Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 57 (2015)

Simplified Separation of Information and Communication

Anup Rao* Makrand Sinha†

April 13, 2015

Abstract

We give an example of a boolean function whose information complexity is exponentially smaller than its communication complexity. Our result simplifies recent work of Ganor, Kol and Raz [GKR14a, GKR14b].

1 Introduction

A fundamental question in the study of communication complexity is whether the information complexity of a communication problem is the same as its communication complexity. If the messages in a protocol reveal a small amount of information, does that mean that the protocol can be simulated using few bits of communication? When the protocol is deterministic and one-way, a precise answer was given by Shannon [Sha48]. He defined the notion of the *entropy*, $H(M)$, to measure the information content of a message M, and showed that the number of bits of communication can always be made at most $H(M)+1$ in expectation, which is tight.

For randomized and interactive (multi-round) protocols, this question was made explicit in a sequence of works. Chakrabarti, Shi, Wirth and Yao [CSWY01] defined what we now call the *external information cost* of a protocol, which measures the information learned about the inputs by an external observer of the messages. If M denotes the messages, R denotes the shared randomness and X, Y the inputs, the external information cost is defined to be the mutual information $\mathbf{I}(XY : M|R)$. Barak, Braverman, Chen and Rao [BBCR13] defined the *internal information cost* of a protocol as $\mathbf{I}(X : M|YR) + \mathbf{I}(Y : M|XR)$, the information learned by the parties about each others' inputs. The internal information cost is never larger than the external information cost.

For bounded-round protocols, Harsha, Jain, McAllester and Radhakrishnan [HJMR10] (see also [BG14]) showed how to give optimal simulations in terms of their external information cost, and Braverman and Rao [BR11] gave near optimal simulations in terms of internal information cost (upto a $1+o(1)$ factor). However, many of the best known simulations for interactive protocols are not known to be optimal. We know how to simulate any interactive protocol with external information cost I and communication C using a protocol with communication $O(I \log^2 C)$ [BBCR13].

*Computer Science and Engineering, University of Washington, anuprao@cs.washington.edu. Supported by an Alfred P. Sloan Fellowship, the National Science Foundation under agreement CCF-1016565, an NSF Career award, and by the Binational Science Foundation under agreement 2010089.

†Computer Science and Engineering, University of Washington, makrand@cs.washington.edu. Supported by the National Science Foundation under agreement CCF-1016565, and by the Binational Science Foundation under agreement 2010089.

ISSN 1433-8092

We also know how to simulate a protocol with internal information I and communication C using a protocol with communication $O(\sqrt{IC}\log C)$ [BBCR13]. Braverman [Bra12] showed how to simulate any protocol with internal information cost I using communication $2^{O(I)}$. Very recently, Ramamoorthy and Rao [RR15] gave better simulations when one party learns much less than the other.

These results are closely tied to communication lower bounds. Information theory based methods for proving lower bounds on the communication complexity of disjointness [KS92, Raz92, BYJKS02] can be seen as precursors to some of them. They have been used to give answers to longstanding questions like the direct sum [BBCR13] and direct product [BRWY13] questions in communication complexity.

Braverman and Weinstein [BW12] (see also [KLL$^+$12]) showed that any boolean function $f(x, y)$ that can be computed with internal information cost I must have a (nearly) monochromatic rectangle (namely a subset $R = S \times T$ of the inputs where the function is essentially constant), and so large discrepancy. This means that upper bounds on the size of monochromatic rectangles cannot be used to prove lower bounds on information complexity[1]. So for a long time, all known methods for proving lower bounds for communication failed to prove lower bounds on functions that have large (0 and 1) monochromatic rectangles. This pointed to a significant weakness in our ability to prove new lower bounds in communication complexity, since there are certainly functions with high communication complexity that do have large monochromatic rectangles. One can plant large monochromatic rectangles into a random function to obtain such an example with high probability.

In a remarkable sequence of papers, Ganor, Kol and Raz [GKR14a, GKR14b] showed that there is a function with internal information cost I that requires $2^{\Omega(I)}$ communication. This proved that Braverman's simulation [Bra12] is tight. Their proof gives a method to prove communication lower bounds on functions that have many large monochromatic rectangles, potentially leading to fundamentally different methods to prove lower bounds on communication problems. Subsequently, Fontes et. al. [FJK$^+$15] showed that the techniques used in [GKR14b] can not be used to separate information and communication in the non-distributional setting as defined in [Bra12].

Building on the work of [GKR14b], we give a new proof of their main result. Our proofs are shorter, and we find them more intuitive. We use the notion of a *fooling distribution* to prove the communication lower bounds. We define a distribution on inputs $p(x, y)$, two disjoint equi-probable events $\mathcal{E}_0$, $\mathcal{E}_1$, and a boolean function $h(x, y)$, such that $h(x, y) = b$ when the inputs are sampled conditioned on $\mathcal{E}_b$. The inputs to the communication protocol are sampled conditioned on the event $\mathcal{E} = \mathcal{E}_0 \vee \mathcal{E}_1$. We show that there is a protocol with internal information cost $O(\log k)$ that computes h on this distribution, yet no protocol with communication significantly smaller than k can compute the same function. The communication lower bound is proved by showing that if the length of the messages m computed by a protocol operating on these inputs is much less than k, then $p(m|\mathcal{E}_0) \approx p(m) \approx p(m|\mathcal{E}_1)$, where $\approx$ denotes closeness of the distributions in statistical distance. This proves that the communication complexity of h is close to k, since if the protocol computes h, the support of $p(m|\mathcal{E}_0)$ must be nearly disjoint from the support of $p(m|\mathcal{E}_1)$. The key concept introduced in [GKR14b] to prove lower bounds is the notion of the *relative discrepancy*. Low relative discrepancy does imply the existence of a fooling distribution, but the converse does not appear to be true (more details in Appendix C).

Next we define the function for which we give a separation on information and communication,

[1]The information based methods for proving lower bounds on disjointness also prove that disjointness does not have large 1-monochromatic rectangles.

Probability space: $J \in [n]$ is uniformly random. $X, Y : [k]^{<n} \to [k]$ are sampled uniformly at random subject to the constraint that for any $z \in [k]^{<J}$, $X(z) = Y(z)$. $F, G : [k]^n \to \{0,1\}$ are uniformly random.

Events $\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}$: Let $\mathcal{E}_0$ denote the event that for all consistent z, $X(z) = Y(z)$ and $F(z) = G(z)$ (when $|z| = n$). Let $\mathcal{E}_1$ denote the event that for all consistent z, $X(z) = Y(z)$ and $F(z) \neq G(z)$ (when $|z| = n$). Note that $\mathcal{E}_0$ and $\mathcal{E}_1$ are disjoint, and are equally likely. Let $\mathcal{E} = \mathcal{E}_0 \vee \mathcal{E}_1$.

Input distribution: J, X, Y, F, G are sampled conditioned on the event $\mathcal{E}$.

Figure 1: Distribution $p(j, x, y, f, g)$ for the k-ary pointer jumping problem

and give a high level sketch of the proof of the communication lower bound.

2 k-ary Pointer Jumping

For a parameter k, we work with the alphabet $[k] = \{1, 2, \ldots, k\}$. Let $X, Y : [k]^{<n} \to [k]$ be functions mapping strings of length less than n to a single character, and let $J \in [n]$ be a number. Let $z_{<j}$ denote the prefix of z of length $j-1$. Let $F, G : [k]^n \to \{0,1\}$ be boolean functions.

Definition 1. *$z \in [k]^{<n}$ is* consistent *with X, Y, J, if $|z| \geq J$, and $X(z_{<J}) + Y(z_{<J}) = z_J \mod k$.*

In the k-ary pointer jumping problem, two parties are given (X, F) and (Y, G) and need to communicate to compute $F(z) + G(z) \mod 2$ at *any* consistent z. The distribution on inputs, described in Figure 1, ensures that $F(z) + G(z) \mod 2$ is the same for *every* consistent z.

There is a trivial protocol for this problem that has communication $O(n \log k)$: in each step Alice and Bob send each other $X(z_{<r}), Y(z_{<r})$ and set z_r so that $z_r = X(z_{<r}) + Y(z_{<r}) \mod k$. Even though neither party knows the value of J, they compute a consistent z, and so can compute $F(z)+G(z) \mod 2$ with two more bits of communication. We prove that there is a low information solution for this task, but no low communication solution. Setting $n = 2^k$, we exhibit a correct protocol with information cost $O(\log k)$ (Theorem 2), even though no protocol with communication $\Omega(k)$ can succeed (Theorem 3).

The low information protocol for the problem is quite similar to the trivial protocol. In each step, the parties send each other the value $X(z_{<r}), Y(z_{<r})$ with probability $1 - \frac{1}{\log n}$ and send a uniformly random value otherwise. The parties abort the protocol if they experience $\frac{\log n}{\log \log n}$ rounds where the messages they sent were not the same. The distribution on inputs ensures that they will sample a consistent z with high probability. When the parties sample a consistent z, the messages sent are almost always sampled from a distribution that the receiving party knows, while if they sample a z that is not consistent, the protocol aborts shortly after the inconsistency. These properties can be used to show that the information cost of the protocol is small. In Section 5, we show:

Theorem 2. *The internal information cost of the k-ary pointer jumping problem is at most $O(\log(k \log n) \cdot 2^{\frac{2 \log n}{k}})$.*

To prove the communication lower bound, consider any protocol with ℓ bits of communication that solves the k-ary pointer jumping problem. Without loss of generality, we may assume that the protocol is deterministic, since any randomness can always be fixed to obtain a deterministic protocol that succeeds with high probability. Let M denote the messages of the protocol. If the protocol solved the k-ary pointer jumping problem, then the statistical distance between $p(m|\mathcal{E}_0)$ and $p(m|\mathcal{E}_1)$ would be close to 1, since these distributions would have nearly disjoint supports. We prove a lower bound on the communication by showing that if ℓ is small, then the distance between these distributions is close to 0. It will be convenient to state our results in terms of the function $\eta : [0, \infty) \to [0, 1]$ defined as

$$\eta(\alpha) = \begin{cases} 0 & \text{if } \alpha = 0, \\ \alpha \log(1/\alpha) & \text{if } \alpha \in (0, 1/e), \\ \frac{\log e}{e} & \text{if } \alpha \geq 1/e. \end{cases} \tag{1}$$

One can check that η is non-decreasing, continuous, and concave. We prove:

Theorem 3. *If the protocol has communication complexity ℓ, then $p(m|\mathcal{E}_0) \overset{\gamma}{\approx} p(m) \overset{\gamma}{\approx} p(m|\mathcal{E}_1)$, with $\gamma = 4(2e\ell/k + 2\ell\sqrt{2^\ell/n} + \eta(\sqrt{2^\ell/n}))^{1/3}$.*

Theorem 3 implies that $\ell = \Omega(\min\{k, \log n\})$ for any protocol that solves the k-ary pointer jumping problem for the given distribution, which implies that $\ell = \Omega(k)$, if we choose $n = 2^k$.

2.1 High-level Proof Sketch for Theorem 3

The low information protocol, described above, works by computing a consistent string with high probability. The first step of the lower bound proof is to show that neither player can can learn much information about which strings are consistent, *when the inputs are sampled from p*. Let S be the set of consistent strings $z \in [k]^{\leq n}$, and let $X_{\leq J}, Y_{\leq J}$ denote the restriction of X, Y to inputs of length at most J. Let X_J, Y_J denote the restriction of X, Y to inputs of length J. Then we prove:

Lemma 4.

$$\left.\begin{aligned} \mathbf{I}\left(M : S | Y_{\leq J} J\right) &\leq \mathbf{I}\left(M : X_J | Y_{\leq J} J\right) \\ \mathbf{I}\left(M : S | X_{\leq J} J\right) &\leq \mathbf{I}\left(M : Y_J | X_{\leq J} J\right) \end{aligned}\right\} \leq \frac{2^\ell}{n}.$$

These equations bound the information learned by each party about S. Lemma 4 is proved via a somewhat subtle application of the chain rule. The proof is delicate because J is essentially determined by X, Y. The bound we obtain here is more or less tight. The parties can use $O(\ell)$ bits of communication and hashing to compute a set of size $n/2^\ell$ that contains J. If the parties then reveal one bit of information about a random coordinate in this set, the information revealed about X_J will be $\Omega(2^\ell/n)$. Lemma 4 does not yet complete the proof, because the protocol may learn a lot of information when the inputs are conditioned on the event $\mathcal{E} = \mathcal{E}_0 \vee \mathcal{E}_1$.

Next, we show that the parties do not learn much information about the values of X, F, Y, G restricted to S (denoted X_S, F_S, Y_S, G_S). Since for every z, $p(z \in S | x_{\leq j} j) \leq 1/k$, one might hope that $\mathbf{I}\left(M : X_S F_S\right)$ can be bounded by $\mathbf{I}\left(M : XF\right)/k \leq \ell/k$. In fact, if S was independent of M, X, F, such a bound would be easy to prove using the chain rule for information. We prove a generalization of Shearer's Lemma [CGFS86, Rad03], which allows us to make such a claim even

when M and S are dependent. This lemma may be of independent interest. Below U_S denotes the restriction of U to the coordinates in S. We show[2]:

Lemma 5. *Suppose $U = U_1, \ldots, U_t$ are mutually independent, $C \in \{0,1\}^\ell, S \subseteq [t]$, and V are such that U is independent of SV, $U - C - SV$ and for all $i \in [t]$, $p(i \in S) \leq 1/k$. Then*

$$\mathbf{I}\left(C : U_S | VS\right) \leq \ell \cdot \left(\frac{2e}{k} + 2\sqrt{\mathbf{I}\left(C : S\right)}\right) + \eta\left(\sqrt{\mathbf{I}\left(C : S\right)}\right).$$

Conditioned on any fixing of $X_{\leq J} Y_{\leq J} J$, we have that XF and YG are independent, and since M are the messages in a communication protocol, $XF - M - YG$ holds (see Proposition 17). Thus Lemmas 4, 5, and convexity can be used to show:

$$\left.\begin{array}{l}\mathbf{I}\left(M : X_S F_S | X_{\leq J} Y_{\leq J} J\right) \\ \mathbf{I}\left(M : Y_S G_S | X_{\leq J} Y_{\leq J} J\right)\end{array}\right\} \leq 2e\ell/k + 2\ell\sqrt{2^\ell/n} + \eta\left(\sqrt{2^\ell/n}\right). \tag{2}$$

The events $\mathcal{E}_0$ and $\mathcal{E}_1$ both assert that two random variables are equal — $\mathcal{E}_0$ is the event that $X_S F_S = Y_S G_S$, and $\mathcal{E}_1$ is the event that $X_S F_S = Y_S, \overline{G}_S$, where $\overline{G}$ is the function $1 - G$. To complete the proof, we show:

Lemma 6. *If A, B are uniform and independent, and $A - C - B$, then $p(c) \overset{\epsilon}{\approx} p(c|a = b)$, with $\epsilon = 2\mathbf{I}\left(C : A\right)^{1/3} + 2\mathbf{I}\left(C : B\right)^{1/3}$.*

Lemma 6 together with (2) and another convexity argument completes the proof of Theorem 3.

2.2 Organization

Following the preliminaries, we give a detailed proof of Theorem 3 in Section 4, assuming the 3 lemmas described in the proof overview. In Section 4.1 we prove the three lemmas. In Section 5, we bound the information complexity of the k-ary pointer jumping problem.

3 Preliminaries

3.1 Probability Spaces and Variables

Unless otherwise stated, logarithms in this text are computed base two. Random variables are denoted by capital letters (e.g. A) and values they attain are denoted by lower-case letters (e.g. a). Events in a probability space will be denoted by calligraphic letters (e.g. $\mathcal{E}$). Given $a = a_1, a_2, \ldots, a_n$, we write $a_{\leq i}$ to denote $a_1, \ldots, a_i$. We define $a_{<i}$ similarly. We write a_S to denote the projection of a to the coordinates specified in the set $S \subseteq [n]$. $[k]$ to denotes the set $\{1, 2, \ldots, k\}$, and $[k]^{<n}$ denotes the set of all strings of length less than n over the alphabet $[k]$, including the empty string. $|z|$ denotes the length of the string z.

We use the notation $p(a)$ to denote both the distribution on the variable a, and the number $\Pr_p[A = a]$. The meaning will be clear from context. We write $p(a|b)$ to denote either the distribution of A conditioned on the event $B = b$, or the number $\Pr[A = a | B = b]$. Given a distribution $p(a, b, c, d)$, we write $p(a, b, c)$ to denote the marginal distribution on the variables a, b, c (or the corresponding probability). We often write $p(ab)$ instead of $p(a, b)$ for conciseness of notation. If $\mathcal{W}$ is an event, we write $p(\mathcal{W})$ to denote its probability according to p. We denote by $\mathbb{E}_{p(a)}\left[g(a)\right]$ the expected value of $g(a)$ in p. We write $A - M - B$ to assert that $p(amb) = p(m) \cdot p(a|m) \cdot p(b|m)$.

[2]Recall that $A - C - B$ means that $p(acb) = p(c) \cdot p(a|c) \cdot p(b|c)$; in words, after fixing C, A and B are independent.

3.2 Statistical Distance

For two distributions p, q, the statistical distance $|p(a) - q(a)|$ between them is defined to be $|p(a) - q(a)| = \max_Q (p(a \in Q) - q(a \in Q))$. We say p and q are ϵ-close if $|p - q| \leq \epsilon$ and write $p \overset{\epsilon}{\approx} q$.

Proposition 7. *If $p(ab), q(ab)$ are such that $p(a) = q(a)$, then $|p(b) - q(b)| = \mathbb{E}_{p(a)} [|p(b|a) - q(b|a)|]$.*

3.3 Divergence and Mutual Information

The *divergence* between p, q is defined to be $\dfrac{p(A)}{q(A)} = \sum_a p(a) \log \frac{p(a)}{q(a)}$. For three random variables A, B, C with underlying probability distribution $p(a, b, c)$, and an event $\mathcal{E}$ in the same probability space, we will use the shorthand $\dfrac{A|bc\mathcal{E}}{A|c} = \dfrac{p(A|bc\mathcal{E})}{p(A|c)}$, when p is clear from context. The *mutual information* between A, B conditioned on C is defined as

$$\mathbf{I}(A : B|C) = \mathop{\mathbb{E}}_{c,b} \left[\frac{A|bc}{A|c} \right] = \mathop{\mathbb{E}}_{c,a} \left[\frac{B|ac}{B|c} \right] = \sum_{a,b,c} p(abc) \log \frac{p(a|bc)}{p(a|c)}.$$

3.4 Basic Divergence Facts

The proofs of the following basic facts can be found in [CT06]:

Proposition 8. *If $A \in \{0,1\}^\ell$, then $\mathbf{I}(A : B) \leq \ell$.*

Proposition 9 (Chain Rule)**.** *If $a = a_1, \ldots, a_s$, then $\dfrac{p(A)}{q(A)} = \sum_{i=1}^{s} \mathbb{E}_{p(a)} \left[\dfrac{p(A_i|a_{<i})}{q(A_i|a_{<i})} \right]$.*

Proposition 10 (Pinsker's Inequality)**.** $|p(a) - q(a)|^2 \leq \dfrac{p(A)}{q(A)}$.

Proposition 11. $\dfrac{p(A)}{q(A)} \geq 0$.

3.5 Divergence Inequalities

The following propositions bound the change in divergence when extra conditioning is involved. These were proved in [BRWY13, GKR14b]. For completeness, we give full proofs of all of them in Appendix A.

Proposition 12. *For an event $\mathcal{W}$ and variables A, M, $\mathbb{E}_{m|\mathcal{W}} \left[\dfrac{A|m\mathcal{W}}{A} \right] \leq \log \frac{1}{p(\mathcal{W})} + \mathbf{I}(A : M|\mathcal{W})$.*

Proposition 13. $\mathbb{E}_{p(b)} \left[\dfrac{p(A|b)}{q(A)} \right] \geq \dfrac{p(A)}{q(A)}$.

Proposition 14. $\mathbb{E}_{p(b)}\left[\frac{p(A|b)}{p(A)}\right] \leq \mathbb{E}_{p(b)}\left[\frac{p(A|b)}{q(A)}\right]$.

Proposition 15. *Let p, q be distributions on bits. Then $\frac{p}{q} \geq p(1)\log\frac{p(1)}{e\cdot q(1)}$.*

3.6 Communication Complexity

We briefly describe basic properties of communication protocols that we need. For more details see the book by Kushilevitz and Nisan [KN97]. The *communication complexity* of a protocol is the maximum number of bits that may be exchanged by the protocol. For a deterministic protocol π, let $\pi(x, y)$ denote the messages of the protocol on inputs x, y and define events,

$$\mathcal{S}_m = \{x|\exists y \text{ such that } \pi(x,y) = m\}, \qquad \mathcal{T}_m = \{y|\exists x \text{ such that } \pi(x,y) = m\}.$$

Proposition 16 (Messages Correspond to Rectangles)**.** $\pi(x, y) = m \iff x \in \mathcal{S}_m$ *and* $y \in \mathcal{T}_m$.

Proposition 16 implies:

Proposition 17 (Markov Property of Protocols)**.** *Let X and Y be inputs to a deterministic communication protocol with messages M. If X and Y are independent then $X - M - Y$.*

4 Communication lower bound

We shall prove that $p(m) \overset{\gamma}{\approx} p(m|\mathcal{E}_0)$. The bound for the event $\mathcal{E}_1$ is analogous. We first give the proof assuming Lemmas 6, 4, and 5. Then we prove the lemmas.

By Lemma 4, $\mathbf{I}(M : Y_J S|X_{\leq J}J) = \mathbf{I}(M : Y_J|X_{\leq J}J) \leq 2^\ell/n$. After fixing $x_{\leq j}, j$, S is determined by Y_J. For any such fixing, we have $p(z \in S|x_{\leq j}j) \leq 1/k$, XF is independent of YG and (by Proposition 17), $XF - M - SY_J$. Thus we can apply Lemma 5 to conclude that

$$\mathbf{I}(M : X_S F_S|SY_J x_{\leq j}j) \leq 2e\ell/k + 2\ell\sqrt{\mathbf{I}(M : Y_J S|x_{\leq j}j)} + \eta\left(\sqrt{\mathbf{I}(M : Y_J S|x_{\leq j}j)}\right).$$

Taking the expectation over the choice of $x_{\leq j}j$, and using the concavity of the square-root and η:

$$\begin{aligned}\mathbf{I}(M : X_S F_S|Y_{\leq J}X_{\leq j}J) &\leq 2e\ell/k + 2\ell\sqrt{\mathbf{I}(M : Y_J S|X_{\leq J}J)} + \eta\left(\sqrt{\mathbf{I}(M : Y_J S|X_{\leq J}J)}\right)\\ &\leq 2e\ell/k + 2\ell\sqrt{2^\ell/n} + \eta\left(\sqrt{2^\ell/n}\right).\end{aligned}$$

The same bound applies to $\mathbf{I}(M : Y_S G_S|Y_{\leq J}X_{\leq J}J)$. For each fixing of $x_{\leq j}y_{\leq j}j$, we have $X_S F_S - M - Y_S G_S$. Thus we can apply Lemma 6 to conclude that

$$|p(m|x_{\leq j}y_{\leq j}j) - p(m|x_{\leq j}y_{\leq j}j, x_s f_s = y_s g_s)| \leq 2\sqrt[3]{\mathbf{I}(M : X_S F_S|x_{\leq j}y_{\leq j}j)} + 2\sqrt[3]{\mathbf{I}(M : Y_S G_S|x_{\leq j}y_{\leq j}j)}.$$

Since $p(x_{\le j}y_{\le j}j) = p(x_{\le j}y_{\le j}j|x_sf_s = y_sg_s)$, we can use Proposition 7 to bound

$$\begin{aligned}
|p(m) - p(m|\mathcal{E}_0)| &= |p(m) - p(m|x_sf_s = y_sg_s)| \\
&\le \mathop{\mathbb{E}}_{p(x_{\le j}y_{\le j}j)} [|p(m|x_{\le j}y_{\le j}j) - p(m|x_{\le j}y_{\le j}j, x_sf_s = y_sg_s)|] \\
&\le \mathop{\mathbb{E}}_{p(x_{\le j}y_{\le j}j)} \left[2\sqrt[3]{\mathbf{I}\left(M : X_SF_S|x_{\le j}y_{\le j}j\right)} + 2\sqrt[3]{\mathbf{I}\left(M : Y_SG_S|x_{\le j}y_{\le j}j\right)}\right] \\
&\le 2\sqrt[3]{\mathbf{I}\left(M : X_SF_S|X_{\le j}Y_{\le J}J\right)} + 2\sqrt[3]{\mathbf{I}\left(M : Y_SG_S|X_{\le J}Y_{\le J}J\right)} \\
&\le 4\sqrt[3]{2e\ell/k + 2\ell\sqrt{2^\ell/n} + \eta\left(\sqrt{2^\ell/n}\right)},
\end{aligned}$$

where the second last inequality follows from the concavity of 3^{rd}-root over non-negative reals.

4.1 Proofs of the Lemmas

4.1.1 Proof of Lemma 4

Once $Y_{<J}J$ are fixed, S is determined by X_J. Thus $I(M;S|Y_{\le J}J) \le I(M;X_J|Y_{\le J}J)$, and similarly $I(M;S|X_{\le J}J) \le I(M;Y_J|X_{\le J}J)$. Since $X_{<J} = Y_{<J}$, we have

$$\mathbf{I}\left(M : X_J|Y_{\le J}J\right) = \mathbf{I}\left(M : X_J|Y_JX_{<J}J\right) = \sum_m p(m) \mathop{\mathbb{E}}_{yxj|m} \left[\frac{X_j|my_jx_{<j}j}{X_j|y_jx_{<j}j}\right]. \tag{3}$$

Recall that $M = m$ is equivalent to the events $\mathcal{S}_m \wedge \mathcal{T}_m$, where $\mathcal{S}_m, \mathcal{T}_m$ are as in Proposition 16. After fixing $x_{<j}j$, X_j is independent of Y_j (and hence $\mathcal{T}_m$). So by Proposition 17, (3) can be rewritten

$$\sum_m p(\mathcal{S}_m)p(\mathcal{T}_m|\mathcal{S}_m) \mathop{\mathbb{E}}_{xj|\mathcal{S}_m\mathcal{T}_m} \left[\frac{X_j|\mathcal{S}_mx_{<j}j}{X_j|x_{<j}j}\right] \le \sum_m p(\mathcal{S}_m) \mathop{\mathbb{E}}_{xj|\mathcal{S}_m} \left[\frac{X_j|\mathcal{S}_mx_{<j}j}{X_j|x_{<j}j}\right],$$

where the inequality follows from the fact that $\mathbb{E}_a[h(a)] \ge p(\mathcal{W})\,\mathbb{E}_{a|\mathcal{W}}[h(a)]$, for any non-negative function h. Since J is independent of X (and hence $\mathcal{S}_m$), we can use the chain rule to write the inner expectation as

$$\frac{1}{n}\sum_m p(\mathcal{S}_m)\frac{X|\mathcal{S}_m}{X} \le \frac{1}{n}\sum_m p(\mathcal{S}_m)\log\frac{1}{p(\mathcal{S}_m)} \le \frac{2^\ell}{n},$$

where the second inequality follows from Proposition 12 and the third from the fact that for $0 \le \gamma \le 1$, it holds that $\gamma\log(1/\gamma) \le \frac{\log e}{e} < 1$.

4.1.2 Proof of Lemma 5

We shall prove:

Claim 18. $\mathbf{I}\left(C : U_S|VS\right) \le \sum_{i=1}^t \mathbb{E}_{cu}\left[p(i \in S|c) \cdot \frac{U_i|cu_{<i}}{U_i|u_{<i}}\right]$.

Let $\mathcal{W}$ denote the event that $p(i \in S|c) \geq 2e/k + \sqrt{\mathbf{I}(C:S)}$ for some $i \in [t]$. $\mathcal{W}$ is determined by C. We show:

Claim 19. $p(\mathcal{W}) \leq \sqrt{\mathbf{I}(C:S)}$.

Using the Claims 18 and 19:

$$\mathbf{I}(C:U_S|VS) \leq p(\mathcal{W}) \sum_{i=1}^{t} \mathop{\mathbb{E}}_{cu|\mathcal{W}} \left[\frac{U_i|cu_{<i}}{U_i|u_{<i}} \right] + (2e/k + \sqrt{\mathbf{I}(C:S)}) \cdot \sum_{i=1}^{t} \mathop{\mathbb{E}}_{cu} \left[\frac{U_i|cu_{<i}}{U_i|u_{<i}} \right].$$

Since $\mathcal{W}$ is determined by c, we have $p(u_i|cu_{<i}\mathcal{W}) = p(u_i|cu_{<i})$. Thus the chain rule gives:

$$\begin{aligned}\mathbf{I}(C:U_S|VS) &\leq p(\mathcal{W}) \sum_{i=1}^{t} \mathop{\mathbb{E}}_{cu|\mathcal{W}} \left[\frac{U_i|cu_{<i}\mathcal{W}}{U_i|u_{<i}} \right] + (2e/k + \sqrt{\mathbf{I}(C:S)}) \sum_{i=1}^{t} \mathop{\mathbb{E}}_{cu} \left[\frac{U_i|cu_{<i}}{U_i|u_{<i}} \right] \\ &= p(\mathcal{W}) \mathop{\mathbb{E}}_{c|\mathcal{W}} \left[\frac{U|c\mathcal{W}}{U} \right] + (2e/k + \sqrt{\mathbf{I}(C:S)}) \mathop{\mathbb{E}}_{c} \left[\frac{U|c}{U} \right]\end{aligned}$$

By Proposition 12 and Claim 19,

$$\begin{aligned}\mathbf{I}(C:U_S|VS) &\leq p(\mathcal{W})(\log(1/p(\mathcal{W})) + \mathbf{I}(U:C|\mathcal{W})) + (2e/k + \sqrt{\mathbf{I}(C:S)}) \cdot \mathbf{I}(U:C) \\ &\leq \eta(p(\mathcal{W})) + p(\mathcal{W}) \cdot \mathbf{I}(U:C|\mathcal{W}) + \sqrt{\mathbf{I}(C:S)} \cdot \mathbf{I}(U:C) + 2e\mathbf{I}(U:C)/k \\ &\leq \eta(\sqrt{\mathbf{I}(C:S)}) + 2\sqrt{\mathbf{I}(C:S)}\ell + 2e\ell/k,\end{aligned}$$

where we used the fact that η (see (1)) is a non-decreasing function. It only remains to prove the claims:

Proof of Claim 18. For $i \in [t]$, set $U_i' = U_i$ if $i \in S$ and set $U_i' = \perp$ otherwise. We have that $\mathbf{I}(C:U_S|VS) = \mathbf{I}(C:U_1'U_2'\ldots U_t'|VS)$, so by the chain rule, we get

$$\mathbf{I}(C:U_S|VS) = \mathop{\mathbb{E}}_{cuvs} \left[\sum_{i=1}^{t} \frac{U_i'|cu_{<i}'vs}{U_i'|u_{<i}'vs} \right] = \mathop{\mathbb{E}}_{cuvs} \left[\sum_{i \in s} \frac{U_i|cu_{<i}'vs}{U_i|u_{<i}'vs} \right],$$

since when $i \notin S$, $U_i' = \perp$ (and so the divergence is 0) and when $i \in S$, $U_i' = U_i$. By assumption, U is independent of VS and $U - C - VS$, so we can write:

$$= \sum_{i=1}^{t} \mathop{\mathbb{E}}_{cu} \left[p(i \in S|c) \cdot \frac{U_i|cu_{<i}'}{U_i|u_{<i}'} \right] \leq \sum_{i=1}^{t} \mathop{\mathbb{E}}_{cu} \left[p(i \in S|c) \cdot \frac{U_i|cu_{<i}}{U_i|u_{<i}} \right], \tag{4}$$

by Proposition 13, and the fact that $p(u_i|u_{<i}') = p(u_i) = p(u_i|u_{<i})$. □

Proof of Claim 19. Define $S_i = 1$ if $i \in S$ and 0 otherwise. By Proposition 15, whenever c is bad,

$$\frac{S|c}{S} \geq \frac{S_i|c}{S_i} \geq p(i \in S|c) \left(\log \frac{k \cdot p(i \in S|c)}{e} \right) \geq \sqrt{\mathbf{I}(C:S)}.$$

Since $\mathbf{I}(C:S) = \mathbb{E}_c \left[\frac{S|c}{S} \right]$, the claim follows from Markov's inequality. □

4.1.3 Proof of Lemma 6

We assume $\mathbf{I}(C:A), \mathbf{I}(C:B) \leq 1$, since otherwise the Lemma is trivially true. For brevity, set

$$\alpha^3 = \mathbf{I}(C:A) = \mathop{\mathbb{E}}_c \left[\frac{A|c}{A} \right] \text{ and } \beta^3 = \mathbf{I}(C:B) = \mathop{\mathbb{E}}_c \left[\frac{B|c}{B} \right].$$

Call c *bad* if $\frac{A|c}{A} \geq \alpha^2$ or $\frac{B|c}{B} \geq \beta^2$, and good otherwise. By Markov's inequality, the probability that C is bad is at most $\alpha + \beta$. To prove Lemma 6, we need the following Claim proved in [GKR14b]. For completeness, we include the short proof in Appendix B:

Claim 20. *If $A, B \in [T]$ are independent, A is γ_1-close to uniform and B is γ_2-close to uniform. Then, $p(a=b) \geq \frac{1-\gamma_1-\gamma_2}{T}$.*

When c is good, Pinsker's inequality (Proposition 10) and Claim 20 imply that

$$p(c|a=b) = \frac{p(c) \cdot p(a=b|c)}{p(a=b)} \geq (1-\alpha-\beta) \cdot p(c). \tag{5}$$

For any set Q, (5) implies that

$$\begin{aligned} p(c \in Q) - p(c \in Q | a=b) &\leq \sum_{c \in Q, c \text{ bad}} p(c) + \sum_{c \in Q, c \text{ good}} (p(c) - p(c|a=b)) \\ &\leq \alpha + \beta + \sum_c p(c)(\alpha+\beta) \leq 2\alpha + 2\beta, \end{aligned}$$

as required.

5 Information upper bound

The low information protocol is given in Figure 2. The protocol is parameterized by ϵ.

Lemma 21. *The protocol outputs the correct answer with probability at least $1 - 4\epsilon$.*

Proof. Since whether a sample z with $|z| \geq J$ is consistent or not is determined solely by the J^{th} step, the probability that the parties sample a z with $|z| \geq J$ and z inconsistent is at most 2ϵ. Given that this event does not happen, the probability that the parties abort in a particular step is at most $(2\epsilon)^{r-1}$, since to abort one of them must choose to send uniform values in at least $r-1$ steps where their inputs are equal. By the union bound, the probability that the parties abort in any step is at most $n(2\epsilon)^{r-1}$. If neither of these bad events happens, the protocol computes the correct answer. Thus the probability of making an error is at most $2\epsilon + n(2\epsilon)^{r-1} = 4\epsilon$, by the choice of r. □

The following theorem proves that the information cost of the protocol is small.

Theorem 22. *If M denotes the messages in the protocol of Figure 2,*

$$\left.\begin{matrix} \mathbf{I}(M : XF | YG\mathcal{E}) \\ \mathbf{I}(M : YG | XF\mathcal{E}) \end{matrix}\right\} \leq 2\log(k/\epsilon) \cdot (1 + 2\epsilon \cdot \log n \cdot 2^{\frac{2\log n}{k \log(1/2\epsilon)}}).$$

Input: Alice is given (x, f), Bob is given (y, g). Both know a parameter $\epsilon \in (0, 1)$.
Output: $f(z) + g(z) \mod 2$ for some consistent z.

Set z to be the empty string;
Set $r = \lceil \frac{\log n}{\log(1/2\epsilon)} + 2 \rceil$;
for $i = 1, 2, \ldots, n$ **do**
  Alice sets $a_i = \begin{cases} x(z_{<i}) & \text{with probability } 1-\epsilon \\ \text{uniform element of } [k] & \text{with probability } \epsilon \end{cases}$,
  Bob sets $b_i = \begin{cases} y(z_{<i}) & \text{with probability } 1-\epsilon \\ \text{uniform element of } [k] & \text{with probability } \epsilon \end{cases}$;
  Send $m_i = a_i, b_i$ to each other;
  Set $w \in [k]$ so that $w = a_i + b_i \mod k$, and append w to the string z;
  if $i \geq r$ *and* $a_{i'} \neq b_{i'}$ *for all* i' *with* $i - r + 1 \leq i' \leq i$ **then**
    Terminate the protocol;
  end
end
Send $f(z), g(z)$;

Figure 2: Protocol π_ϵ

Setting $\epsilon = 1/\log n$ gives Theorem 2. To prove the Theorem 22, we bound $\mathbf{I}(M : XF|YG\mathcal{E})$. The second term is bounded in the same way. By the chain rule, we can write

$$\mathbf{I}(M : XF|YG\mathcal{E}) = \mathop{\mathbb{E}}_{xfyg|\mathcal{E}} \left[\frac{M|xfyg\mathcal{E}}{M|yg\mathcal{E}} \right] = \mathop{\mathbb{E}}_{mxfyg|\mathcal{E}} \left[\sum_{i=1}^{|m|} \frac{M_i|m_{<i}xfyg\mathcal{E}}{M_i|m_{<i}yg\mathcal{E}} \right].$$

We prove the following claim, which bounds the contribution to the divergence of each possible message:

Claim 23.

$$\frac{M_i|m_{<i}xfyg\mathcal{E}}{M_i|m_{<i}yg\mathcal{E}} \begin{cases} = 0 & \text{if } i \leq n \text{ and } x(z_{<i}) = y(z_{<i}), \\ \leq \log(k/\epsilon) & \text{if } i \leq n \text{ and } x(z_{<i}) \neq y(z_{<i}), \\ = 1 & \text{if } i = n+1. \end{cases}$$

Before proving Claim 23, we show how to use it to bound the information. Set $Q_i = 1$ if $|M| \geq i$ and $X(Z_{<i}) \neq Y(Z_{<i})$, and 0 otherwise. Claim 23 implies that $\mathbf{I}(M : FX|YG\mathcal{E}) \leq 1 + \log(k/\epsilon) \cdot \mathbb{E}\left[\sum_{i=1}^n Q_i \middle| \mathcal{E}\right]$, so it only remains to bound $\mathbb{E}\left[\sum_{i=1}^n Q_i \middle| \mathcal{E}\right]$.

Claim 24.

$$\mathbb{E}\left[\sum_{i=1}^n Q_i \middle| \mathcal{E}\right] \leq 1 + \frac{2r\epsilon}{(1-1/k)^r} \leq 1 + 2\epsilon \cdot \log n \cdot 2^{\frac{2\log n}{k \log(1/2\epsilon)}}.$$

Claims 23 and 24 complete the proof of Theorem 22. We prove them next.

Proof of Claim 23. When $i = n+1$, a direct calculation shows

$$\frac{M_{n+1}|m_{\le n}xfyg\mathcal{E}}{M_{n+1}|m_{\le n}yg\mathcal{E}} = 1 \cdot \log(1/2) = 1.$$

For the other two cases, we use Proposition 14 to bound the terms corresponding to each i. Set $q(myg|\mathcal{E}) = p(m|y\mathcal{E}, x = y)$. By Proposition 14, when $i \le n$ we have

$$\frac{M_i|m_{<i}xfyg\mathcal{E}}{M_i|m_{<i}yg\mathcal{E}} \le \frac{p(M_i|m_{<i}xfyg\mathcal{E})}{q(M_i|m_{<i}yg\mathcal{E})}.$$

If $x(z_{<i}) = y(z_{<i})$, $q(m_i|m_{<i}yg\mathcal{E}) = p(m_i|m_{<i}xfyg\mathcal{E})$, proving the first bound. If $x(z_{<i}) \neq y(z_{<i})$, then $p(A_i = a|m_{<i}xfyg\mathcal{E}) = q(A_i = a|m_{<i}yg\mathcal{E})$, except when $a = x(z_{<i})$ or $a = y(z_{<i})$. Thus the divergence can be bounded by the contribution of these two values:

$$\begin{aligned}\frac{p(M_i|m_{<i}xfyg\mathcal{E})}{q(M_i|m_{<i}yg\mathcal{E})} &= (1-\epsilon+\epsilon/k)\log\frac{1-\epsilon+\epsilon/k}{\epsilon/k} + (\epsilon/k)\log\frac{\epsilon/k}{1-\epsilon+\epsilon/k} \\ &\le \log\frac{1-\epsilon+\epsilon/k}{\epsilon/k} \le \log(k/\epsilon),\end{aligned}$$

as required. □

Proof of Claim 24. Let T be such that M_T is the last message sent in the protocol. Let $\mathcal{W}$ be the event that $T \ge J$ and Z sampled by the protocol is not consistent. Since $p(\mathcal{W}|\mathcal{E}) \le 2\epsilon$, and $\mathbb{E}\left[\sum_{i=1}^n Q_i \middle| \neg\mathcal{W}\mathcal{E}\right] \le 1$,

$$\mathbb{E}\left[\sum_{i=1}^n Q_i \middle| \mathcal{E}\right] \le 2\epsilon \cdot \mathbb{E}\left[\sum_{i=1}^n Q_i \middle| \mathcal{W}\mathcal{E}\right] + 1.$$

If $T \ge J$, $\sum_{i=1}^n Q_i = \sum_{i=J}^T Q_i \le T - J$, so we get $\mathbb{E}\left[\sum_{i=1}^n Q_i \middle| \mathcal{W}\mathcal{E}\right] \le 2\epsilon \cdot \mathbb{E}[T - J|\mathcal{W}\mathcal{E}]$. Note that $T - J$ roughly behaves like a geometric random variable. We have

$$\begin{aligned}\mathbb{E}[T-J|\mathcal{W}\mathcal{E}] &\le p(T-J \le r|\mathcal{W}\mathcal{E})r + p(T-J > r|\mathcal{W}\mathcal{E})(r + \mathbb{E}[T-J|\mathcal{W}\mathcal{E}]) \\ &\le (1-1/k)^r r + (1-(1-1/k)^r)(r + \mathbb{E}[T-J|\mathcal{W}\mathcal{E}]) \\ \Rightarrow \mathbb{E}[T-J|\mathcal{W}\mathcal{E}] &\le \frac{r}{(1-1/k)^r},\end{aligned}$$

as required. The second inequality follows from the fact that $1/(1-1/k) \le 2^{2/k}$, for $k \ge 2$, and by the choice of r. □

6 Acknowledgements

We thank Paul Beame and Sivaramakrishnan Ramamoorthy for useful conversations.

References

[BYJKS02] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An Information Statistics Approach to Data Stream and Communication Complexity. In *FOCS*, pages 209–218, 2002.

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to Compress Interactive Communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[Bra12] Mark Braverman. Interactive Information Complexity. In *STOC*, pages 505–524, 2012.

[BG14] Mark Braverman and Ankit Garg. Public vs Private Coin in Bounded-Round Information. In *ICALP*, pages 502–513, 2014.

[BR11] Mark Braverman and Anup Rao. Information Equals Amortized Communication. In *FOCS*, pages 748–757, 2011.

[BRWY13] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct Products in Communication Complexity. In *FOCS*, pages 746–755, 2013.

[BW12] Mark Braverman and Omri Weinstein. A Discrepancy Lower Bound for Information Complexity. In *APPROX-RANDOM*, pages 459–470, 2012.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational Complexity and the Direct Sum Problem for Simultaneous Message Complexity. In *FOCS*, pages 270–278, 2001.

[CGFS86] Fan R. K. Chung, Ronald L. Graham, Peter Frankl, and James B. Shearer. Some intersection theorems for ordered sets and graphs. *Journal of Combinatorial Theory, Series A*, 43(1):23–37, 1986.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[FJK+15] Lila Fontes, Rahul Jain, Iordanis Kerenidis, Mathieu Laurière, Sophie Laplante, and Jérémie Roland. Relative Discrepancy does not separate Information and Communication Complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 2015.

[GKR14a] Anat Ganor, Gillat Kol, and Ran Raz. Exponential Separation of Information and Communication. In *FOCS*, pages 176–185, 2014.

[GKR14b] Anat Ganor, Gillat Kol, and Ran Raz. Exponential Separation of Information and Communication for Boolean Functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:113, 2014.

[HJMR10] Prahladh Harsha, Rahul Jain, David A. McAllester, and Jaikumar Radhakrishnan. The Communication Complexity of Correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2010.

[KS92] Bala Kalyanasundaram and Georg Schnitger. The Probabilistic Communication Complexity of Set Intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992.

[KLL+12] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower Bounds on Information Complexity via Zero-Communication Protocols and Applications. In *FOCS*, pages 500–509, 2012.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, New York, NY, USA, 1997.

[Rad03] Jaikumar Radhakrishnan. Entropy and Counting. In *Computational Mathematics, Modelling and Algorithms (Ed. J.C. Misra)*, pages 146–168. Narosa Publishing House, New Delhi, 2003.

[RR15] Sivaramakrishnan Natarajan Ramamoorthy and Anup Rao. How to Compress Asymmetric Communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 2015.

[Raz92] A.A. Razborov. On the Distributional Complexity of Disjointness. *Theoretical Computer Science*, 106(2):385 – 390, 1992.

[Sha48] Claude E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, 623–, July, October 1948.

A Proofs of Basic Divergence Inequalities

Proof of Proposition 12.

$$\begin{aligned}\mathop{\mathbb{E}}_{m|\mathcal{W}}\left[\frac{A|m\mathcal{W}}{\overline{\overline{a}}}\right] - \mathbf{I}\left(A:M|\mathcal{W}\right) &= \sum_{am} p(am|\mathcal{W}) \log \frac{p(a|m\mathcal{W})\cdot p(a|\mathcal{W})}{p(a)\cdot p(a|m\mathcal{W})} \\ &= \sum_{a} p(a|\mathcal{W}) \log \frac{p(a|\mathcal{W})}{p(a)} \\ &\leq \sum_{a} p(a|\mathcal{W}) \log \frac{p(\mathcal{W}|a)}{p(\mathcal{W})} \leq \log \frac{1}{p(\mathcal{W})}.\end{aligned}$$

□

Proof of Proposition 13.

$$\begin{aligned}\mathop{\mathbb{E}}_{p(b)}\left[\frac{p(A|b)}{\overline{\overline{q(A)}}}\right] - \frac{p(A)}{\overline{\overline{q(A)}}} &= \sum_{a,b} p(ab) \log \frac{p(a|b)\cdot q(a)}{q(a)\cdot p(a)} \\ &= \mathop{\mathbb{E}}_{p(b)}\left[\frac{p(A|b)}{\overline{\overline{p(A)}}}\right] \geq 0,\end{aligned}$$

by Proposition 11. □

Proof of Proposition 14.

$$\begin{aligned}\mathop{\mathbb{E}}_{p(b)}\left[\frac{p(A|b)}{\overline{\overline{q(A)}}}\right] - \mathop{\mathbb{E}}_{p(b)}\left[\frac{p(A|b)}{\overline{\overline{p(A)}}}\right] &= \sum_{a,b} p(ab)\left(\log \frac{p(a|b)}{q(a)} - \log \frac{p(a|b)}{p(a)}\right) \\ &= \sum_{a,b} p(ab) \log \frac{p(a)}{q(a)} = \frac{p(A)}{\overline{\overline{q(A)}}} \geq 0,\end{aligned}$$

by Proposition 11. □

Proof of Proposition 15. Let $p(1) = \gamma, q(1) = \epsilon$. Then

$$\begin{aligned}\frac{p}{\overline{\overline{q}}} &= \gamma \log(\gamma/\epsilon) + (1-\gamma)\log(1-\gamma)/(1-\epsilon) \\ &\geq \gamma \log(\gamma/\epsilon) + (1-\gamma)(-\gamma \log e) \\ &\geq \gamma(\log(\gamma/\epsilon) - \log e) = \gamma \log \frac{\gamma}{e\epsilon},\end{aligned}$$

where we used the fact that $\log(1-\gamma)/\log e = \ln(1-\gamma) \geq -\gamma$. □

B Proof of Claim 20

For each i, let $p(A = i) = 1/T + \alpha_i$ and $p(B = i) = 1/T + \beta_i$. Then $\sum_i \alpha_i = \sum_i \beta_i = 0$, and $\alpha_i, \beta_i \geq -1/T$. Using these facts,

$$\begin{aligned} p(a = b) &= \sum_i (1/T + \alpha_i)(1/T + \beta_i) \\ &= 1/T + \frac{\sum_i \alpha_i}{T} + \frac{\sum_i \beta_i}{T} + \sum_i \alpha_i \beta_i = 1/T + \sum_i \alpha_i \beta_i. \end{aligned}$$

To lower bound the above, we will only consider the negative terms in the summation:

$$p(a = b) \geq 1/T + \sum_{i:\alpha_i>0,\beta_i<0} \alpha_i\beta_i + \sum_{i:\alpha_i<0,\beta_i>0} \alpha_i\beta_i \geq 1/T - (1/T)\sum_{i:\alpha_i>0} \alpha_i - (1/T)\sum_{i:\beta_i>0} \beta_i.$$

$\sum_{i:\alpha_i>0} \alpha_i$ is the statistical distance between A and uniform and likewise for B. So, we get,

$$p(a = b) \geq \frac{1 - \gamma_1 - \gamma_2}{T}.$$

C Relative Discrepancy and Fooling Distributions

Let $f(x, y)$ be a boolean function and $q(x, y)$ be a distribution such that the events $f = 0$ and $f = 1$ are equi-probable under q. Then, f has (ϵ, δ) *relative discrepancy* under q if there exists a distribution $u(x, y)$ such that for every rectangle $S \times T$,

$$\left.\begin{array}{l} q(x \in S, y \in T | f = 0) \\ q(x \in S, y \in T | f = 1) \end{array}\right\} \geq (1 - \epsilon)u(x \in S, y \in T) \quad \text{if } u(x \in S, y \in T) \geq \delta. \tag{6}$$

[GKR14b] proved:

Lemma 25 ([GKR14b])**.** *If f has $(\epsilon = 1/3, \delta)$ relative discrepancy under q, then any protocol that computes f has communication $\Omega(\log 1/\delta)$.*

Here we show that low relative discrepancy implies the existence of a fooling distribution.

Claim 26 (Relative Discepancy implies Fooling Distribution)**.** *If f has relative discrepancy $(\epsilon, 2^{-2\ell})$ then there exists a distribution u such that if $m \in \{0,1\}^\ell$ denote the messages of a protocol, then $q(m|f = 0) \overset{\gamma}{\approx} u(m) \overset{\gamma}{\approx} q(m|f = 1)$ where $\gamma = 2^{-\ell} + \epsilon$.*

Proof. Let u be the distribution that satisfies (6) with $\delta = 2^{-2\ell}$. We will show that $u(m) \approx q(m|f = 0)$. The proof for $u(m) \approx q(m|f = 1)$ is similar. Define $B = \{m | u(m) < 2^{-2\ell}\}$ and note that $u(m \in B) < 2^\ell 2^{-2\ell} < 2^{-\ell}$. Also observe that when $m \notin B$, then by (6), $u(m) - q(m|f = 0) \leq \epsilon u(m)$. Now for any set Q, we have

$$\begin{aligned} u(m \in Q) - u(m \in Q | f = 0) &\leq \sum_{m \in Q \cap B} u(m) + \sum_{m \in Q \cap \bar{B}} (u(m) - u(m|f = 0)) \\ &\leq u(B) + \epsilon u(\bar{B}) \leq 2^{-\ell} + \epsilon. \end{aligned}$$

Hence, $|u(m) - q(m|f = 0)| \leq 2^{-\ell} + \epsilon$. □

We remark that the existence of fooling distributions implies that the *adaptive* relative discrepancy (as defined in [GKR14b]) is small.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de