsilon} = \sigma^2(\alpha, p)\alpha^{1-2/p}\alpha^{-\delta/\ln\alpha} \ge (1 - o(1))e^{-\delta}.$$

In particular, the limit of the right hand side is strictly smaller than one, and so $\sigma^2(\alpha, p) \ge 1 - \epsilon$ for sufficiently small α. □

Proof of Theorem 2.4. Fix $\epsilon, \delta > 0$. Let A and p be as in Lemma 4.2 and define $S = 1/A$. If $s \ge S$ then $\alpha = 1/s \le A$ and so Lemma 4.2 implies that $\sigma^2(\alpha, p) \ge 1 - \epsilon$. Thus, (5) and Theorem 4.1 imply that

$$\mathbb{P}_\epsilon(X, Y \in A) = \|T_{\sqrt{1-\epsilon}} 1_A\|_2^2 \le \|1_A\|_p^2 = \mathbb{P}(A)^{\frac{2}{p}}.$$

Hence, $\mathbb{P}_\epsilon(Y \in A | X \in A) \le \mathbb{P}(A)^{2/p-1}$. Taking the logarithm and dividing by $\ln \mathbb{P}(A)$ (which is negative), we have

$$M_\epsilon(A) = \frac{\ln \mathbb{P}_\epsilon(X, Y \in A)}{\ln \mathbb{P}(A)} \ge \frac{2}{p} - 1 = \frac{\ln \frac{1}{1-\epsilon}}{\ln s} - \frac{\delta}{\ln s}. \qquad \square$$

5 Hamming ball

In this section, we consider the example of the Hamming ball $A_{s,\alpha,n}$ consisting of $x \in [s]^n$ such that $\#\{i : x_i = 0\} \le \frac{n}{s} - \alpha\sqrt{n}$. This is an interesting example because [1] showed that if α is sufficiently large (depending on ϵ), then as $n \to \infty$, $A_{2,\alpha,n}$ achieves the upper bound of Theorem 2.4. We will show, however, that this is no longer true for large s.

Note that $1_{X_1=0}$ has mean $\frac{1}{s}$ and variance $\frac{s-1}{s^2}$. Thus, the Berry-Esséen theorem implies that for any fixed α and s,

$$\mathbb{P}(A_{s,\alpha,n}) \to \mathbb{P}\Big(Z \le -\frac{\alpha s}{\sqrt{s-1}}\Big) \tag{7}$$

as $n \to \infty$, where $Z \sim \mathcal{N}(0,1)$. Moreover, if $(Z_1, Z_2) \sim \mathcal{N}(0, \left(\begin{smallmatrix} 1 & 1-\epsilon \\ 1-\epsilon & 1 \end{smallmatrix}\right))$ then

$$\mathbb{P}_\epsilon(X, Y \in A_{s,\alpha,n}) \to \mathbb{P}\Big(Z_1, Z_2 \le -\frac{\alpha s}{\sqrt{s-1}}\Big). \tag{8}$$

In particular, by studying normal probabilities we can use (7) and (8) to compute $\lim_{n\to\infty} M_\epsilon(A_{s,\alpha,n})$.

Lemma 5.1. *Suppose that $(Z_1, Z_2) \sim \mathcal{N}(0, \left(\begin{smallmatrix} 1 & 1-\epsilon \\ 1-\epsilon & 1 \end{smallmatrix}\right))$. There is a sufficiently small constant c such that for all $t > 0$ and $0 < \epsilon < 1$,*

$$\mathbb{P}(Z_1 \ge t \mid Z_2 \ge t) \le \mathbb{P}(Z_1 \ge t)^{c\epsilon}.$$

Lemma 5.1 has the following immediate consequence for $M_\epsilon(A_{s,\alpha,n})$:

Corollary 5.2. *There exists a constant c such that for any s and α,*

$$\lim_{n\to\infty} M_\epsilon(A_{s,\alpha,n}) \ge c\epsilon.$$

By comparison, the trivial protocol $A = \{x : x_1 = \cdots = x_k = 0\}$ has

$$M_\epsilon(A) = \frac{1}{\ln s} \ln\left(\frac{1}{1-(1-s^{-1})\epsilon}\right) \le \frac{C'\epsilon}{\ln s}.$$

In particular, for a fixed success probability and a sufficiently large alphabet s, the trivial protocol recovers $c \ln s$ times as many symbols as the Hamming ball protocol.

Proof of Corollary 5.2. According to (7) and (8),

$$M_\epsilon(A_{s,\alpha,n}) \to \frac{\log \mathbb{P}\Big(Z_1 \le -\frac{\alpha s}{\sqrt{s-1}} \Big| Z_2 \le -\frac{\alpha s}{\sqrt{s-1}}\Big)}{\log \mathbb{P}\Big(Z_1 \le -\frac{\alpha s}{\sqrt{s-1}}\Big)}.$$

Now apply Lemma 5.1 to the numerator (recalling that the denominator is negative):

$$\lim M_\epsilon(A_{s,\alpha,n}) \ge \frac{\log \mathbb{P}\Big(Z_1 \le -\frac{\alpha s}{\sqrt{s-1}}\Big)^{c\epsilon}}{\log \mathbb{P}\Big(Z_1 \le -\frac{\alpha s}{\sqrt{s-1}}\Big)} = c\epsilon. \qquad \square$$

Proof of Lemma 5.1. The proof makes use of the Ornstein-Uhlenbeck semigroup P_t, defined by

$$(P_\tau f)(x) = \mathbb{E} f(e^{-\tau}x + \sqrt{1-e^{-2\tau}}Z),$$

where $Z \sim \mathcal{N}(0,1)$. The Nelson-Gross [2, 7] hypercontractive inequality states that

$$\big(\mathbb{E} P_\tau |f(Z)|^q\big)^{1/q} \le \big(\mathbb{E}|f(Z)|^p\big)^{1/p} \tag{9}$$

whenever $q \le 1 + e^{2\tau}(p-1)$. If we set $f(x) = 1_{x \ge t}$ and $\tau = -\log(1-\epsilon)$, then

$$\mathbb{P}(Z_1, Z_2 \ge t) = \mathbb{E} f(Z_1) f(Z_2) = \mathbb{E} f(Z) P_\tau f(Z) = \mathbb{E}(P_{\tau/2} f(Z))^2.$$

Thus, (9) with $q = 2$ and $p = 1 + e^{-2\tau} = 1 + (1-\epsilon)^2$ implies that

$$\mathbb{P}(Z_1, Z_2 \ge t) \le (\mathbb{E} f(Z))^{\frac{2}{1+(1-\epsilon)^2}} = \mathbb{P}(Z_2 \ge t)^{\frac{2}{1+(1-\epsilon)^2}} \le \mathbb{P}(Z_2 \ge t)^{1+c\epsilon}.$$

Hence,

$$\mathbb{P}(Z_1 \ge t | Z_2 \ge t) \le \frac{\mathbb{P}(Z_1 \ge t, Z_2 \ge t)}{\mathbb{P}(Z_2 \ge t)} \le \mathbb{P}(Z_2 \ge t)^{c\epsilon}. \qquad \square$$

References

[1] A. Bogdanov and E. Mossel. On extracting common random bits from correlated sources. *IEEE Transactions on information theory*, 57(10):6351–6355, 2011. Arxiv 1007.2135.

[2] Leonard Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[3] D. Lim, J.W. Lee, B. Gassend, G.E. Suh, M. Van Dijk, and S. Devadas. Extracting secret keys from integrated circuits. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 13(10):1200–1205, 2005.

[4] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality (extended abstract). In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*, pages 21–30. IEEE Computer Society, 2005.

[5] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov. Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality. *Israel J. Math.*, 154:299–336, 2006.

[6] E. Mossel, K. Oleszkiewicz, and A. Sen. On reverse hypercontractivity. 2011.

[7] Edward Nelson. The free Markoff field. *J. Functional Analysis*, 12:211–227, 1973.

[8] K. Oleszkiewicz. On a nonsymmetric version of the Khinchine-Kahane inequality. *Progress In Probability*, 56:156–168, 2003.

[9] Y. Su, J. Holleman, and B.P. Otis. A digital 1.6 pJ/bit chip identification circuit using process variations. *Solid-State Circuits, IEEE Journal of*, 43(1):69–77, 2008.

[10] P. Wolff. Hypercontractivity of simple random variables. *Studia Mathematica*, pages 219–326, 2007.

[11] Ke Yang. On the (im)possibility of non-interactive correlation distillation. *Theoretical Computer Science*, 382(2):157–166, 2007.

[12] H. Yu, P.H.W. Leong, H. Hinkelmann, L. Moller, M. Glesner, and P. Zipf. Towards a unique FPGA-based identification circuit using process variations. In *19th International Conference on Field Programmable Logic and Applications*, pages 397–402. IEEE, 2009.