2013 IEEE 54th Annual Symposium on Foundations of Computer Science

Direct Products in Communication Complexity

Mark Braverman[*]
Department of Computer Science
Princeton University
Princeton, NJ
mbraverm@cs.princeton.edu

Anup Rao[†]
Computer Science and Engineering
University of Washington
Seattle, WA
anuprao@cs.washington.edu

Omri Weinstein[‡]
Department of Computer Science
Princeton University
Princeton, NJ
oweinste@cs.princeton.edu.

Amir Yehudayoff[§]
Department of Mathematics
Technion-IIT
Haifa, Israel
amir.yehudayoff@gmail.com

***Abstract*—We give exponentially small upper bounds on the success probability for computing the direct product of any function over any distribution using a communication protocol. Let $\mathsf{suc}(\mu, f, C)$ denote the maximum success probability of a 2-party communication protocol for computing the boolean function $f(x, y)$ with C bits of communication, when the inputs (x, y) are drawn from the distribution μ. Let μ^n be the product distribution on n inputs and f^n denote the function that computes n copies of f on these inputs.**

We prove that if $T \log^{3/2} T \ll (C-1)\sqrt{n}$ and $\mathsf{suc}(\mu, f, C) < \frac{2}{3}$, then $\mathsf{suc}(\mu^n, f^n, T) \leq \exp(-\Omega(n))$. When μ is a product distribution, we prove a nearly optimal result: as long as $T \log^2 T \ll Cn$, we must have $\mathsf{suc}(\mu^n, f^n, T) \leq \exp(-\Omega(n))$.

I. Introduction

The *direct sum* question is about quantifying the resources needed to compute n independent copies of a function in terms of the resources needed to compute one copy of it. If one copy can be computed with C resources, then n copies can be computed using nC resources, but is this optimal?

When the inputs are drawn from a distribution (or the computational model is randomized), one can also measure the probability of success of computing the function. The *direct product* question is about understanding what the maximum probability of success of computing n copies of the function is. If there is a way to compute one copy with C resources and success probability ρ, then n copies can be computed using nC resources with success probability ρ^n, but is this optimal?

In this work, we study the direct product question in the model of distributional communication complexity [Yao79]. Direct sum theorems for this model were proved in [BBCR10], and we strengthen their results to give direct product theorems. For a longer introduction to direct sums and direct products in communication complexity and their significance, we refer the reader to the introductions of [BBCR10], [JPY12].

We say that a communication protocol with inputs x, y computes a function f if the messages and public randomness of the protocol determine the value of f correctly. Let $\mathsf{suc}(\mu, f, C)$ denote the maximum success probability of a 2-party communication protocol of communication complexity C for computing function $f(x, y)$ when the inputs are drawn from the distribution μ. Let $f^n(x_1, \dots, x_n, y_1, \dots, y_n)$ denote the function that maps its inputs to the n bits $(f(x_1, y_1), f(x_2, y_2), \dots, f(x_n, y_n))$ and μ^n denote the product distribution on n pairs of inputs, where each pair is sampled independently according to μ. Our goal in this work is to prove new upper bounds on $\mathsf{suc}(\mu^n, f^n, T)$ in terms of $\mathsf{suc}(\mu, f, C)$, for $T \gg C$.

It is easy to prove that $\mathsf{suc}(\mu^n, f^n, nC) \geq \mathsf{suc}(\mu^n, f^n, C)^n$ (which is the success probability of the trivial solution which applies the per-copy optimal solution to each coordinate independently), and $\mathsf{suc}(\mu^n, f^n, C) \leq \mathsf{suc}(\mu, f, C)$ (since a protocol for f^n must in particular solve the first coordinate). Shaltiel [Sha03] showed that there exist μ, f, C such that $\mathsf{suc}(\mu^n, f^n, \frac{3}{4}nC) \geq \frac{3}{4}$, even though $\mathsf{suc}(\mu, f, C) \leq \frac{2}{3}$. Roughly, his ideas show that if $T \geq 2(1 - \mathsf{suc}(\mu, f, C))Cn$, there are examples where $\mathsf{suc}(\mu^n, f^n, T) > \mathsf{suc}(\mu, f, C)$. A counterexample due to Feige [Fei00], originally designed to show the limitations of parallel repetition, can be easily extended to show that under a slightly different (yet meaningful) definition of success of the protocol, there are problems whose communication complexity does not increase at all with

0272-5428/13 $26.00 © 2013 IEEE
DOI 10.1109/FOCS.2013.85

n. We elaborate on this issue in the full version of this paper [BRWY12] (see Appendix A).

Much past work has found success in proving upper bounds on $\mathsf{suc}(\mu^n, f^n, T)$ in special cases: for example, when f is the disjointness function [Kla10], or f is known to have small discrepancy [Sha03], [LSS08], [She11], or have a smooth rectangle bound [JY12], or the protocols computing f^n and f are restricted to using a bounded number of rounds of interaction [JPY12], [MWY13], or restricted to behaving somewhat independently on each coordinate of the input [PRW97]. The work of [PRW97] does imply a bound that behaves roughly like $\mathsf{suc}(\mu^n, f, C) < \exp(-\Omega(n - C))$. Note that the bound is meaningful only when $n > C$ and the protocol for n copies is not allowed to communicate more bits than the protocol for 1 copy. We refer the reader to [BBCR10], [JPY12] for more references.

Prior to our work, the only known general upper bounds on $\mathsf{suc}(\mu^n, f^n, T)$, for $T > C$, are a consequence of the direct sum theorem proved in [BBCR10]: If $\mathsf{suc}(\mu, f, C) \leq \frac{2}{3}$, then $\mathsf{suc}(\mu^n, f^n, T) \leq \frac{2}{3}$, as long as[1] $T \log T \ll (C-1)\sqrt{n}$. They also proved the same upper bound when $T\mathsf{polylog}(T) \ll Cn$ and μ is a product distribution.

In this work, we give new upper bounds that are exponentially small in n. When $\mathsf{suc}(\mu, f, C) \leq \frac{2}{3}$, we prove that $\mathsf{suc}(\mu^n, f^n, T) \leq \exp(-\Omega(n))$, as long as $T \log^{3/2} T \ll (C-1)\sqrt{n}$. By Yao's minimax principle [Yao79], we get an analogous statement for randomized worst case computation. If $\mathsf{suc}(f, C)$ denotes the maximum success probability for the best C-bit public coin randomized protocol computing f in the worst case, and if $\mathsf{suc}(f, C) \leq \frac{2}{3}$, then $\mathsf{suc}(f^n, T) \leq \exp(-\Omega(n))$ as long as $T \log^{3/2} T \ll (C-1)\sqrt{n}$. Formally, we prove:

Theorem 1 (Main Theorem)**.** *There is a universal constant $\alpha > 0$ such that if f is boolean, $\gamma = 1 - \mathsf{suc}(\mu, f, C)$, $T \geq 2$, and $T \log^{3/2} T < \alpha\gamma^{5/2}(C-1)\sqrt{n}$, then $\mathsf{suc}(\mu^n, f^n, T) \leq \exp\left(-\alpha\gamma^2 n\right)$.*

We remark that when f is a function that has a k-bit output, the above theorem is true with $(C-1)$ replaced by $(C-k)$. For simplicity, we focus on the case $k = 1$ throughout this paper. When μ is a product distribution, we prove an almost optimal result. We show that if $\mathsf{suc}(\mu, f, C) \leq \frac{2}{3}$ and $T \log^2 T \ll Cn$, then $\mathsf{suc}(\mu^n, f^n, T) \leq \exp(-\Omega(n))$.

Theorem 2 (Main Theorem for Product Distributions)**.** *There is a universal constant $\alpha > 0$ such that for every product distribution μ, if $\gamma = 1 - \mathsf{suc}(\mu, f, C)$, $T \geq 2$, and $T \log^2 T \leq \alpha\gamma^6 Cn$, then $\mathsf{suc}(\mu^n, f^n, T) \leq \exp\left(-\alpha\gamma^2 n\right)$.*

Our proofs heavily rely on methods from information theory [Sha48] which have been applied to a variety of problems in communication complexity [Raz92], [NW93], [Abl96], [CSWY01], [BYJKS04], [BBCR10], and ideas developed to prove the parallel repetition theorem [Raz98], [Hol07]. We give an overview of our proofs next.

A. Overview of the Proofs

The notation used below is formally defined in Section III. Before we describe our proof in detail, we give a high level overview of the proof of the direct sum theorem proved in [BBCR10]. The theorem is proved by reduction. For T, C roughly as in the theorems above, they show that any protocol π for computing n copies of f with communication complexity $\|\pi\| = T$ can be used to obtain a protocol for computing one copy, with communication complexity less than C. This proves that computing n copies requires communication complexity more than T. The reduction itself has two steps. In the first step, they show that π can be used to obtain a protocol for computing f with small *information cost* (which we discuss below). In the second step, they show that any protocol with small information cost can be compressed to obtain a protocol that actually has small communication.

[CSWY01] were the first to define the (external) information cost of protocols. Let the inputs to a protocol be X, Y, the messages be M and the public randomness be R. The *external information cost* [CSWY01] of the protocol is the mutual information between the inputs and the messages, conditioned on the public randomness: $I(XY; M|R)$. It is the information that an observer learns about the inputs by watching the execution of the protocol. The *internal information cost* [BYJKS04], [BBCR10] of the protocol is defined to be $I(X; M|YR) + I(Y; M|XR)$. It is the information learnt by the parties about each others inputs during the execution of the protocol. The external information is always at least as large as the internal information.

[1] The statement in [BBCR10] is seemingly stronger than is written here (the assumption there was $T \log T \ll C\sqrt{n}$). This difference arises from a different definition of *success* for protocols. Roughly speaking, here we require f to be determined by the messages and the public randomness, whereas [BBCR10] allowed each player separately to also use her input in determining f. If one uses the definition from [BBCR10] then direct product fails (for non-boolean relations), as Feige's counterexample (discussed in Appendix **??**) shows. However, the proof of [BBCR10] with the definition we use here yields a quantitatively weaker direct sum theorem, as stated above. We note that for Boolean functions, the two definitions of success are equivalent up to 1 additive bit of communication, as one party can always write the output $\pi(x, y)$ using one bit at the end of the protocol.

The first step of the reduction in [BBCR10] gives a protocol with internal information cost bounded by $\sim T/n$ and communication bounded by T. In the second step, they show that any protocol with internal information I and communication N can be compressed to get a protocol with communication $\sim \sqrt{I \cdot N}$. Thus one obtains a protocol with communication $\sim T/\sqrt{n}$ for computing f. When μ is a product distribution, the first step of the reduction gives a protocol with external information cost bounded by $\sim T/n$. They show how to compress any protocol with small external information almost optimally, and so obtain a protocol with communication $\sim T/n$ for computing f. In both cases, the intuition for the first step of the reduction is that the T bits of the messages can reveal at most $\sim T/n$ bits of information about an average input coordinate.

To prove our direct product theorems, we modify the approach above using ideas inspired by the proof of the parallel repetition theorem [Raz98]. Let E be the event that π correctly computes f^n. For $i \in [n]$, let W_i denote the event that the protocol π correctly computes $f(x_i, y_i)$. Let $\pi(E)$ denote the probability of E, and let $\pi(W_i|E)$ denote the conditional probability of the event W_i given E. We shall prove that if $\pi(E)$ is not very small, then $(1/n)\sum_i \pi(W_i|E) < 1$, which is a contradiction (since $\pi(W_i|E) = 1 \; \forall \; i$). In fact, we shall prove that this holds for an arbitrary event W, not just E.

Lemma 3 (Main Lemma)**.** *There is a universal constant $\alpha > 0$ so that the following holds. For every $\gamma > 0$, and event W such that $\pi(W) \geq 2^{-\gamma^2 n}$, if $\|\pi\| \geq 2$, and $\|\pi\| \log^{3/2}\|\pi\| < \alpha\gamma^{5/2}(C-1)\sqrt{n}$, then $(1/n)\sum_{i\in[n]} \pi(W_i|W) \leq \mathsf{suc}(\mu, f, C) + \gamma/\alpha$.*

Lemma 4 (Main Lemma for Product Distributions)**.** *There is a universal constant $\alpha > 0$ such that if μ is a product distribution, the following holds. For every $\gamma > 0$, and event W such that $\pi(W) \geq 2^{-\gamma^2 n}$, if $\|\pi\| \geq 2$, and $\|\pi\| \log^2\|\pi\| \leq \alpha\gamma^6 Cn$, then $(1/n)\sum_{i\in[n]} \pi(W_i|W) \leq \mathsf{suc}(\mu, f, C) + \gamma/\alpha$.*

The proofs of the lemmas proceed by reduction, and can be broken up into two steps as in [BBCR10]. However there are substantial differences in our proof, which are discussed in detail below. First let us see how Lemma 3 implies Theorem 1. Theorem 2 follows from Lemma 4 in the same way.

Proof of Theorem 1: Let E denote the event that π computes f correctly in all n coordinates. So, $(1/n)\sum_{i\in[n]} \pi(W_i|E) = 1$. Set $\gamma = \alpha(1 - \mathsf{suc}(\mu, f, C))/2$ so that $\mathsf{suc}(\mu, f, C) + \gamma/\alpha < 1$. Then by Lemma 3, either $\|\pi\| < 2$, $\|\pi\| \log^{3/2}\|\pi\| \geq \alpha^{7/2}2^{-5/2}(1-\mathsf{suc}(\mu, f, C))^{5/2}C\sqrt{n}$, or $\pi(E) < 2^{-\gamma^2 n}$. ■

Due to space constraints, we leave out the formal proofs of the main lemmas (these can be found in Section 3 in the full version of this paper [BRWY12]). At a high level, the proofs of the lemmas are quite similar to each other, though there are some technical differences. We discuss Lemma 4 first, which avoids some complications that come from the fact that the inputs are correlated under μ. We give a protocol with communication complexity C that computes f correctly with probability at least $(1/n)\sum_i \pi(W_i|W) - O(\gamma)$. Let m denote the messages of π, and $\pi(x_i y_i m)$ denote the joint distribution of x_i, y_i, m. For fixed x_i, y_i, let $\pi(m|x_i y_i W)$ denote the conditional distribution of m.

Using standard subadditivity based arguments, one can show that for average i, $\pi(x_i y_i|W) \overset{\gamma}{\approx} \pi(x_i y_i) = \mu(x_i y_i)$, where here the approximation is in terms of the ℓ_1 distance of the distributions. Intuitively, since W has probability $2^{-\gamma^2 n}$, it cannot significantly alter all n of the inputs. We can hope to obtain a protocol that computes $f(x, y)$ by picking a random i, setting $x_i = x, y_i = y$ and simulating the execution of π conditioned on the event W. There are two challenges that need to be overcome:

1) **The protocol must simulate** $\pi(m|x_i y_i W)$**.** In the probability space of π conditioned on W, the messages sent by the first party can become correlated with the input of the second party, even though they were initially independent. Thus (unlike in [BBCR10]), $\pi(m|x_i y_i W)$ is no longer distributed like the messages of a communication protocol, and it is non-trivial for the parties to sample a message from this distribution.
2) **The protocol must communicate at most** $C \ll |m|$ **bits.** To prove the lemma, the parties need to sample m using communication that is much smaller than the length of m.

To solve the first challenge, we use a protocol θ. The parties publicly sample a uniformly random coordinate i in $[n]$ and set $x_i = x, y_i = y$. They also publicly sample a variable r_i that contains a subset of the variables $x_1, \ldots, x_n, y_1, \ldots, y_n$. Each message m_j sent by the first party in π is sampled according to the distribution $\pi(m_j|m_{<j} x_i r_i W)$, and each message sent by the second party is sampled according to the distribution $\pi(m_j|m_{<j} y_i r_i W)$. We prove that for average i, $\theta(x_i y_i r_i m) \overset{\gamma}{\approx} \pi(x_i y_i r_i m|W)$. [JPY12] analyzed a different protocol θ, which used a different definition of r_i, and showed that for average i, $\theta(x_i y_i r_i m) \overset{\gamma t}{\approx} \pi(x_i y_i r_i m|W)$, where here t is the number of rounds

of communication in π. Our bound is independent of t, a feature that is essential to our results. A crucial technical feature of our protocol is the definition of r_i, which allows us to split the dependencies between inputs to π in a new way. This allows us to control the effect of the dependencies introduced by W using a bound that is independent of the number of rounds in π.

To solve the second challenge, we need to come up with a way to *compress* the protocol θ. To use the compression methods of [BBCR10], we need to bound the *external information cost* of θ. We did not succeed in bounding this quantity, and so cannot apply the compression methods of [BBCR10] directly. Instead, we are able to bound $I_\pi(X_iY_i; M|W)$ for average i, the corresponding quantity for the variables in the probability space of π.

This does not show that the information cost of θ is small, even though the distribution of the variables in θ is close in ℓ_1 distance to the distribution of the corresponding variables of π conditioned on W. For example, suppose θ is such that with small probability the first party sends her own input, and otherwise she sends a random string. Then θ is close to a protocol that reveals 0 information, but its information cost may be arbitrarily large.

Nevertheless, we show that any protocol that is close to having small external information cost can be simulated by a protocol that actually has small external information cost. In our example from above, the first party can simulate the protocol θ bit by bit and decide to abort it if she sees that her transmissions are significantly correlated with her input. This does not change the protocol most of the time, but does significantly reduce the amount of information that is revealed. Our general solution is very similar to this. The parties simulate θ and abort the simulation if they find that they are revealing too much information. We prove that any protocol that is close to having low information can be simulated with small communication (the term "δ-simulates" in the theorem statement is formally defined and discussed in Section 2 of the full version of this paper):

Theorem 5 (Simulation for External Information)**.** *Suppose θ is a protocol with inputs x, y, public randomness r, and messages m, and q is another distribution on these variables such that $\theta(xyrm) \overset{\epsilon}{\approx} q(xyrm)$. Then, there exists a protocol τ that strongly $O(\epsilon)$-simulates θ with $\|\tau\| \leq 2\|\theta\|$ and*

$$\begin{aligned}
&I_\tau(XY; M|R) \leq \\
&2\left(\frac{I_q(XY; M|R) + 1/(e\ln 2) + 2\log(\|\theta\|+1)}{\epsilon}\right) + \\
&+ \log(\|\theta\|+1) + 2\log(1/\epsilon) + 4.
\end{aligned}$$

Again, due to space constraints, we defer the formal proof of Theorem 5 to the full version of this paper (see section 4 in [BRWY12]). The final protocol computing f is obtained by compressing τ using the methods of [BBCR10].

The high level outline of the proof of Lemma 3 is similar to the proof of Lemma 4. When μ is not a product distribution, we obtain a bound on the internal information cost associated with π conditioned on W, namely we bound $I_\pi(X_i; M|Y_iR_iW) + I_\pi(Y_i; M|X_iR_iW)$. Proving an analogue of Theorem 5 for the *internal* information cost is beyond the reach of this paper (and it remains an interesting open question whether such a theorem is true or not). Instead, to prove Lemma 3, we reanalyze the compression method of [BBCR10] for internal information cost, and show that it can be used here. We prove:

Theorem 6 (Compression for Internal Information)**.** *Suppose θ is a protocol so that $\|\theta\| \geq 2$ with inputs x, y and messages m, and q is another distribution on these variables such that $\theta(xym) \overset{\epsilon}{\approx} q(xym)$. Let $I_q := I_q(X; M|Y) + I_q(Y; M|X)$. Then, there exists a protocol τ that $O(\epsilon)$-simulates θ such that*

$$\|\tau\| \leq \frac{\log\|\theta\|\sqrt{(I_q + 1 + \log\|\theta\|)\cdot\|\theta\|}}{\epsilon^{3/2}}.$$

Remark 7. *Theorem 6 can also be used to compress protocols θ that have public randomness. Indeed if the inputs are x', y', the public randomness is r and the messages are m, one can set $x = x'r, y = y'r$. Then $I_q(X; M|Y) + I_q(Y; M|X) = I_q(X'; M|Y'R) + I_q(Y'; M|X'R)$, so one can apply the theorem.*

The intuition for the proof is quite similar to the intuition for the proof of Theorem 5. We show that the compression goes well most of the time, and there is a small probability that the messages of the protocol will lead to a failure in the simulation, but this does not affect the outcome of the simulation by much. A formal proof of Theorem 6 can be found in Section 4 of the full version of this paper [BRWY12].

II. Organization

In Section III we introduce notations, definitions and technical claims which are used throughout the paper.

In Section IV we prove Lemma 3 (and outline the proof of Lemma 4). Due to space constraints, the proof of our simulation/compression results (Lemma 5 and Lemma 6) are deferred to the full version of the paper. We conclude the paper with discussion and an interesting related open problem in Section V.

III. Preliminaries

A. Notation

Unless otherwise stated, logarithms in this text are computed base two. Random variables are denoted by capital letters and values they attain are denoted by lower-case letters. For example, A may be a random variable and then a denotes a value A may attain and we may consider the event $A = a$. Given $a = a_1, a_2, \dots, a_n$, we write $a_{\leq i}$ to denote $a_1, \dots, a_i$. We define $a_{>i}$ and $a_{\leq i}$ similarly.

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 usually be clear from context, but in cases where there may be confusion we shall be more explicit about which meaning is being used. We write $p(a|b)$ to denote either the distribution of A conditioned on the event $B = b$, or the number $\Pr[A = a|B = b]$. Again, the meaning will usually be clear from context. 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 W is an event, we write $p(W)$ to denote its probability according to p. We denote by $\mathbb{E}_{p(a)}[g(a)]$ the expected value of $g(a)$ with respect to a distributed according to p.

For two distributions p, q, we write $|p(a) - q(a)|$ to denote the ℓ_1 distance between the distributions p and q. We write $p \overset{\epsilon}{\approx} q$ if $|p - q| \leq \epsilon$. Given distributions $p_1, \dots, p_n$ and $q_1, \dots, q_n$, we sometimes say "in expectation over i sampled according to $\eta(i)$, $p_i \overset{\gamma}{\approx} q_i$" when we mean that $\mathbb{E}_{\eta(i)}[|p_i - q_i|] \leq \gamma$.

The *divergence* between p, q is defined to be

$$\mathsf{D}\left(\frac{p(a)}{q(a)}\right) = \sum_a p(a) \log \frac{p(a)}{q(a)}.$$

For three random variables A, B, C with underlying probability distribution $p(a, b, c)$, the *mutual information* between A, B conditioned on C is defined as

$$I_p(A; B|C) = \mathop{\mathbb{E}}_{p(cb)}\left[\mathsf{D}\left(\frac{p(a|bc)}{p(a|c)}\right)\right] =$$
$$= \mathop{\mathbb{E}}_{p(ca)}\left[\mathsf{D}\left(\frac{p(b|ac)}{p(b|c)}\right)\right] = \sum_{a,b,c} p(abc) \log \frac{p(a|bc)}{p(a|c)}.$$

We shall often work with multiple distributions over the same space. To avoid confusion, we shall always explicitly specify the distribution being used when computing the mutual information. We shall sometimes work with an event W. In this case, we denote $I_p(A; B|CW) = I_q(A; B|C)$ where $q(abc) = p(abc|W)$.

B. Communication Complexity

Given a protocol π that operates on inputs x, y drawn from a distribution μ using public randomness[2] r and messages m, we write $\pi(xymr)$ to denote the joint distribution of these variables. We write $\|\pi\|$ to denote the *communication complexity* of π, namely the maximum number of bits that may be exchanged by the protocol.

Our work relies heavily on ways to measure the information complexity of a protocol (see [BBCR10], [Bra12] and references within for a more detailed overview). The *internal information cost* of π is defined to be $I_\pi(X; M|YR) + I_\pi(Y; M|XR)$. The *external information cost* is $I_\pi(XY; M|R)$. The internal information cost is always at most the external information cost, and the two measures are equal when $\pi(xy) = \pi(x)\pi(y)$ is a product distribution. Both measures are at most the communication complexity of the protocol.

Let $q(x, y, a)$ be an arbitrary distribution. We say that π *δ-simulates* q, if there is a function g and a function h such that

$$\pi(x, y, g(x, r, m), h(y, r, m)) \overset{\delta}{\approx} q(x, y, a, a),$$

where $q(x, y, a, a)$ is the distribution on 4-tuples (x, y, a, a) where (x, y, a) are distributed according to q. Thus if π δ-simulates q, the protocol allows the parties to sample a according to $q(a|xy)$. If in addition $g(x, r, m)$ does not depend on x, we say that π *strongly* δ-simulates q. Thus if π strongly simulates q, then the outcome of the simulation is apparent even to an observer that does not know x or y.

If λ is a protocol with inputs x, y, public randomness r' and messages m', we say that π δ-simulates λ if π δ-simulates $\lambda(x, y, (r', m'))$. Similarly, we say that π strongly δ-simulates λ if π strongly δ-simulates $\lambda(x, y, (r', m'))$. We say that π computes f with success probability $1 - \delta$, if π strongly δ-simulates $\pi(x, y, f(x, y))$.

The following lemma will be useful in our simulation protocols. It shows that messages sent by each party

[2] In our paper we define protocols where the public randomness is sampled from a continuous (i.e. non-discrete) set. Nevertheless, we often treat the randomness as if it were supported on a discrete set, for example by taking the sum over the set rather than the integral. This simplifies notation throughout our proofs, and does not affect correctness in any way, since all of our public randomness can be approximated to arbitrary accuracy by sufficiently dense finite sets.

remain independent of the other party's input even after some part of the input is fixed. A formal proof can be found in the full version of this paper.

Lemma 8. *Let x, y be inputs to a protocol π with public randomness r and let r' be a variable such that $\pi(xy|rr') = \pi(x|rr')\pi(y|rr')$. Let $m_1, \ldots, m_j$ be messages in π such that m_j is transmitted by Alice. Then $\pi(m_j|m_{<j}rr') = \pi(m_j|m_{<j}rr'y)$.*

C. Useful Protocols

The following lemma was proved by Holenstein [Hol07].

Lemma 9 (Correlated Sampling)**.** *Suppose Alice is given a distribution p and Bob a distribution q over a common universe. Then there is a randomized sampling procedure that allows Alice and Bob to use shared randomness to jointly sample elements A, B such that A is distributed according to p, B is distributed according to q, and $\Pr[A \neq B] = |p - q|$.*

The following compression theorem from [BBCR10] will be useful:

Theorem 10. *For every protocol π, and every $\epsilon > 0$, there exists a protocol λ that strongly ϵ-simulates π with*

$$\|\lambda\| \leq O\left(\frac{I_\pi(XY; M|R) \cdot \log(\|\pi\|/\epsilon)}{\epsilon^2}\right).$$

D. Basic Lemmas

The proofs of the following two lemmas can be found in [CT91]:

Lemma 11 (Divergence is Non-negative)**.** $\mathsf{D}\left(\frac{p(a)}{q(a)}\right) \geq 0$.

Lemma 12 (Chain Rule)**.** *If $a = a_1, \ldots, a_s$, then*

$$\mathsf{D}\left(\frac{p(a)}{q(a)}\right) = \sum_{i=1}^{s} \mathop{\mathbb{E}}_{p(a_{<i})}\left[\mathsf{D}\left(\frac{p(a_i|a_{<i})}{q(a_i|a_{<i})}\right)\right].$$

In this section we give some basic lemmas which will be used repeatedly throughout the rest of the paper. Due to lack of space, we omit proofs. For the formal proofs see the full version of this paper [BRWY12].

Pinsker's inequality bounds statistical distance in terms of the divergence:

Lemma 13 (Pinsker)**.** *If $p(b) = q(b)$, then*

$$|p(a,b) - q(a,b)|^2 \leq \mathop{\mathbb{E}}_{p(b)}\left[\mathsf{D}\left(\frac{p(a|b)}{q(a|b)}\right)\right].$$

The following lemma bounds the probability of getting a large term in the divergence:

Lemma 14 (Reverse Pinsker)**.** *Let $S = \left\{(a,b) : \log\frac{p(a|b)}{q(a|b)} > 1\right\}$. Then, $p(S) < 2|p(a,b) - q(a,b)|$.*

The following bounds the contribution of the negative terms to the divergence:

Lemma 15. *Let $S = \{a : p(a) < q(a)\}$. Then, $\sum_{a \in S} p(a) \log\frac{p(a)}{q(a)} \geq -1/(e \ln 2)$.*

E. Inequalities that Involve Conditioning

The following lemmas bound the change in divergence when extra conditioning is involved. Due to lack of space we omit all proofs. We note that these claims are central to our results and are used in a subtle way, and we encourage the reader to consult the full version of this paper for the complete proofs.

Lemma 16. *Let W be an event and A, B, M be random variables in the probability space p. Then,*

$$\mathop{\mathbb{E}}_{p(bm|W)}\left[\mathsf{D}\left(\frac{p(a|bmW)}{p(a|b)}\right)\right] \leq$$
$$\log\frac{1}{p(W)} + I_p(A; M|BW).$$

Lemma 17 (Conditioning does not decrease divergence)**.**

$$\mathop{\mathbb{E}}_{p(b)}\left[\mathsf{D}\left(\frac{p(a|b)}{q(a)}\right)\right] \geq \mathsf{D}\left(\frac{p(a)}{q(a)}\right).$$

The following lemma gives a key estimate that is used crucially in our proof. It allows us to remove the effect of conditioning on an event W on the second argument of a divergence expression. The lemma states that, on average, $\mathsf{D}\left(\frac{p(a|brW)}{p(a|rW)}\right)$ cannot be larger than $\mathsf{D}\left(\frac{p(a|brW)}{p(a|r)}\right)$. Intuitively this is true because in both cases the first distribution is conditioned on W, but in the second case the second distribution is not conditioned on W. The second part of the lemma shows that conditioning on an event W of probability 2^{-s} can create a mutual information of up to s between two formerly independent random variables.

Lemma 18. *Let W be an event and A, B, R be random variables. Then,*

$$I_p(A; B|RW) \leq \mathop{\mathbb{E}}_{p(br|W)}\left[\mathsf{D}\left(\frac{p(a|brW)}{p(a|r)}\right)\right].$$

If in addition $p(abr) = p(r)p(a|r)p(b|r)$, then

$$I_p(A; B|RW) \leq$$
$$\leq \mathop{\mathbb{E}}_{p(br|W)}\left[\mathsf{D}\left(\frac{p(a|brW)}{p(a|br)}\right)\right] \leq \log\frac{1}{p(W)}.$$

F. Variable Truncation

We shall need to analyze protocols that are statistically close to having low information. The following lemmas show that if a variable A is statistically close to having low information, then some prefix $A_{\leq K}$ of A usually has low information. By truncating the variable to $A_{\leq K}$, we obtain a new variable that is statistically close to the old one, yet has low information. We defer the proof to the full version of this paper.

Lemma 19 (Truncation Lemma). *Let* $p(a,b,c) \overset{\epsilon}{\approx} q(a,b,c)$ *where* $a = a_1,\ldots,a_s$. *For every* a,b,c, *define* k *to be the minimum number* j *in* $[s]$ *such that*

$$\log \frac{p(a_{\leq j}|bc)}{p(a_{\leq j}|c)} > \beta.$$

If no such index exists, set $k = s+1$. *Then,*

$$p(k < s+1) \\ < \frac{I_q(A;B|C) + \log(s+1) + 1/(e\ln 2)}{\beta - 2} + 9\epsilon/2.$$

Remark 20. *One can also prove that* $I_p(A_{<K}, B|C) \leq \beta + \log(s+1)$, *in Lemma 19. We do not need this conclusion, so we omit its proof.*

IV. Proof Outline of the Main Lemma

In this section we give a more detailed outline of the proof of Lemma 3, though still leaving out most of the technical proofs. Lemma 4 is proved in a similar fashion. The formal proofs for all the claims written below can be found in the full version of this paper [BRWY12].

We write $M = M_1, M_2, \ldots, M_{2t}$ to denote the messages in π. Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be the inputs. We write $\overline{X} = X_1,\ldots,X_n$ and $\overline{Y} = Y_1,\ldots,Y_n$. Without loss of generality, we assume that n is even.

Consider the protocol η in Figure 1. We show that η computes f with good probability, although with a lot of communication. The protocol η has public randomness $i,\mathbf{g},\mathbf{h}$ and runs protocol $\theta_{i,\mathbf{g},\mathbf{h}}$ given in Figure 2 as a subroutine with inputs $(x_i, r'_{i,\mathbf{g},\mathbf{h}}), (y_i, r''_{i,\mathbf{g},\mathbf{h}})$. Eventually, we shall argue that in expectation over $i,\mathbf{g},\mathbf{h}$ sampled according to $\eta(i\mathbf{gh})$,

$$\eta\big((x_i, r'_{i,\mathbf{g},\mathbf{h}}), (y_i, r''_{i,\mathbf{g},\mathbf{h}})\big) \overset{O(\gamma)}{\approx} \\ \theta_{i,\mathbf{g},\mathbf{h}}((x_i, r'_{i,\mathbf{g},\mathbf{h}}), (y_i, r''_{i,\mathbf{g},\mathbf{h}})),$$

and that, on average, $\theta_{i,\mathbf{g},\mathbf{h}}$ is statistically close to having small internal information, and statistically close to having small external information in the case that μ is product. We shall apply Theorem 6 to compress the communication so as to obtain our final protocol for computing f and conclude the proof of Lemma 3 (Similarly, Theorem 5 and Theorem 10 are used to obtain the protocol that proves Lemma 4).

Our first goal is to show that conditioning on the event W does not change the distribution in a typical coordinate. The following lemma is rather standard and follows from subadditivity of divergence and its relation to the ℓ_1 norm (Pinsker's inequality):

Lemma 21. *In expectation over* i *sampled according to* $\eta(i)$, $\pi(x_iy_i) \overset{\gamma}{\approx} \pi(x_iy_i|W)$.

Next we eliminate a corner case:

Lemma 22. *If* $\|\pi\| \leq \gamma^2 n$, *then in expectation over* i *sampled according to* $\eta(i)$, $\pi(mx_iy_i|W) \overset{\sqrt{2}\gamma}{\approx} \pi(m|W) \cdot \pi(x_iy_i)$.

The proof of Lemma 22 is also a straightforward application of subadditivity. Lemma 22 implies that if $\|\pi\| \leq \gamma^2 n$, then a protocol with 0 communication can approximate the messages of π conditioned on W, and so compute f with 1 additional bit of communication. So

$$(1/n)\sum_{i=1}^{n} \pi(W_i|W) - \gamma/\sqrt{2} \\ \leq \mathsf{suc}(\mu, f, 1) \leq \mathsf{suc}(\mu, f, C),$$

which completes the proof. The more interesting case is when $\|\pi\| \geq \gamma^2 n$, and so we assume that this holds in the rest of this section.

Given subsets $\mathbf{g},\mathbf{h} \subset [n]$, let $\overline{X}_{\mathbf{h}}$ and $\overline{Y}_{\mathbf{g}}$ denote $\overline{X}$ and $\overline{Y}$ projected on to the relevant coordinates. Define

$$R_{i,\mathbf{g},\mathbf{h}} = \overline{X}_{\mathbf{h}\setminus\{i\}}, \overline{Y}_{\mathbf{g}\setminus\{i\}}.$$

The random variable $R_{i,\mathbf{g},\mathbf{h}}$ helps to break the dependencies between Alice and Bob.

It turns out that choosing the right distribution for $i,\mathbf{g},\mathbf{h}$ in η is crucial to our proofs. We need the distribution to be symmetric in $\mathbf{g},\mathbf{h}$. It is important that $\mathbf{g} \cup \mathbf{h} = [n]$ so that $x_i, y_i, r_{i,\mathbf{g},\mathbf{h}}$ split the dependences between $\overline{x}, \overline{y}$. In the analysis we shall repeatedly use the fact that for every fixing of $\mathbf{h}$, $\eta(i\mathbf{g}|\mathbf{h})$ has the property that i is distributed uniformly over a large set, and $i \in \mathbf{g} \cap \mathbf{h}$. This allows us to apply the chain rule. For more intuition on the choice of the variables $r_{i,\mathbf{g},\mathbf{h}}$, see Section 3.3 in [BRWY12].

Now we argue that $\eta(i\mathbf{gh})$ has the properties we need. Observe that we can sample $\eta(i\mathbf{gh})$ by the following different yet equivalent process. Let $\mathbf{h}$ be distributed as in η. For fixed $\mathbf{h}$, let $\kappa_{\mathbf{h}} : [n] \to [n]$ be a permutation sampled uniformly from the set of permutations that map $[|\mathbf{h}|]$ to $\mathbf{h}$. Let ℓ be a uniformly random

Protocol η for computing $f(x, y)$ when inputs are sampled according to μ.

1) Let s_h, s_g be uniformly random numbers from the set $\{n/2+1, \ldots, n\}$. Let $\kappa : [n] \to [n]$ be a uniformly random permutation. Set $\mathbf{h} = \kappa([s_h])$ and $\mathbf{g} = \kappa(\{n - s_g + 1, \ldots, n\})$. Let i be a uniformly random element of $\mathbf{g} \cap \mathbf{h}$ (which must be non-empty by the choice of s_g, s_h).
2) Alice sets $x_i = x$ and Bob sets $y_i = y$.
3) Alice and Bob use Lemma 9 to sample $r_{i,\mathbf{g},\mathbf{h}}$: Alice uses the distribution $\pi(r_{i,\mathbf{g},\mathbf{h}}|x_i W)$ and Bob uses the distribution $\pi(r_{i,\mathbf{g},\mathbf{h}}|y_i W)$. Write $r'_{i,\mathbf{g},h}$ to denote Alice's sample and $r''_{i,\mathbf{g},\mathbf{h}}$ to denote Bob's sample.
4) Alice and Bob run protocol $\theta_{i,\mathbf{g},\mathbf{h}}$ from Figure 2 with inputs $(x_i, r'_{i,\mathbf{g},\mathbf{h}})$ and $(y_i, r''_{i,\mathbf{g},\mathbf{h}})$.

Fig. 1. Protocol for computing f.

Protocol $\theta_{i,\mathbf{g},\mathbf{h}}$ for computing $f(x_i, y_i)$ when inputs $(x_i, r'_{i,\mathbf{g},\mathbf{h}}), (y_i, r''_{i,\mathbf{g},\mathbf{h}})$ are sampled according to $\pi((x_i, r_{i,\mathbf{g},\mathbf{h}}), (y_i, r_{i,\mathbf{g},\mathbf{h}})|W)$.

Alice sends each message M_j, j odd, according to the distribution $\pi(m_j|x_i r'_{i,\mathbf{g},\mathbf{h}} m_{<j} W)$. Bob sends each message M_j, j even, according to the distribution $\pi(m_j|y_i r''_{i,\mathbf{g},\mathbf{h}} m_{<j} W)$.

Fig. 2. Simulation in the i'th coordinate.

element of $[n/2]$. Given $\mathbf{h}, \kappa_{\mathbf{h}}, \ell$, set $i = \kappa_{\mathbf{h}}(\ell)$ and $\mathbf{g} = \kappa_{\mathbf{h}}(\{\ell, \ell+1, \ldots, n\})$. Then note that $\mathbf{g}, \mathbf{h}, i$ are distributed as defined in the protocol η. Further, note that $(i, x_i, r_{i,\mathbf{g},\mathbf{h}})$ and $(\kappa_{\mathbf{h}}(\ell), \overline{x}_{\mathbf{h}}, \overline{y}_{\kappa_{\mathbf{h}}(\{\ell+1,\ldots,n\})})$ determine each other.

The following lemma asserts that the distribution of the public randomness $R_{i,\mathbf{g},\mathbf{h}}$ of π doesn't change much when conditioning on W:

Lemma 23. *In expectation over $i, \mathbf{g}, \mathbf{h}$ sampled according to $\eta(i\mathbf{gh})$,*
$\pi(x_i y_i)\pi(r_{i,\mathbf{g},\mathbf{h}}|x_i W) \overset{3\gamma}{\approx} \pi(x_i y_i r_{i,\mathbf{g},\mathbf{h}}|W) \overset{3\gamma}{\approx} \pi(x_i y_i)\pi(r_{i,\mathbf{g},\mathbf{h}}|y_i W)$.

The following claim is the heart of the proof. It asserts that indeed the distribution $(\pi|W)$, on an average coordinate i, is well approximated by the protocol θ.

Claim 24. *In expectation over $i, \mathbf{g}, \mathbf{h}$ sampled according to $\eta(i\mathbf{gh})$,*

$$\theta_{i,\mathbf{g},\mathbf{h}}(x_i y_i r_{i,\mathbf{g},\mathbf{h}} m) \overset{2\gamma}{\approx} \pi(x_i y_i r_{i,\mathbf{g},\mathbf{h}} m|W).$$

Proof: Consider

$$\mathop{\mathbb{E}}_{\eta(i\mathbf{gh})}\left[\mathop{\mathbb{E}}_{\pi(x_i y_i r_{i,\mathbf{g},\mathbf{h}}|W)}\left[\mathsf{D}\left(\frac{\pi(m|x_i y_i r_{i,\mathbf{g},\mathbf{h}} W)}{\theta_{i,\mathbf{g},\mathbf{h}}(m|x_i y_i r_{i,\mathbf{g},\mathbf{h}})}\right)\right]\right]$$

$$= \sum_{j=1}^{2t} \mathbb{E}_{\eta(i\mathbf{gh})}\Bigg[$$

$$\mathop{\mathbb{E}}_{\pi(m_{<j} x_i y_i r_{i,\mathbf{g},\mathbf{h}}|W)}\left[\mathsf{D}\left(\frac{\pi(m_j|x_i y_i r_{i,\mathbf{g},\mathbf{h}} m_{<j} W)}{\theta_{i,\mathbf{g},\mathbf{h}}(m_j|x_i y_i r_{i,\mathbf{g},\mathbf{h}} m_{<j})}\right)\right]\Bigg] \quad (1)$$

The odd j's correspond to the cases when Alice speaks. These terms contribute:

$$\sum_{\text{odd } j} \mathop{\mathbb{E}}_{\eta(i\mathbf{gh})}\left[I_\pi(M_j; Y_i|X_i R_{i,\mathbf{g},\mathbf{h}} M_{<j} W)\right].$$

As in the proof of Lemma 23, we can express this as

$$\frac{2}{n}\sum_{\text{odd } j} \mathbb{E}_{\eta(\mathbf{h}\kappa_{\mathbf{h}})}\Bigg[$$

$$I_\pi(M_j; \overline{Y}_{\kappa_{\mathbf{h}}([n/2])}|\overline{X}_{\mathbf{h}}\overline{Y}_{\kappa_{\mathbf{h}}(\{n/2+1,\ldots,n\})} M_{<j} W)\Bigg]$$

by the chain rule. By Lemma 18, we can upper bound this by

$$\leq \frac{2}{n}\sum_{\text{odd } j} \mathbb{E}_{\eta(\mathbf{h}\kappa_{\mathbf{h}})}\Bigg[$$

$$\mathop{\mathbb{E}}_{\pi(m_{<j}\overline{x}_{\mathbf{h}}\overline{y}|W)}\left[\mathsf{D}\left(\frac{\pi(m_j|m_{<j}\overline{x}_{\mathbf{h}}\overline{y}W)}{\pi(m_j|m_{<j}\overline{x}_{\mathbf{h}}\overline{y}_{\kappa_{\mathbf{h}}(\{n/2+1,\ldots,n\})})}\right)\right]\Bigg].$$

Conditioned on $\overline{x}_{\mathbf{h}}\overline{y}_{\kappa_{\mathbf{h}}(\{n/2+1,\ldots,n\})}$, the inputs $\overline{x}, \overline{y}$ are independent. Thus Lemma 8 gives

$$\pi(m_j|m_{<j}\overline{x}_{\mathbf{h}}\overline{y}_{\kappa_{\mathbf{h}}(\{n/2+1,\ldots,n\})}) = \pi(m_j|m_{<j}\overline{x}_{\mathbf{h}}\overline{y}),$$

and we can continue to bound

$$= \frac{2}{n}\sum_{\text{odd } j} \mathbb{E}_{\eta(\mathbf{h}\kappa_{\mathbf{h}})}\Bigg[$$

$$\mathop{\mathbb{E}}_{\pi(m_{<j}\overline{x}_{\mathbf{h}}\overline{y}|W)}\left[\mathsf{D}\left(\frac{\pi(m_j|m_{<j}\overline{x}_{\mathbf{h}}\overline{y}W)}{\pi(m_j|m_{<j}\overline{x}_{\mathbf{h}}\overline{y})}\right)\right]\Bigg].$$

Since the divergence is always non-negative, we can add in the even terms in the sum over j to bound

$$\leq \frac{2}{n}\sum_{j=1}^{2t}\mathbb{E}_{\eta(\mathbf{h}\kappa_{\mathbf{h}})}\Bigg[$$

$$\mathop{\mathbb{E}}_{\pi(m_{<j}\overline{x}_{\mathbf{h}}\overline{y}|W)}\left[\mathsf{D}\left(\frac{\pi(m_j|m_{<j}\overline{x}_{\mathbf{h}}\overline{y}W)}{\pi(m_j|m_{<j}\overline{x}_{\mathbf{h}}\overline{y})}\right)\right]\Bigg]$$

$$= \frac{2}{n}\mathop{\mathbb{E}}_{\eta(\mathbf{h}\kappa_{\mathbf{h}})}\left[\mathop{\mathbb{E}}_{\pi(\overline{x}_{\mathbf{h}}\overline{y}|W)}\left[\mathsf{D}\left(\frac{\pi(m|\overline{x}_{\mathbf{h}}\overline{y}W)}{\pi(m|\overline{x}_{\mathbf{h}}\overline{y})}\right)\right]\right]$$

(by the chain rule)

$$\leq \frac{2}{n}\mathop{\mathbb{E}}_{\eta(\mathbf{h}\kappa_{\mathbf{h}})}\left[\gamma^2 n\right] = 2\gamma^2,$$

by Lemma 16. Repeating the same argument for even j gives (1) $\leq 4\gamma^2$. We apply Lemma 13 to conclude the proof. ■

A. Completing the Proof of Lemma 3

Claim 25. *The expected value of the expression for the internal information cost according to π conditioned on W can be bounded:*

$$\mathbb{E}_{\eta(i\mathbf{gh})}[(I_\pi(X_i;M|Y_iR_{i,\mathbf{g},\mathbf{h}}W)+ I_\pi(Y_i;M|X_iR_{i,\mathbf{g},\mathbf{h}}W))] \leq 4\|\pi\|/n.$$

In the probability space of π, let $i,\mathbf{g},\mathbf{h}$ be independent of all other variables, and distributed as in η. Let $x' = (i,\mathbf{g},\mathbf{h},x_i,r_{i,\mathbf{g},\mathbf{h}})$ and $y' = (i,\mathbf{g},\mathbf{h},y_i,r_{i,\mathbf{g},\mathbf{h}})$. Define the protocol θ that gets inputs $(i,\mathbf{g},\mathbf{h},x_i,r'_{i,\mathbf{g},\mathbf{h}})$ and $(i,\mathbf{g},\mathbf{h},y_i,r''_{i,\mathbf{g},\mathbf{h}})$, where the inputs are distributed according to

$$\pi((i,\mathbf{g},\mathbf{h},x_i,r_{i,\mathbf{g},\mathbf{h}}),(i,\mathbf{g},\mathbf{h},y_i,r_{i,\mathbf{g},\mathbf{h}})|W),$$

and executes $\theta_{i,\mathbf{g},\mathbf{h}}((x_i,r'_{i,\mathbf{g},\mathbf{h}}),(y_i,r''_{i,\mathbf{g},\mathbf{h}}))$.

By Lemma 9 and Lemma 23, $\Pr_\eta[R'_{i,\mathbf{g},\mathbf{h}} \neq R''_{i,\mathbf{g},\mathbf{h}}] \leq O(\gamma)$. Thus in expectation over $i,\mathbf{g},\mathbf{h}$ sampled according to $\eta(i\mathbf{gh})$,

$$\eta((x_i,r'_{i,\mathbf{g},\mathbf{h}}),(y_i,r''_{i,\mathbf{g},\mathbf{h}})) \overset{O(\gamma)}{\approx} \eta((x_i,r'_{i,\mathbf{g},\mathbf{h}}),(y_i,r'_{i,\mathbf{g},\mathbf{h}})),$$

where here $\eta((x_i,r'_{i,\mathbf{g},\mathbf{h}}),(y_i,r'_{i,\mathbf{g},\mathbf{h}}))$ denotes the distribution where Bob's sample for $r''_{i,\mathbf{g},\mathbf{h}}$ is set to be the same as Alice's sample. By Lemma 23 and Lemma 21,

$$\eta(i\mathbf{gh}xyr'_{i,\mathbf{g},\mathbf{h}}) \overset{O(\gamma)}{\approx} \pi(i\mathbf{gh}x_iy_ir_{i,\mathbf{g},\mathbf{h}}|W).$$

Therefore the protocol η can be viewed as executing θ as a subroutine with inputs that are $O(\gamma)$-close to $\theta(x',y')$. Claim 24 implies that $\theta(x'y'm) \overset{O(\gamma)}{\approx} \pi(x'y'm|W)$. Claim 25 implies that

$$\begin{aligned} &I_\pi(X';M|Y'W)+I_\pi(Y';M|X'W)\\ &= \mathop{\mathbb{E}}_{\eta(i\mathbf{gh})}[I_\pi(X_i;M|Y_iR_{i,\mathbf{g},\mathbf{h}}W)+I_\pi(Y_i;M|X_iR_{i,\mathbf{g},\mathbf{h}}W)]\\ &\leq 4\|\pi\|/n \quad \text{(since } \|\pi\| \geq \gamma^2 n\text{)}. \end{aligned}$$

To prove Lemma 3, we apply Theorem 6 to conclude that there exists a protocol that $O(\gamma)$-simulates θ with communication at most

$$\frac{\log\|\pi\|\sqrt{(4\|\pi\|/n+1+\log\|\pi\|)\|\pi\|}}{\gamma^{3/2}}$$

$$< O\left(\frac{\|\pi\|\cdot\log^{3/2}\|\pi\|}{\sqrt{n}\gamma^{5/2}}\right) < C-1,$$

where the first inequality appealed to the fact that $\|\pi\|/n > \gamma^2$ and the second is by our choice of α in the statement of Lemma 3. The proof of Lemma 3 is complete, since with one additional bit of communication to send the value of f, the protocol η computes f with probability of success at least $(1/n)\sum_{i=1}^n \pi(W_i|W) - O(\gamma)$. □

V. Open Problem: Direct Products for Information Complexity

Both the direct sum result of [BBCR10] and our direct product result rely on methods to compress protocols. So it is natural to ask whether our ability to prove direct product results is limited only by our ability to compress protocols with low information cost. In fact, information cost can be made into a meaningful complexity measure. The *information complexity* of a function f with respect to a distribution μ is the lowest internal information cost attainable by a protocol computing f with respect to μ and error $1/3$ [BR11], [Bra12]. It turns out that the amortized communication complexity of f is exactly equal to its information complexity [BR11]. [BW11], [KLL$^+$12] showed that many communication lower bound techniques actually give lower bounds on the information complexity.

Given this new complexity measure, we might have hoped that direct sum and direct product theorems holds with respect to it. Indeed [BBCR10] show that an optimal direct sum theorem holds for information complexity. However, a direct product theorem (with small success probability) cannot hold, because of the following counterexample. Let f be a function with information complexity I. Consider the protocol that computes f^n as follows. Let $\epsilon > 0$ be an arbitrary parameter. With probability ϵ, the protocol executes n copies of the optimal protocol for computing f. With probability $1-\epsilon$ the protocol transmits nothing and fails.

This protocol computes f^n with probability ϵ, yet its information complexity is at most $\epsilon I n$. For example, setting $\epsilon = 1/n$ shows that even without increasing the information complexity, one can compute f^n with success probability $1/n$.

The following question is still interesting, and may be easier than proving new direct product results for communication complexity:

Open Problem 26. *Let* $\gamma = 1 - \mathsf{suc}(\mu, f, C)$. *Is there a universal constant* α *such that if the information complexity of* f *with respect to the distribution* μ *is* I, $T \geq 2$, *and* $T < \alpha I n$, *then* $\mathsf{suc}(\mu^n, f^n, T) \leq \exp\left(-\alpha\gamma^2 n\right)$*?*

Acknowledgments and authors credits

We thank Penghui Yao for finding an error in an earlier version of this paper. We thank Paul Beame, Ankit Garg, and David Xiao for useful comments.

[*] Research supported in part by an Alfred P. Sloan Fellowship, an NSF CAREER award, and a Turing Centenary Fellowship.

[†] 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.

[‡] Supported by a Simons Fellowship in Theoretical Computer Science.

[§] Horev Fellow – supported by the Taub Foundation. Supported by the Israel Science Foundation and by the Binational Science Foundation under agreement 2010089.

References

[Abl96] F. Ablayev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoretical Computer Science*, 157(2):139–159, 1996.

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 2010 ACM International Symposium on Theory of Computing*, pages 67–76, 2010.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In Rafail Ostrovsky, editor, *FOCS*, pages 748–757. IEEE, 2011.

[Bra12] Mark Braverman. Interactive information complexity. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 505–524, New York, NY, USA, 2012. ACM.

[BRWY12] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:143, 2012.

[BW11] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:164, 2011.

[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.

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

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley series in telecommunications. J. Wiley and Sons, New York, 1991.

[Fei00] Uriel Feige. Error reduction by parallel repetition – the state of the art, June 20 2000.

[FPRU94] Uriel Feige, David Peleg, Prabhakar Raghavan, and Eli Upfal. Computing with noisy information. *SIAM Journal on Computing*, 23(5):1001–1018, 1994.

[Hol07] Thomas Holenstein. Parallel repetition: Simplifications and the no-signaling case. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, 2007.

[JPY12] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for bounded-round public-coin randomized communication complexity. *CoRR*, abs/1201.1666, 2012.

[JY12] Rahul Jain and Penghui Yao. A strong direct product theorem in terms of the smooth rectangle bound. *CoRR*, abs/1209.0263, 2012.

[Kla10] Hartmut Klauck. A strong direct product theorem for disjointness. In *STOC*, pages 77–86, 2010.

[KLL$^+$12] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:38, 2012.

[LSS08] Troy Lee, Adi Shraibman, and Robert Spalek. A direct product theorem for discrepancy. In *CCC*, pages 71–80, 2008.

[MWY13] Marco Molinaro, David Woodruff, and Grigory Yaroslavtsev. Beating the direct sum theorem in communication complexity with implications for sketching. In *SODA*, page to appear, 2013.

[NW93] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM Journal on Computing*, 22(1):211–219, February 1993.

[PRW97] Itzhak Parnafes, Ran Raz, and Avi Wigderson. Direct product results and the GCD problem, in old and new communication models. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computing (STOC '97)*, pages 363–372, New York, May 1997. Association for Computing Machinery.

[Raz92] Razborov. On the distributed complexity of disjointness. *TCS: Theoretical Computer Science*, 106, 1992.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998. Prelim version in STOC '95.

[Sha48] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27, 1948. Monograph B-1598.

[Sha03] Ronen Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1-2):1–22, 2003. Prelim version CCC 2001.

[She11] Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. In *STOC*, pages 41–50, 2011.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *STOC*, pages 209–213, 1979.