Electronic Colloquium on Computational Complexity, Report No. 164 (2011)

A discrepancy lower bound for information complexity

Mark Braverman* Omri Weinstein†

December 8, 2011

Abstract

This paper provides the first general technique for proving information lower bounds on two-party unbounded-rounds communication problems. We show that the discrepancy lower bound, which applies to randomized communication complexity, also applies to information complexity. More precisely, if the discrepancy of a two-party function f with respect to a distribution μ is $Disc_\mu f$, then any two party randomized protocol computing f must reveal at least $\Omega(\log(1/Disc_\mu f))$ bits of information to the participants. As a corollary, we obtain that any two-party protocol for computing a random function on $\{0,1\}^n \times \{0,1\}^n$ must reveal $\Omega(n)$ bits of information to the participants. The proof develops a new simulation result that may be of an independent interest.

*Princeton University and the University of Toronto, mbraverm@cs.princeton.edu. Partially supported by an NSERC Discovery Grant and an Alfred P. Sloan Fellowship.

†Princeton University, oweinste@cs.princeton.edu.

ISSN 1433-8092

1 Introduction

The main objective of this paper is to expand the available techniques for proving information complexity lower bounds for communication problems. Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function, and μ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Informally, the information complexity of f is the least amount of *information* the Alice and Bob need to exchange on average to compute $f(x,y)$ using a randomized communication protocol if initially x is given to Alice, y is given to Bob, and $(x,y) \sim \mu$. Note that information here is measured in the Shannon's sense, and the amount of information may be much smaller than the number of bits exchanged. Thus the randomized communication complexity of f is an upper bound on its information complexity, but may not be a lower bound.

Information complexity has first been introduced in the context of direct sum theorems for randomized communication complexity [CSWY01, BYJKS04, BBCR10]. These techniques are also being used in the related direction of direct product theorems [KSDW04, LSS08, Jai10, Kla10]. The direct sum line of work [HJMR07, JSR08, BBCR10, BR11, Bra11] has eventually led to a tight connection (=equality) between amortized communication complexity and information complexity. Thus proving lower bounds on the communication complexity of k copies of f for a growing k is equivalent to proving lower bounds on the information complexity of f. In particular if f satisfies $IC(f) = \Omega(CC(f))$, i.e. that its information cost is asymptotically equal to its communication complexity, then a strong direct sum theorem holds for f. In addition to the intrinsic interest of understanding the amount of information exchange that needs to be involved in computing f, direct sum theorems motivate the development of techniques for proving lower bounds on the information complexity of functions.

Another important motivation for seeking lower bounds on the information complexity of functions stems from understanding the limits of security in two-party computation. In a celebrated results Ben-Or et al. [BOGW88] (see also [AL11]) showed how a multi-party computation (with three or more parties) may be carried out in a way that reveals no information to the participants except for the computation's output. The protocol relies heavily on the use of random bits that are shared between some, but not all, parties. Such a resource can clearly not exist in the two-party setting. While it can be shown that a perfect information security is unattainable by two-party protocols [CK89, BYCKO93], quantitatively it is not clear just how much information must the parties "leak" to each other to compute f. The quantitative answer depends on the model in which the leakage occurs, and whether quantum computation is allowed [Kla04]. Information complexity answers this question in the strongest possible sense for classical protocols: the parties are allowed to use private randomness to help them "hide" their information, and the information revealed is measured on average. Thus an information complexity lower bound of I on a problem implies that the *average* (as opposed to worst-case) amount of information revealed to the parties is I.

As mentioned above, the information complexity is always lower bounded by the communication complexity of f. The converse is unknown to be true. Moreover, lower bound techniques for communication complexity do not readily translate into lower bound techniques for information complexity. The key difference is that a low-information protocol is not limited in the amount of communication it uses, and thus rectangle-based communication bounds do not readily convert into information lower bound. No general technique has been known to yield sharp information complexity lower bounds. A linear lower bound on the communication complexity of the disjointness function has been shown in [Raz92]. An information-theoretic proof of this lower bound [BYJKS04] can be adapted to prove a linear *information* lower bound on disjointness [Bra11]. One general technique for obtaining (weak) information complexity lower bounds was introduced in [Bra11],

where it has been shown that any function that has I bits of information complexity, has communication complexity bounded by $2^{O(I)}$. This immediately implies that the information complexity of a function f is at least the log of its communication complexity ($IC(f) \geq \Omega(\log(CC(f)))$). In fact, this result easily follows from the stronger result we prove in this paper (Theorem 3.1).

1.1 Our results

In this paper we prove that the discrepancy method – a general communication complexity lower bound technique – generalizes to information complexity. The discrepancy of f with respect to a distribution μ on inputs, denoted $Disc_\mu(f)$, measures how "unbalanced" can f get on any rectangle, where the balancedness is measured with respect to μ:

$$Disc_\mu(f) = \max_{R \text{ is a rectangle}} \left| \Pr_\mu[f(x,y) = 0 \wedge (x,y) \in R] - \Pr_\mu[f(x,y) = 1 \wedge (x,y) \in R] \right|. \quad (1)$$

A well-known lower bound (see e.g [KN97]) asserts that the distributional communication complexity of f, when required to predict f with advantage ε over a random guess (with respect to μ), is bounded from below by $\Omega(\log 1/Disc_\mu(f))$:

$$D^\mu_{1/2-\epsilon}(f) \geq \log(2\epsilon/Disc_\mu(f)).$$

Note that the lower bound holds even if we are merely trying to get an advantage of $\varepsilon = \sqrt{Disc_\mu(f)}$ over random guessing in computing f. We prove that the information complexity of computing f with probability $9/10$ with respect to μ is also bounded from below by $\Omega(\log(1/Disc_\mu(f)))$.

Theorem 1.1. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function and let μ be any probability distribution on $\mathcal{X} \times \mathcal{Y}$. Then*

$$\mathsf{IC}_\mu(f, 1/10) \geq \Omega(\log(1/Disc_\mu(f))).$$

Remark 1.2. The choice of $9/10$ is somewhat arbitrary. For randomized worst-case protocols, we may replace the success probability with $1/2 + \delta$ for a constant δ, since repeating the protocol constantly many times would yield the aforementioned success rate, while the information cost of the repeated protocol differs only by a constant factor from the original one. In particular, using prior-free information cost [Bra11] this implies $\mathsf{IC}(f, 1/2 - \delta) \geq \Omega_\delta(\log(1/Disc_\mu(f)))$.

In particular, Theorem 1.1 implies a linear lower bound on the information complexity of the inner product function $IP(x,y) = \sum_{i=1}^n x_i y_i \mod 2$, and on a random boolean function $f_r : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, expanding the (limited) list of functions for which good information-complexity lower bounds are known:

Corollary 1.3. *The information complexity $\mathsf{IC}_{uniform}(IP, 1/10)$ of $IP(x,y)$ is $\Omega(n)$. The information complexity $\mathsf{IC}_{uniform}(f_r, 1/10)$ of a random function f_r is $\Omega(n)$, except with probability $2^{-O(n)}$.*

The key technical idea in the proof of the theorem is a new simulation procedure that allows us to convert any protocol that has information cost I into a (two-round) protocol that has communication complexity $O(I)$ and succeeds with probability $> 1/2 + 2^{-O(I)}$, yielding a $2^{-O(I)}$ advantage over random guessing. Combined with the discrepancy lower bound for communication complexity, this proves Theorem 1.1.

1.2 Comparison and connections to prior results

The most relevant prior work is an article by Lee, Shraibman, and Špalek [LSS08]. Improving on an earlier work of Shaltiel [Sha03], Lee et al. show a direct product theorem for discrepancy, proving that the discrepancy of $f^{\otimes k}$ — the k-wise XOR of a function f with itself — behaves as $Disc(f)^{\Omega(k)}$. This implies in particular that the communication complexity of $f^{\otimes k}$ scales at least as $\Omega(k \cdot \log Disc(f))$. Using the fact that the limit as $k \to \infty$ of the amortized communication complexity of f is equal to the information cost of f [BR10], the result of Lee et al. "almost" implies the bound of Theorem 1.1. Unfortunately, the amortized communication complexity in the sense of [BR10] is the amortized cost of k copies of f, where *each* copy is allowed to err with some probability (say 1/10). Generally speaking, this task is much easier than computing the XOR (which requires *all* copies to be evaluated correctly with high probability. This specific problem can be addressed, but the reduction in [BR10] is not strong enough to be able to replace repetition with XOR in this context. Thus the lower bound that follows from Lee et al. applies to a more difficult problem, and does not imply the information complexity lower bound.

Our result can be viewed as a weak compression result for protocols, where a protocol for computing f that conveys I bits of information is converted into a protocol that uses $O(I)$ bits of *communication* and giving an advantage of $2^{-O(I)}$ in computing f. This strengthens the result in [Bra11] where a compression to $2^{O(I)}$ bits of communication has been shown. We still do not know whether compression to a protocol that uses $O(I)$ bits of communication and succeeds with high probability (as opposed to getting a small advantage over random) is possible.

2 Preliminaries

In an effort to make this paper as self-contained as possible, we provide some background on information theory and communication complexity, which is essential to proving our results. For further details and a more thorough treatment of these subjects see [BR10] and references therein.

Notation. We reserve capital letters for random variables and distributions, calligraphic letters for sets, and small letters for elements of sets. Throughout this paper, we often use the notation $|b$ to denote conditioning on the event $B = b$. Thus $A|b$ is shorthand for $A|B = b$.

We use the standard notion of *statistical/total variation* distance between two distributions.

Definition 2.1. Let D and F be two random variables taking values in a set $\mathcal{S}$. Their *statistical distance* is $|D - F| \stackrel{def}{=} \max_{\mathcal{T} \subseteq \mathcal{S}}(|\Pr[D \in \mathcal{T}] - \Pr[F \in \mathcal{T}]|) = \frac{1}{2}\sum_{s \in \mathcal{S}} |\Pr[D = s] - \Pr[F = s]|$

2.1 Information Theory

Definition 2.2. The *entropy* of a random variable X is $H(X) \stackrel{def}{=} \sum_x \Pr[X = x] \log(1/\Pr[X = x])$. The *conditional entropy* $H(X|Y)$ is defined to be $\mathbf{E}_{y \in_R Y}[H(X|Y = y)]$.

Definition 2.3 (Mutual Information)**.** The *mutual information* between two random variables A, B, denoted $I(A; B)$ is defined to be the quantity $H(A) - H(A|B) = H(B) - H(B|A)$. The *conditional mutual information* $I(A; B|C)$ is $H(A|C) - H(A|BC)$.

We also use the notion of *divergence* (also known as Kullback-Leibler distance or relative entropy), which is a different way to measure the distance between two distributions:

Definition 2.4 (Divergence)**.** The informational divergence between two distributions is

$$\mathbf{D}\left(A||B\right) \stackrel{def}{=} \sum_x A(x)\log(A(x)/B(x)).$$

Proposition 2.5. *Let A, B, C be random variables in the same probability space. For every a in the support of A and c in the support of C, let B_a denote $B|A=a$ and B_{ac} denote $B|A=a, C=c$. Then $I(A;B|C) = \mathbf{E}_{a,c\in_R A,C}[\mathbf{D}\left(B_{ac}||B_c\right)]$.*

2.2 Communication Complexity

Let $\mathcal{X}, \mathcal{Y}$ denote the set of possible inputs to the two players, who we name A and B. We view a *private coins protocol* for computing a function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{Z}_K$ as a rooted tree with the following structure:

- Each non-leaf node is *owned* by A or by B.
- Each non-leaf node owned by a particular player has a set of children that are owned by the other player. Each of these children is labeled by a binary string, in such a way that this coding is prefix free: no child has a label that is a prefix of another child.
- Every node is associated with a function mapping $\mathcal{X}$ to distributions on children of the node and a function mapping $\mathcal{Y}$ to distributions on children of the node.
- The leaves of the protocol are labeled by output values.

A public coin protocol is a distribution on private coins protocols, run by first using shared randomness to sample an index r and then running the corresponding private coin protocol π_r. Every private coin protocol is thus a public coin protocol. The protocol is called deterministic if all distributions labeling the nodes have support size 1.

Definition 2.6. The *communication cost* (or communication complexity) of a public coin protocol π, denoted $\mathsf{CC}(\pi)$, is the maximum number of bits that can be transmitted in any run of the protocol.

Definition 2.7. The *number of rounds* of a public coin protocol is the maximum depth of the protocol tree π_r over all choices of the public randomness.

Given a protocol π, $\pi(x,y)$ denotes the concatenation of the public randomness with all the messages that are sent during the execution of π. We call this the *transcript* of the protocol. When referring to the random variable denoting the transcript, rather than a specific transcript, we will use the notation $\Pi(x,y)$ — or simply Π when x and y are clear from the context, thus $\pi(x,y) \in_R \Pi(x,y)$. When x and y are random variables themselves, we will denote the transcript by $\Pi(X,Y)$, or just Π.

Definition 2.8 (Communication Complexity notation)**.** For a function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{Z}_K$, a distribution μ supported on $\mathcal{X} \times \mathcal{Y}$, and a parameter $\epsilon > 0$, $D^{\mu}_{\epsilon}(f)$ denotes the communication complexity of the cheapest deterministic protocol computing f on inputs sampled according to μ with error ϵ.

Definition 2.9 (Combinatorial Rectangle)**.** A *Rectangle* in $\mathcal{X} \times \mathcal{Y}$ is a subset $R \subseteq \mathcal{X} \times \mathcal{Y}$ which satisfies

$$(x_1, y_1) \in R \text{ and } (x_2, y_2) \in R \Longrightarrow (x_1, y_2) \in R$$

2.3 Information + Communication: The information cost of a protocol

The following quantity, which is implicit in [BYJKS04] and was explicitly defined in [BBCR10], is the central notion of this paper.

Definition 2.10. The *information cost* of a protocol π over inputs drawn from a distribution μ on $\mathcal{X} \times \mathcal{Y}$, is given by:

$$\mathsf{IC}_\mu(\pi) := I(\Pi; X|Y) + I(\Pi; Y|X).$$

Intuitively, Definition 2.10 captures what the two parties learn about each other's inputs from the execution transcript of the protocol π. The first term captures what the second player learns about X from Π – the mutual information between the input X and the transcript Π given the input Y. Similarly, the second term captures what the first player learns about Y from Π.

Note that the information of a protocol π depends on the prior distribution μ, as the mutual information between the transcript Π and the inputs depends on the prior distribution on the inputs. To give an extreme example, if μ is a singleton distribution, i.e. one with $\mu(\{(x, y)\}) = 1$ for some $(x, y) \in \mathcal{X} \times \mathcal{Y}$, then $\mathsf{IC}_\mu(\pi) = 0$ for all possible π, as no protocol can reveal anything to the players about the inputs that the do not already know *a-priori*. Similarly, $\mathsf{IC}_\mu(\pi) = 0$ if $\mathcal{X} = \mathcal{Y}$ and μ is supported on the diagonal $\{(x, x) : x \in \mathcal{X}\}$. As expected, one can show that the communication cost $\mathsf{CC}(\pi)$ of π is an upper bound on its information cost over *any* distribution μ:

Lemma 2.11. *[BR10] For any distribution μ, $\mathsf{IC}_\mu(\pi) \leq \mathsf{CC}(\pi)$.*

On the other hand, as noted in the introduction, the converse need not hold.

As one might expect, the information cost of a function f with respect to μ and error ρ is the least amount of information that needs to be revealed by a protocol computing f with error $\leq \rho$:

$$\mathsf{IC}_\mu(f, \rho) := \inf_{\pi:\ \mathbf{P}_\mu[\pi(x,y) \neq f(x,y)] \leq \rho} \mathsf{IC}_\mu(\pi).$$

The (prior-free) information cost was defined in [Bra11] as the minimum amount of information that a worst-case error-ρ randomized protocol can reveal against *all* possible prior distributions.

$$\mathsf{IC}\,(f, \rho) := \inf_{\pi \text{ is a protocol with } \mathbf{P}[\pi(x,y) \neq f(x,y)] \leq \rho \text{ for all } (x,y)} \max_{\mu} \mathsf{IC}_\mu(\pi).$$

This information cost matches the amortized *randomized* communication complexity of f [Bra11]. It is clear that lower bounds on $\mathsf{IC}_\mu(f, \rho)$ *for any distribution* μ also apply to $\mathsf{IC}\,(f, \rho)$.

3 Proof of Theorem 1.1

To establish the correctness of Theorem 1.1, we prove the following theorem, which is the central result of this paper:

Theorem 3.1. *Suppose that $\mathsf{IC}_\mu(f, 1/10) = I_\mu$. Then there exist a protocol π' such that*

- $\mathsf{CC}(\pi') = O(I_\mu)$.
- $\mathbf{P}_{(x,y)\sim\mu}[\pi'(x,y) = f(x,y)] \geq 1/2 + 2^{-O(I_\mu)}$

We first show how Theorem 1.1 follows from Theorem 3.1:

Proof of Theorem 1.1. Let f, μ be as in theorem 1.1, and let $\mathsf{IC}_\mu(f, 1/10) = I_\mu$. By theorem 3.1, there exists a protocol π' computing f with error probability $1/2 - 2^{-O(I_\mu)}$ using $O(I_\mu)$ bits of communication. Applying the discrepancy lower bound for communication complexity we obtain

$$O(I_\mu) \geq D^\mu_{1/2-2^{-O(I_\mu)}}(f) \geq \log(2 \cdot 2^{O(I_\mu)}/Disc_\mu(f)) \tag{2}$$

which after rearranging gives $I_\mu \geq \Omega(\log(1/Disc_\mu(f)))$, as desired.

We now turn to prove Theorem 3.1. The main step is the following sampling lemma.

Lemma 3.2. *Let μ be any distribution over a universe $\mathcal{U}$ and let $I \geq 0$ be a parameter that is known to both A and B. Further, let ν_A and ν_B be two distributions over $\mathcal{U}$ such that $\mathbf{D}(\mu||\nu_A) \leq I$ and $\mathbf{D}(\mu||\nu_B) \leq I$. The players are each given a pair of real functions (p_A, q_A), (p_B, q_B), $p_A, q_A, p_B, q_B : \mathcal{U} \to [0,1]$ such that for all $x \in \mathcal{U}$, $\mu(x) = p_A(x) \cdot p_B(x)$, $\nu_A(x) = p_A(x) \cdot q_A(x)$, and $\nu_B(x) = p_B(x) \cdot q_B(x)$. Then there is a (two round) sampling protocol $\Pi_1 = \Pi_1(p_A, p_B, q_A, q_B, I)$ which has the following properties:*

1. *at the end of the protocol, the players either declare that the protocol "fails", or output $x_A \in \mathcal{U}$ and $x_B \in \mathcal{U}$, respectively ("success");*
2. *let $\mathcal{S}$ be the event that the players output "success". Then $\mathcal{S} \Rightarrow x_A = x_B$, and $0.9 \cdot 2^{-50(I+1)} \leq \Pr[\mathcal{S}] \leq 2^{-50(I+1)}$.*
3. *if μ_1 is the distribution of x_A conditioned on $\mathcal{S}$, then $|\mu - \mu_1| < 2/9$.*

Furthermore, Π_1 can be "compressed" to a protocol Π_2 such that $\mathsf{CC}(\Pi_2) = 211I + 1$, whereas $|\Pi_1 - \Pi_2| \leq 2^{-59I}$ (by an abuse of notation, here we identify Π_i with the random variable representing its output).

We will use the following technical fact about the information divergence of distributions.

Claim 3.3. *[Claim 5.1 in [Bra11]] Suppose that $\mathbf{D}(\mu||\nu) \leq I$. Let ε be any parameter. Then*

$$\mu\left\{x : \ 2^{(I+1)/\varepsilon} \cdot \nu(x) < \mu(x)\right\} < \varepsilon.$$

For completeness, we repeat the short proof in the appendix.

Proof of Lemma 3.2 . Throughout the execution of Π_1, Alice and Bob interpret their shared random tape as a source of points (x_i, α_i, β_i) uniformly distributed in $\mathcal{U} \times [0, 2^{50(I+1)}] \times [0, 2^{50(I+1)}]$. Alice and Bob consider the first $T = |\mathcal{U}| \cdot 2^{100(I+1)} \cdot 60I$ such points. Their goal will be to discover the first index τ such that $\alpha_\tau \leq p_A(x_\tau)$ and $\beta_\tau \leq p_B(x_\tau)$ (where they wish to find it using a

minimal amount of communication, even if they are most likely to fail). First, we note that the probability that an index t satisfies $\alpha_t \le p_A(x_t)$ and $\beta_t \le p_B(x_t)$ is exactly $1/|\mathcal{U}|2^{50(I+1)}2^{50(I+1)} = 1/|\mathcal{U}|2^{100(I+1)}$. Hence the probability that $\tau > T$ (i.e. that x_τ is not among the T points considered) is bounded by

$$\left(1 - 1/|\mathcal{U}|2^{100(I+1)}\right)^T < e^{-T/|\mathcal{U}|2^{100(I+1)}} = e^{-60I} < 2^{-60I} \tag{3}$$

Denote by $\mathcal{A}$ the following set of indices $\mathcal{A} := \{i \le T : \ \alpha_i \le p_A(x_i) \text{ and } \beta_i \le 2^{50(I+1)} \cdot q_A(x_i)\}$, the set of potential candidates for τ from A's viewpoint. Similarly, denote $\mathcal{B} := \{i \le T : \ \alpha_i \le 2^{50(I+1)} \cdot q_B(x_i) \text{ and } \beta_i \le p_B(x_i)\}$.

The protocol Π_1 is very simple. Alice takes her bet on the first element $a \in \mathcal{A}$ and sends it to Bob. Bob outputs a only if (it just so happens that) $\beta_\tau \le p_B(a)$. The details are given in Figure 1.

Information-cost sampling protocol Π_1

1. Alice computes the set $\mathcal{A}$. Bob computes the set $\mathcal{B}$.
2. If $\mathcal{A} = \emptyset$, the protocol fails, otherwise Alice finds the first element $a \in \mathcal{A}$, and sends a to Bob.
3. Bob checks if $a \in \mathcal{B}$. If not, the protocol fails.
4. Alice and Bob output a ("success").

Figure 1: The sampling protocol Π_1 from Lemma 3.2

We turn to analyze Π_1. Denote the set of "Good" elements by

$$\mathcal{G} \stackrel{def}{=} \{x : \ 2^{50(I+1)} \cdot \nu_A(x) \ge \mu(x) \ \text{ and } \ 2^{50(I+1)} \cdot \nu_B(x) \ge \mu(x)\}\}.$$

Then by Claim 3.3, $\mu(\mathcal{G}) \ge 48/50 = 24/25$. The following claim asserts that if it succeeds, the output of Π_1 has the "correct" distribution on elements in $\mathcal{G}$.

Claim 3.4. *Assume $\mathcal{A}$ is nonempty. Then for any $x_i \in \mathcal{U}$, the probability that Π_1 outputs x_i is at most $\mu(x_i) \cdot 2^{-50(I+1)}$. If $x_i \in \mathcal{G}$, then this probability is* exactly $\mu(x_i) \cdot 2^{-50(I+1)}$.

Proof. Note that if $\mathcal{A}$ is nonempty, then for any $x_i \in \mathcal{U}$, the probability that x_i is the first element in $\mathcal{A}$ (i.e, $a = x_i$) is $p_A(x_i)q_A(x_i) = \nu_A(x_i)$. By construction, the probability that $\beta_i \le p_B(a)$ is $\min\{p_B(x_i)/(2^{50(I+1)}q_A(x_i)), 1\}$, and thus

$$\Pr[\Pi_1 \text{ outputs } x_i] \le p_A(x_i)q_A(x_i) \cdot \frac{p_B(x_i)}{2^{50(I+1)}q_A(x_i)} = \mu(x_i) \cdot 2^{-50(I+1)}.$$

On the other hand, if $x_i \in \mathcal{G}$, then we know that $p_B(x_i)/q_A(x_i) = \mu(x_i)/\nu_A(x_i) \le 2^{50(I+1)}$, and so the probability that $\beta_i \le p_B(a)$ is *exactly* $p_B(x_i)/(2^{50(I+1)}q_A(x_i))$. Since $\Pr[\Pi_1 \text{ outputs } x_i] = \Pr[a = x_i]\Pr[\beta_i \le p_B(a)]$ (assuming $\mathcal{A}$ is nonempty), we conclude that:

$$x_i \in \mathcal{G} \Longrightarrow \Pr[\Pi_1 \text{ outputs } x_i] = p_A(x_i)q_A(x_i) \cdot \frac{p_B(x_i)}{2^{50(I+1)}q_A(x_i)} = \mu(x_i) \cdot 2^{-50(I+1)}.$$

□

We are now ready to estimate the success probability of the protocol.

Proposition 3.5. *Let $\mathcal{S}$ denote the event that $\mathcal{A} \neq 0$ and $a \in \mathcal{B}$ (i.e, that the protocol succeeds). Then*

$$0.9 \cdot 2^{-50(I+1)} \leq \Pr[\mathcal{S}] \leq 2^{-50(I+1)}.$$

Proof. Using Claim 3.4, we have that

$$\Pr[\mathcal{S}] \leq \mathbf{P}[a \in \mathcal{B} \mid \mathcal{A} \neq \emptyset] = \sum_{i \in \mathcal{U}} \Pr[a = x_i] \Pr[\beta_i \leq p_B(a)] \leq \sum_{i \in \mathcal{U}} \mu(x_i) \cdot 2^{-50(I+1)} = 2^{-50(I+1)} \quad (4)$$

For the lower bound, we have

$$\begin{aligned}
\Pr[\mathcal{S}] &\geq \Pr[\beta_i \leq p_B(a) \mid \mathcal{A} \neq \emptyset] \cdot \Pr[\mathcal{A} \neq \emptyset] \geq \\
&\geq (1 - 2^{-60I}) \left(\sum_{i \in \mathcal{U}} \Pr[a = x_i] \Pr[\beta_i \leq p_B(a)] \right) \geq \\
&\geq (1 - 2^{-60I}) \left(\sum_{i \in \mathcal{G}} \Pr[a = x_i] \Pr[\beta_i \leq p_B(a)] \right) = \\
&= (1 - 2^{-60I}) \left(2^{-50(I+1)} \sum_{i \in \mathcal{G}} \mu(x_i) \right) = (1 - 2^{-60I}) \left(2^{-50(I+1)} \mu(\mathcal{G}) \right) \geq \\
&\geq \frac{24}{25} (1 - 2^{-60I}) 2^{-50(I+1)} \geq 0.9 \cdot 2^{-50(I+1)} \quad (5)
\end{aligned}$$

where the equality follows again from claim 3.4. This proves the second claim of Lemma 3.2. □

The following claim asserts that if $\mathcal{S}$ occurs, then the distribution of a is indeed close to μ.

Claim 3.6. *Let μ_1 be the distribution of $a|\mathcal{S}$. Then $|\mu_1 - \mu| \leq 2/9$.*

Proof. The claim follows directly from proposition 3.5. We defer the proof to the appendix.

We turn to the "Furthermore" part of of Lemma 3.2. The protocol Π_1 satisfies the premises of the lemma, except it has a high communication cost. This is due to the fact that Alice explicitly sends a to Bob. To reduce the communication, Alice will instead send $O(I)$ random hash values of a, and Bob will add corresponding consistency constraints to his set of candidates. The final protocol Π_2 is given in Figure 2.

Let $\mathcal{E}$ denote the event that in step 3 of the protocol, Bob finds an element $x_i \neq a$ (that is, the probability that the protocol outputs "success" but $x_A \neq x_B$). We upper bound the probability of $\mathcal{E}$. Given $a \in \mathcal{A}$ and $x_i \in \mathcal{B}$ such that $a \neq x_i$, the probability (over possible choices of the hash functions) that $h_j(a) = h_j(x_i)$ for $j = 1..d$ is $2^{-d} \leq 2^{-211I}$. For any t, $\mathbf{P}[t \in \mathcal{B}] \leq \frac{1}{|\mathcal{U}|} \sum_{x_i \in \mathcal{U}} p_B(x_i) q_B(x_i) \cdot 2^{50(I+1)} = \frac{1}{|\mathcal{U}|} \sum_{x_i \in \mathcal{U}} \nu_B(x_i) \cdot 2^{50(I+1)} = 2^{50(I+1)}/|\mathcal{U}|$. Thus, by a union bound we have

Information-cost sampling protocol Π_2

1. Alice computes the set $\mathcal{A}$. Bob computes the set $\mathcal{B}$.
2. If $\mathcal{A} = \emptyset$, the protocol fails. Otherwise, Alice finds the first element $a \in \mathcal{A}$ and sets $x_A = a$. She then computes $d = \lceil 211I \rceil$ random hash values $h_1(a), \ldots, h_d(a)$, where the hash functions are evaluated using public randomness.
3. Alice sends the values $\{h_j(a)\}_{1 \le j \le d}$ to Bob.
4. Bob finds the first index τ such that there is a $b \in \mathcal{B}$ for which $h_j(b) = h_j(a)$ for $j = 1..d$ (if such an τ exists). Bob outputs $x_B = x_\tau$. If there is no such index, the protocol fails.
5. Bob outputs x_B ("success").
6. Alice outputs x_A.

Figure 2: The sampling protocol Π_2 from Lemma 3.2

$$\begin{aligned} \mathbf{P}[\mathcal{E}] &\le \mathbf{P}[\exists x_i \in \mathcal{B}\ s.t\ x_i \neq a\ \wedge\ h_j(a) = h_j(x_i)\ \forall\ j = 1, \ldots, d] \le \\ &\le T \cdot 2^{50(I+1)} \cdot 2^{-d}/|\mathcal{U}| = 2^{150(I+1)} \cdot 60I \cdot 2^{-211I} \ll 2^{-60I}. \end{aligned} \tag{6}$$

By a slight abuse of notation, let Π_2 be the distribution of Π_2's output. Similarly, denote by Π_1 the distribution of the output of protocol Π_1. Note that if $\mathcal{E}$ does not occur, then the outcome of the execution of Π_2 is identical to the outcome of Π_1. Since $\mathbf{P}[\mathcal{E}] \le 2^{-60I}$, we have

$$|\Pi_2 - \Pi_1| = \Pr[\mathcal{E}] \cdot |[\Pi_2|\mathcal{E}] - [\Pi_1|\mathcal{E}]| \le 2 \cdot 2^{-60I} \ll 2^{-59I}$$

which finishes the proof of the lemma.

□

Using the above lemma, we are now ready to prove our main theorem.

Proof of Theorem 3.1 . Let π be a protocol that realizes the value $I_\mu := \mathsf{IC}_\mu(f, 1/10)$. In other words, π has an error rate of at most $1/10$ and information cost of at most I_μ with respect to μ. Denote by π_{xy} the random variable that represents that transcript π given the inputs (x, y), and by π_x (resp. π_y) the protocol conditioned on only the input x (resp. y). We denote by π_{XY} the transcripts where (X, Y) are also a pair of random variables. By Claim 3.3, we know that

$$I_\mu = I(X; \pi_{XY}|Y) + I(Y; \pi_{XY}|X) = \mathbf{E}_{(x,y)\sim\mu}[\mathbf{D}\left(\pi_{xy}||\pi_x\right) + \mathbf{D}\left(\pi_{xy}||\pi_y\right)]. \tag{7}$$

Let us now run the sampling algorithm Π_1 from Lemma 3.2, with the distribution μ taken to be π_{xy}, the distributions ν_A and ν_B taken to be π_x and π_y respectively, and I taken to be $20I_\mu$.

At each node v of the protocol tree that is owned by player X let $p_0(v)$ and $p_1(v) = 1 - p_0(v)$ denote the probabilities that the next bit sent by X is 0 and 1, respectively. For nodes owned by

player Y, let $q_0(v)$ and $q_1(v) = 1 - q_0(v)$ denote the probabilities that the next bit sent by Y is 0 and 1, respectively, *as estimated by player X given the input x*. For each leaf ℓ let $p_X(\ell)$ be the product of all the values of $p_b(v)$ from the nodes that are owned by X along the path from the root to ℓ; let $q_X(\ell)$ be the product of all the values of $q_b(v)$ from the nodes that are owned by Y along the path from the root to ℓ. The values $p_Y(\ell)$ and $q_Y(\ell)$ are defined similarly. For each ℓ we have $\mathbf{P}[\pi_{xy} = \ell] = p_X(\ell) \cdot p_Y(\ell)$, $\mathbf{P}[\pi_x = \ell] = p_X(\ell) \cdot q_X(\ell)$, and $\mathbf{P}[\pi_y = \ell] = p_Y(\ell) \cdot q_Y(\ell)$. Thus we can apply Lemma 3.2 so as to obtain the following protocol π' for computing f:

- If Π_1 fails, we return a random unbiased coin flip.
- If Π_1 succeeds, we return the final bit of the transcript sample T. Denote this bit by T_{out}.

To prove the correctness of the protocol, let $\mathcal{Z}$ denote the event that both $\mathbf{D}\left(\pi_{xy}||\pi_x\right) \leq 20I_\mu$ and $\mathbf{D}\left(\pi_{xy}||\pi_y\right) \leq 20I_\mu$. By (7) and Markov inequality, $\Pr[\mathcal{Z}] \geq 19/20$ (where the probability is taken with respect to μ). Denote by δ the probability that Π_1 succeeds. By the assertions of Lemma 3.2, $\delta \geq 0.9 \cdot 2^{-50(I+1)}$. Furthermore, if Π_1 succeeds, then we have $|T - \pi_{xy}| \leq 2/9$, which in particular implies that that $\mathbf{P}[T_{out} = \pi_{out}] \geq 7/9$. Finally, $\mathbf{P}[\pi_{out} = f(x,y)] \geq 9/10$, since π has error at most $1/10$ with respect to μ. Now, let $\mathcal{W}$ denote the indicator variable whose value is 1 iff $\pi'(x,y) = f(x,y)$. Putting together the above,

$$\mathbf{E}[\mathcal{W} \mid \mathcal{Z}] = (1-\delta) \cdot \frac{1}{2} + \delta \cdot \left(\frac{7}{9} - \frac{1}{10}\right) > \frac{1}{2} + \delta \cdot \frac{1}{6} > \frac{1}{2} + \frac{1}{8} \cdot 2^{-50(I+1)}. \tag{8}$$

On the other hand, note that by lemma 3.2 the probability that Π_1 succeeds is at most $2^{-50(I+1)}$ (no matter how large $\mathbf{D}\left(\pi_{xy}||\pi_x\right)$ and $\mathbf{D}\left(\pi_{xy}||\pi_y\right)$ are!), and so $\mathbf{E}[\mathcal{W} \mid \neg\mathcal{Z}] \geq (1 - 2^{-50(I+1)})/2$. Hence we conclude that

$$\begin{aligned}\mathbf{E}[\mathcal{W}] &= \mathbf{E}[\mathcal{W} \mid \mathcal{Z}] \cdot \mathbf{P}[\mathcal{Z}] + \mathbf{E}[\mathcal{W} \mid \neg\mathcal{Z}] \cdot \mathbf{P}[\neg\mathcal{Z}] \geq \left(\frac{1}{2} + \frac{1}{8} \cdot 2^{-50(I+1)}\right) \cdot \frac{19}{20} + \left(1 - 2^{-50(I+1)}\right) \cdot \frac{1}{2} \cdot \frac{1}{20} \\ &\geq \frac{1}{2} + \frac{1}{12} \cdot 2^{-50(I+1)} > \frac{1}{2} + \frac{1}{12} \cdot 2^{-1000(I_\mu+1)}. \end{aligned} \tag{9}$$

Finally, Lemma 3.2 asserts that $|\Pi_1 - \Pi_2| < 2^{-59I}$. Thus if we replace Π_1 by Π_2 in the execution of protocol π', the success probability decreases by at most $2^{-59I} \ll \frac{1}{12} \cdot 2^{-50(I+1)}$. Furthermore, the amount of communication used by π' is now

$$211I = 4220I_\mu = O(I_\mu).$$

Hence we conclude that with this modification, π' has the following properties:

- $\mathsf{CC}(\pi') = 4220 \cdot I_\mu$;
- $\mathbf{P}_{(x,y)\sim\mu}[\pi'(x,y) = f(x,y)] \geq 1/2 + 2^{-1000(I_\mu+1)-4}$;

which completes the proof. □

Remark 3.7. Using similar techniques, it was recently shown in [Bra11] that any function f whose information complexity is I has communication cost at most $2^{O(I)}$ [1], thus implying that $IC(f) \geq \Omega(\log(CC(f)))$. We note that this result can be easily derived (up to constant factors) from Theorem 3.1. Indeed, applying the "compressed" protocol $2^{O(I)} \log(1/\epsilon)$ independent times and taking a majority vote guarantees an error of at most ϵ (By a standard Chernoff bound[2]), with communication $O(I) \cdot 2^{O(I)} = 2^{O(I)}$. Thus, our result is strictly stronger than the former one.

References

[AL11] G. Asharov and Y. Lindell. A full proof of the bgw protocol for perfectly-secure multiparty computation. *Advances in CryptologyCRYPTO 2011*, 2011.

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, 2010.

[BOGW88] M. Ben-Or, S. Goldwasser, and A. Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation. In *Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 1–10. ACM, 1988.

[BR10] Mark Braverman and Anup Rao. Information equals amortized communication. *CoRR*, abs/1106.3595, 2010.

[BR11] M. Braverman and A. Rao. Information equals amortized communication. *Arxiv preprint arXiv:1106.3595*, 2011.

[Bra11] Mark Braverman. Interactive information complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:123, 2011.

[BYCKO93] R. Bar-Yehuda, B. Chor, E. Kushilevitz, and A. Orlitsky. Privacy, additional information and communication. *Information Theory, IEEE Transactions on*, 39(6):1930–1943, 1993.

[BYJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[CK89] B. Chor and E. Kushilevitz. A zero-one law for boolean privacy. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 62–72. ACM, 1989.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In Bob Werner, editor, *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, Los Alamitos, CA, October 14–17 2001. IEEE Computer Society.

[1] More precisely, it shows that for any distribution μ, $D^{\mu}_{\varepsilon+\delta}(f) = 2^{O(1+\mathsf{IC}_{\mu}(f,\varepsilon)/\delta^2)}$.

[2] See N.Alon and J. Spencer, "The Probabilistic Method" (Third Edition) ,Corollary A.1.14, p.312.

[HJMR07] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan. The communication complexity of correlation. In *Computational Complexity, 2007. CCC'07. Twenty-Second Annual IEEE Conference on*, pages 10–23. IEEE, 2007.

[Jai10] R. Jain. A strong direct product theorem for two-way public coin communication complexity. *Arxiv preprint arXiv:1010.0846*, 2010.

[JSR08] R. Jain, P. Sen, and J. Radhakrishnan. Optimal direct sum and privacy tradeoff results for quantum and classical communication complexity. *Arxiv preprint arXiv:0807.1267*, 2008.

[Kla04] Hartmut Klauck. Quantum and approximate privacy. *Theory Comput. Syst.*, 37(1):221–246, 2004.

[Kla10] H. Klauck. A strong direct product theorem for disjointness. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 77–86. ACM, 2010.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, New York, 1997. 96012840 96012840 Eyal Kushilevitz, Noam Nisan.

[KSDW04] H. Klauck, R. Spalek, and R. De Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 12–21. IEEE, 2004.

[LSS08] T. Lee, A. Shraibman, and R. Spalek. A direct product theorem for discrepancy. In *Computational Complexity, 2008. CCC'08. 23rd Annual IEEE Conference on*, pages 71–80. IEEE, 2008.

[Raz92] Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[Sha03] R. Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1):1–22, 2003.

A Proof of Claim 3.3 (from [Bra11])

Proof. Recall that $\mathbf{D}(\mu||\nu) = \sum_{x\in\mathcal{U}} \mu(x)\log\frac{\mu(x)}{\nu(x)}$. Denote by $\mathcal{N} = \{x:\ \mu(x) < \nu(x)\}$ – the terms that contribute a negative amount to $\mathbf{D}(\mu||\nu)$. First we observe that for all $0 < x < 1$, $x\log x > -1$, and thus

$$\sum_{x\in\mathcal{N}} \mu(x)\log\frac{\mu(x)}{\nu(x)} = \sum_{x\in\mathcal{N}} \nu(x)\cdot\frac{\mu(x)}{\nu(x)}\log\frac{\mu(x)}{\nu(x)} \geq \sum_{x\in\mathcal{N}} \nu(x)\cdot(-1) > -1.$$

Denote by $\mathcal{L} = \left\{x:\ 2^{(I+1)/\varepsilon}\cdot\nu(x) < \mu(x)\right\}$; we need to show that $\mu(\mathcal{L}) < \varepsilon$. For each $x\in\mathcal{L}$ we have $\log\frac{\mu(x)}{\nu(x)} > (I+1)/\varepsilon$. Thus

$$I \geq \mathbf{D}(\mu||\nu) \geq \sum_{x\in\mathcal{L}} \mu(x)\log\frac{\mu(x)}{\nu(x)} + \sum_{x\in\mathcal{N}} \mu(x)\log\frac{\mu(x)}{\nu(x)} > \mu(\mathcal{L})\cdot(I+1)/\varepsilon - 1,$$

implying $\mu(\mathcal{L}) < \varepsilon$. □

B Proof of Claim 3.6

Proof. For any $x_i \in \mathcal{U}$,

$$\mu_1(x_i) = \Pr(a = x_i \mid \mathcal{S}) \leq \frac{\mu(x_i)2^{-50(I+1)}}{\Pr[\mathcal{S}]} \leq \frac{\mu(x_i)}{0.9} = (1 + 1/9)\mu(x_i) \tag{10}$$

where the last inequality follows from Proposition 3.5. Hence,

$$|\mu_1 - \mu| = 2\Bigg(\sum_{x_i:\mu_1(x_i)\geq\mu(x_i)} \mu_1(x_i) - \mu(x_i)\Bigg) \leq 2\Bigg(\sum_{x_i:\mu_1(x_i)\geq\mu(x_i)} (1 + 1/9)\mu(x_i) - \mu(x_i)\Bigg) \leq 2/9 \tag{11}$$

This proves claim (3) of the lemma. □

ECCC ISSN 1433-8092

http://eccc.hpi-web.de