XOR Lemmas for Communication via Marginal Information*

Siddharth Iyer
University of Washington
Seattle, USA
siyer@cs.washington.edu

Anup Rao
University of Washington
Seattle, USA
anuprao@cs.washington.edu

ABSTRACT

We define the *marginal information* of a communication protocol, and use it to prove XOR lemmas for communication complexity. We show that if every C-bit protocol has bounded advantage for computing a Boolean function f, then every $\widetilde{\Omega}(C\sqrt{n})$-bit protocol has advantage $\exp(-\Omega(n))$ for computing the n-fold xor $f^{\oplus n}$. We prove exponentially small bounds in the average case setting, and near optimal bounds for product distributions and for bounded-round protocols.

CCS CONCEPTS

• **Theory of computation** → **Communication complexity**.

KEYWORDS

Communication Complexity, XOR Lemma

ACM Reference Format:
Siddharth Iyer and Anup Rao. 2024. XOR Lemmas for Communication via Marginal Information. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC '24), June 24–28, 2024, Vancouver, BC, Canada.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3618260.3649726

1 INTRODUCTION

If a function is hard to compute, is it even harder to compute it many times? This old question is often challenging, and new answers are usually accompanied by foundational ideas. We give new answers in the framework of communication complexity, accompanied by a new measure of complexity called *marginal information*. This definition provides a new tool for proving lower bounds in theoretical computer science.

A wide variety of important lower bounds in computer science ultimately rely on information theoretic lower bounds in communication complexity, including lower bounds on the depth of monotone circuits [17], lower bounds on data structures [19] and lower bounds on the extension complexity of polytopes [3, 16, 26, 30], to name a few nice examples. We refer the reader to the textbook [23] for an introduction to the basic definitions and concepts in communication complexity, the role played by the questions we address here, and the connections to other areas.

*This work was supported by NSF award 2131899. The full version of this paper is available at https://arxiv.org/abs/2312.03076.

STOC '24, June 24–28, 2024, Vancouver, BC, Canada

ACM ISBN 979-8-4007-0383-6/24/06.
https://doi.org/10.1145/3618260.3649726

Given a Boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, define the functions $f^n : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}^n$ and $f^{\oplus n} : \mathcal{X}^n \times \mathcal{Y}^n \to \{0,1\}$ as follows[1]:

$$f^n(xy) = (f(x_1y_1), f(x_2y_2), \ldots, f(x_ny_n)),$$
$$f^{\oplus n}(xy) = f(x_1y_1) \oplus f(x_2y_2) \oplus \cdots \oplus f(x_ny_n).$$

So, f^n computes f on n different pairs of inputs, and $f^{\oplus n}$ computes the parity of the outputs of f^n. If f is hard to compute, are f^n and $f^{\oplus n}$ even harder to compute? For deterministic communication complexity, Feder, Kushilevitz, Naor and Nisan [10] proved that if $|\mathcal{X}|, |\mathcal{Y}| \le 2^\ell$ and f requires C bits of communication, then f^n requires at least $n(\sqrt{C} - \log_2 \ell - 1)$ bits of communication. In this work, we study randomized communication complexity. Let $\|\pi\|$ denote the communication complexity of a randomized communication protocol π and define the advantage:

$$\mathrm{adv}(C, f) = \sup_{\|\pi\| \le C} \inf_{xy} \mathbb{E}[(-1)^{\pi(xy)+f(xy)}].$$

This quantity measures the best worst-case advantage achievable by a C-bit protocol over random guessing. We can now state our main result:

Theorem 1. *There is a universal constant $\kappa > 0$ such that if $C > 1/\kappa$ and $\mathrm{adv}(C, f) < 1/2$, then*

$$\mathrm{adv}\left(\frac{\kappa C\sqrt{n}}{\log(Cn)}, f^{\oplus n}\right) < \exp(-\kappa n).$$

The constant $1/2$ is not important, it can be replaced by any constant less than 1. Some assumption of the type $C > 1/\kappa$ is necessary, because if $x, y \in \{0,1\}$ and $f(xy) = x \oplus y$, then $\mathrm{adv}(1, f) = 0$, yet $\mathrm{adv}(2, f^{\oplus n}) = 1$. Prior to our work, the best known upper bound was proved by the second author with Barak, Braverman and Chen [2], who showed that the advantage is bounded by $1/2$ for a similar choice of the other parameters. Our work builds on the work of Yu [31], who proved exponentially small bounds on the advantage in the setting of bounded-round communication protocols.

Our ideas lead to many results similar to Theorem 1. Next, we review the history that led us to the notion of marginal information, explain the intuitions behind the choices made in the definition, and then describe all of our results in Section 1.2.

1.1 The Evolution of Information Complexity

Marginal information is the most recent advance in an evolution of definitions about information. We relate bounds on the communication and advantage for computing f to the corresponding parameters for $f^{\oplus n}$ via a scheme that has been applied many times before. We prove:

[1]Throughout, we drop the delimiters between variables. $f(xy)$ is to be read as $f(x, y)$.

Step 1 Every protocol computing $f^{\oplus n}$ with significant advantage and small communication has small marginal information; see Theorem 5.

Step 2 Marginal information is subadditive, so the marginal information for computing f is smaller by a factor of n; see Theorem 6.

Step 3 Small marginal information can be compressed to give protocols with small communication; see Theorems 7 to 10.

Definitions of information are famously subtle. In order to make this strategy work, the marginal information needs to permit all 3 steps, and even minor changes to the definition can make one of the steps infeasible.

Our current definition builds on important insights and intuitions developed in theoretical computer science over a period of decades. An early precursor to the use of information theory in computer science is the work of Kalyanasundaram and Schnitger, who used Kolmogorov complexity to prove lower bounds on the randomized communication complexity of the disjointness function [27]. The proof was subsequently simplified by Razborov [25], who gave a beautiful short argument that used Shannon's notion of entropy [28] and implicitly followed the outline of the steps 1,2,3 described above. This is related to the questions we study here because the disjointness function can be thought of as a way to compute the AND of 2 bits n times. Step 1 is relatively easy for this problem. Step 2 involved a clever way to split the dependence between random variables, and was accomplished using the subadditivity of entropy. Step 3 is also not too difficult.

The next chapter of the story was written during the study of parallel repetition, a vital tool in the development of probabilistically checkable proofs. Raz [24] proved the first exponentially small bounds in this context using the Kullback-Liebler divergence as a measure of information. Given a distribution $p(xy)$, and a carefully chosen event W, Raz measured the divergence

$$\begin{aligned}&\mathop{\mathbb{E}}_{p(xy|W)}\left[\mathsf{D}(p(x|yW)||p(x|y)) + \mathsf{D}(p(y|xW)||p(y|x))\right]\\ &= \mathop{\mathbb{E}}_{p(xy|W)}\left[\log\left(\frac{p(x|yW)}{p(x|y)}\cdot\frac{p(y|xW)}{p(y|x)}\right)\right].\end{aligned} \tag{1}$$

In the proof, it is crucial that the event W is *rectangular*, meaning that if x, y are independent, then they remain independent even after conditioning on W. Once again, Step 1 is not too difficult. Raz used the subadditivity of divergence and a similar set of clever random variables as in [25] to split the dependence and accomplish Step 2. Later, Holenstein [13] introduced a method called *correlated sampling* to simplify the analogue of Step 3 in Raz's proof, and obtained better bounds. The second author used these tools to prove optimal bounds for parallel repetition in the setting relevant to probabilistically checkable proofs [21].

Chakrabarti, Shi, Wirth and Yao [9] were the first to propose using general measures of information complexity to address the questions we consider in this paper. Let xy denote the inputs, m denote the public randomness and transcript of a communication protocol and $p(xym)$ denote the joint distribution induced by the protocol[2]. [9] proposed to measure the mutual information

$$\mathsf{I}(M:XY) = \mathop{\mathbb{E}}_{p(xym)}\left[\log\frac{p(xy|m)}{p(xy)}\right].$$

Years later, this measure was renamed *external information* by [2]. The external information measures the information learned by an external observer about the parties' inputs. Step 1 is easy for this measure of information. However, the subadditivity of Step 2 does not hold in general; the proof only goes through when the input distribution $p(xy)$ is a product distribution. Jain, Radhakrishnan and Sen [15], and Harsha, Jain, McAllester and Radhakrishnan [12] gave ways to implement Step 3 that led to bounds on the success probability for computing f^n in the setting where the inputs are assumed to come from a product distribution and the communication protocols are restricted to having a bounded number of rounds. Meanwhile, Bar-yossef, Jayram, Kumar and Sivakumar [1] showed how to reframe Razborov's proof using mutual information instead of entropy, and proved other results using this formulation which contained hints of the definition of information that came next.

The first upper bounds on the success probability in the general setting came when the second author together with Barak, Braverman and Chen [2] adapted the methods developed in the study of parallel repetition to these problems. In contrast with the external information, they defined the *internal information*, which is the sum of two mutual information terms

$$\mathsf{I}(M:X|Y) + \mathsf{I}(M:Y|X) = \mathop{\mathbb{E}}_{p(xym)}\left[\log\left(\frac{p(x|ym)}{p(x|y)}\cdot\frac{p(y|xm)}{p(y|x)}\right)\right]. \tag{2}$$

The internal information measures what is learned by each party about the other's input. Equation (1) was the inspiration for Equation (2); indeed, each setting of m corresponds to a rectangular event. When the inputs come from a product distribution, the internal and external information are the same, and [2] proved that subadditivity holds for internal information using an argument similar to the one used in the context of parallel repetition. Moreover, they showed how to leverage the technique of correlated sampling developed by Holenstein to simulate protocols with information I and communication C using $\approx \sqrt{IC}/\log C$ communication. They gave near optimal simulations of $\approx I\log^2 C$ for protocols with small external information using rejection sampling and a variant of Azuma's concentration inequality. These results proved that there is a constant κ such that if $\mathsf{adv}(C, f) < 1/2$, then

$$\mathsf{adv}\left(\frac{\kappa C\sqrt{n}}{\log(Cn)}, f^{\oplus n}\right) < 1/2,$$

which was the first result along the lines of Theorem 1. Later, the second author and Braverman [6] argued that this is the *right* definition of information, because the internal information cost of a function is equal to the amortized communication complexity of that function. This suggested that the internal information might well be the last word in this evolution of definitions, because it could be defined purely using the concept of communication complexity. It seemed like the only path to better results was through better

[2]We often say $p(xym)$ is a protocol when we mean that it is a distribution induced by a protocol.

methods to compress internal information. This is a belief we no longer hold.

Nevertheless, a flurry of ideas about compressing protocols with internal information I and communication C followed. Braverman [4] showed how to obtain protocols with communication $\approx 2^{O(I)}$. The second author and Ramamoorthy [20] showed that if I_A, I_B denote the internal information learned by each party, then you can achieve communication $\approx I_A \cdot 2^{O(I_B)}$ and can also achieve communication $\approx I_A + \sqrt[4]{I_B \cdot C^3}$. Two excellent papers, the first by Kol [18] and the second by Sherstov [29], showed that $\approx I \log^2 I$ communication can be achieved when the inputs come from a product distribution. Ganor, Kol and Raz [11] (see also [22]) gave a nice counterexample: a function that can be computed with communication $\approx 2^{2^{O(I)}}$, and internal information $\approx I$, but cannot be computed with communication $\approx 2^I$.

The next definition to evolve was proposed by the second author together with Braverman, Weinstein and Yehudayoff [7, 8], inspired by the work of Jain, Pereszlényi and Yao [14]. Rather than bounding the information under the distribution $p(xym)$ induced by the protocol, they bounded the infimum of information achieved in the ball of distributions that are close to the protocol. They defined the information to be the infimum

$$\begin{aligned}&\inf_q \mathsf{I}_q(M : X|Y) + \mathsf{I}_q(M : Y|X)\\ &= \inf_q \mathop{\mathbb{E}}_{q(xym)} \left[\log\left(\frac{q(x|ym)}{q(x|y)} \cdot \frac{q(y|xm)}{q(y|x)} \right) \right], \end{aligned} \tag{3}$$

where here the infimum is taken over all distributions $q(xym)$ that are close to $p(xym)$ in statistical distance. This quantity was ultimately bounded by setting $q(xym) = p(xym|W)$, where here W is a reasonably large event (not necessarily rectangular) that implies that the protocol correctly computes the function. The bound on Equation (3) does not lead to a bound on the information according to $p(xym)$, because it is quite possible that the points outside W reveal a huge amount of information. Still, [8] were able to follow all 3 steps of the high-level approach to prove their results. Step 1 remained easy, but Steps 2 and 3 became more difficult using Equation (3). [8] obtained exponentially small upper bounds for the success probability of computing f^n, but did not manage to prove new bounds on the advantage for $f^{\oplus n}$ using this approach. Equation (3) may not seem very different from Equation (2), but it does involve a proxy q, and we pursue the use of such proxies further in the definition of marginal information that we discuss next.

In a paper full of new ideas, Yu [31] recently proved exponentially small bounds on the advantage of bounded-round protocols computing $f^{\oplus n}$. Although Yu's paper involves a potential function that superficially looks like a definition of information, his proof does not involve a method to compress protocols whose potential is small, and we are unable to extract a definition of information from his work. Still, his ideas inspired many of the choices made in our definition. To define the marginal information, we need the concept of a rectangular distribution, which was defined in [31]:

Definition 2. *Given a set Q consisting of triples (xym), we say that Q is* rectangular *if its indicator function can be expressed as*

$$\mathbb{1}_Q(xym) = \mathbb{1}_A(xm) \cdot \mathbb{1}_B(ym),$$

for some Boolean functions $\mathbb{1}_A, \mathbb{1}_B$. Given a distribution $q(xym)$ and a distribution $\mu(xy)$, we say that q is rectangular with respect to μ *if it can be expressed as*

$$q(xym) = \mu(xy) \cdot A(xm) \cdot B(ym),$$

for some functions A, B.

For intuition, it is helpful to think of a rectangular distribution as the result of conditioning a protocol distribution $p(xym)$ on a rectangular event. That would produce a rectangular distribution, but the space of rectangular distributions actually contains other distributions that cannot be obtained in this way.

From our perspective, the most useful insight of Yu's work is that if q is restricted to being rectangular, then one can allow q to be quite far from p in Equation (3) and still carry out a meaningful compression of a protocol p to implement Step 3. That is because the rectangular nature of q allows the parties to use hashing and rejection sampling to convert a protocol that samples from p into a protocol that samples from q. If $q(xym) = p(xym|R)$ for a rectangular event R, this is easy to understand: the parties can communicate 2 bits to compute if $xym \in R$ and output the most likely value of f under q with $xym \in R$. If $xym \notin R$ they can output a random guess for the value of f. So, it is enough to bound the information terms for $xym \in R$, and enough to guarantee that the compression is efficient for such points. This observation is very powerful, because it allows us to throw away problematic points in the support of the distributions we are working with and pass to appropriate sub-rectangles throughout our proofs.

For all of this to work, it is crucial that the protocol retains some advantage within the support of q. For this reason, we need to keep track of the information in the support of q as well as the advantage within the support of q, and so, for the first time, the measure of information is going to depend on the function f that the protocol computes. We are ready to state the definition:

Definition 3. *For $I \geq 1$ and[3] $\delta = 1/15$, the* marginal information *of a protocol p for computing f is defined as*

$$\mathsf{M}_I(p, f) = \inf_q \sup_{xym} \log \left(\frac{q(x|ym)}{p(x|y)} \cdot \frac{q(y|xm)}{p(y|x)} \cdot \left(\frac{q(xym)}{p(xym)} \right)^I \cdot \left| \mathop{\mathbb{E}}_{q(xy|m)} [(-1)^{f(xy)}] \right|^{-12I/\delta} \right),$$

where the infimum is taken over all distributions q that are rectangular with respect to the input distribution $p(xy)$, and the supremum is taken over all xym in the support of q.

We use the letter I above because it turns out that protocols computing f can be efficiently compressed when $\mathsf{M}_I = O(I)$, and any compression must have communication $\Omega(I)$. Compare Definition 3 with Equations (2) and (3). The fact that q must be tethered to p is ensured by including the term $q(xym)/p(xym)$. If $q(xym) = p(xym|R)$ for a rectangular event R, $q(xym)/p(xym)$ will be equal to $1/p(R)$. The last term in the product computes the advantage of q for computing f, because under q and given m, the best guess for the value of f is determined by the sign of

[3] Even though δ is a fixed constant, we choose to write it in the definition because it eases the notation throughout the paper.

$\mathbb{E}_{q(xy|m)}[(-1)^{f(xy)}]$, and its advantage is the absolute value of this quantity. In words, the marginal information measures the supremum over all xym of the information per unit of advantage, of the best rectangular approximation q.

In analogy with the external information, we define the external marginal information:

Definition 4. *For $I \geq 1$ and $\delta = 1/15$, the* external marginal information *of a protocol p for computing f is defined as:*

$$\mathsf{M}_I^{\text{ext}}(p,f) = \inf_q \sup_{xym} \log\left(\frac{q(xy|m)}{p(xy)} \cdot \left(\frac{q(xym)}{p(xym)}\right)^I \cdot \left|\mathop{\mathbb{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|^{-12I/\delta}\right),$$

where the infimum is taken over all distributions q that are rectangular with respect to the input distribution $p(xy)$, and the supremum is taken over all xym in the support of q.

When the distribution on inputs is a product distribution, it turns out that the external marginal information is equal to the marginal information.

To state our results about marginal information, we first define the average-case measure of advantage. Given a distribution $\mu(xy)$ on inputs, define

$$\mathrm{adv}_\mu(C,f) = \sup_{\|\pi\| \leq C} \mathbb{E}[(-1)^{\pi(xy)+f(xy)}],$$

where here the expectation is over the choice of inputs xy as well as the random coins of the communication protocol. To study the more restricted setting where the protocols we are working with have a bounded number of rounds, define the worst-case and average case quantities:

$$\mathrm{adv}^r(C,f) = \sup_{\|\pi\| \leq C} \inf_{xy} \mathbb{E}[(-1)^{\pi(xy)+f(xy)}],$$

$$\mathrm{adv}^r_\mu(C,f) = \sup_{\|\pi\| \leq C} \mathbb{E}[(-1)^{\pi(xy)+f(xy)}],$$

where throughout, the supremums are taken over r-round protocols.

Returning to our high-level approach, we prove the following results about marginal information, which allow us to carry out Steps 1,2,3:

(1) First, we show that a protocol with small communication and large advantage has small marginal information, to handle Step 1:

Theorem 5. *For every Boolean function $f(xy)$ and every protocol p of communication complexity C,*

$$\mathsf{M}_I(p,f) \leq 2C + O(I) - (1+12/\delta)\cdot I \cdot \log\left(\mathop{\mathbb{E}}_{p(m)}\left|\mathop{\mathbb{E}}_{p(xy|m)}\left[(-1)^f\right]\right|\right).$$

For any fixed m, the quantity $|\mathbb{E}_{p(xy|m)}[(-1)^f]|$ measures the advantage of the protocol for computing f conditioned on that value of m. So, if $\mathrm{adv}_\mu(C, f^{\oplus n}) \geq \exp(-m)$ via a protocol corresponding to the distribution p, then the above theorem implies that $\mathsf{M}_I(p, f^{\oplus n}) \leq O(C+Im)$. Unlike all previous definitions, for marginal information Step 1 involves significant work. Our proof crucially uses the fact that the protocol has bounded communication complexity: for example it would not be enough to start with a bound on the internal information.

(2) Next, we prove that marginal information is sub-additive with respect to the n-fold xor of f. If the transcript $m = (m_0, m_1, \ldots, m_C)$, where m_j denotes the j'th message of the protocol, we show

Theorem 6. *There is a universal constant Δ such that if $I \geq 1$ and p is a protocol distribution for computing $f^{\oplus n}$ with $p(xy) = \prod_{i=1}^n p(x_i y_i)$, then there is a protocol p_i for computing f such that $p_i(x_i y_i) = p(x_i y_i)$, p_i has the same number of messages as p, for $j > 1$ the support of m_j is identical in p_i and p, and moreover*

$$\mathsf{M}_I(p_i, f) \leq \frac{\mathsf{M}_I(p, f^{\oplus n})}{n} + \Delta I \cdot \left(1 + \log \frac{\mathsf{M}_I(p, f^{\oplus n})}{n \cdot I}\right).$$

If $\mathsf{M}_I(p, f^{\oplus n}) \leq O(In)$, this theorem proves that $\mathsf{M}_I(p_i, f) \leq O(I)$. This might well be the most technically novel part of our proof; it is certainly where we spent the most time. The main challenge is proving the result for $n = 2$, which is very delicate. If $n = 2$ and $\mathsf{M}_I(p, f^{\oplus 2})$ is small, then there is a rectangular distribution q such that the pair

$$q(x_1x_2y_1y_2m), p(x_1x_2y_1y_2m)$$

leads to a small value of $\mathsf{M}_I(p, f^{\oplus 2})$. We show how to use q, p to generate a new pair

$$q_1(x_1y_1m^{(1)}), p_1(x_1y_1m^{(1)})$$

or a new pair

$$q_2(x_2y_2m^{(2)}), p_2(x_2y_2m^{(2)})$$

proving that either $\mathsf{M}_I(p_1, f)$ or $\mathsf{M}_I(p_2, f)$ is more or less bounded by $\mathsf{M}_I(p, f^{\oplus 2})/2$.
We are unable to bound the length of the first message of p_i in terms of the length of the corresponding message of p in Theorem 6, because in our proof the first message $m_1^{(1)}$ or $m_1^{(2)}$ needs to encode one of the inputs of the original protocol. Fortunately, this is not a significant obstacle for the high-level strategy.

(3) Lastly, we show how to compress marginal information to handle Step 3. We have been able to match many of the prior results [2, 4, 6] about compressing information and external information with corresponding results about compressing marginal information and external marginal information, though our proofs are much more technical. Our most general simulation is captured by the following theorem:

Theorem 7. *For every $\alpha > 0$ there is a $\Delta > 0$ such that if $\mathsf{M}_I(p,f) \leq \alpha I$, $\mu(xy) = p(xy)$ and moreover the messages $m = (m_0, \ldots, m_C)$ are such that $m_2, \ldots, m_C \in \{0,1\}$, then $\mathrm{adv}_\mu(\Delta(I + \sqrt{CI}\log(CI)), f) \geq 1/\Delta$.*

Theorem 7 shows that if the marginal information is $O(I)$, then one can obtain a protocol with communication $\widetilde{O}(\sqrt{CI})$ that has $\Omega(1)$ advantage for computing f. For the external marginal information, we prove:

Theorem 8. *For every $\alpha > 0$ there is a $\Delta > 0$ such that if $\mathsf{M}_I^{\mathrm{ext}}(p, f) \le \alpha I$, $\mu(xy) = p(xy)$, and moreover the messages $m = (m_0, \ldots, m_C)$ are such that $m_2, \ldots, m_C \in \{0,1\}$, then $\mathrm{adv}_\mu(\Delta I \log^2 C, f) \ge 1/\Delta$.*

This theorem gives improved results when the inputs come from a product distribution. It is quite possible that even better simulations can be obtained using the ideas of [5, 18, 29], but we have not managed to obtain such results. We also obtain results that are independent of the communication complexity:

Theorem 9. *For every $\alpha > 0$ there is a $\Delta > 0$ such that if $\mathsf{M}_I(p, f) \le \alpha I$ and $\mu(xy) = p(xy)$, then $\mathrm{adv}_\mu(\Delta I, f) \ge \exp(-\Delta I)$.*

When the number of rounds of the protocol is bounded, we prove:

Theorem 10. *For every $\alpha > 0$ there is a $\Delta > 0$ such that if $\mathsf{M}_I(p, f) \le \alpha I$, $\mu(xy) = p(xy)$, p has r-rounds and $m_r \in \{0,1\}$, then $\mathrm{adv}^r_\mu(\Delta r(I + \log r), f) \ge 1/\Delta$.*

These results about the marginal information cost allow us to prove Theorem 1, as well as several other results of that flavor.

1.2 Using Marginal Information to Prove XOR Lemmas

To state all of our results, let us define the average-case and worst-case measures of success:

$$\begin{aligned}
\mathrm{suc}(C, f) &= \sup_{\|\pi\| \le C} \inf_{xy} \mathbf{Pr}[\pi(xy) = f(xy)] \\
\mathrm{suc}^r(C, f) &= \sup_{\|\pi\| \le C} \inf_{xy} \mathbf{Pr}[\pi(xy) = f(xy)] \\
\mathrm{suc}_\mu(C, f) &= \sup_{\|\pi\| \le C} \mathbf{Pr}[\pi(xy) = f(xy)] \\
\mathrm{suc}^r_\mu(C, f) &= \sup_{\|\pi\| \le C} \mathbf{Pr}[\pi(xy) = f(xy)],
\end{aligned}$$

where in $\mathrm{suc}^r, \mathrm{suc}^r_\mu$ the supremum is taken over r-round protocols, and in $\mathrm{suc}_\mu, \mathrm{suc}^r_\mu$ the probability is over inputs sampled from $\mu(xy)$. Yao's min-max theorem yields

$$\begin{aligned}
\mathrm{adv}(C, f) &= \inf_\mu \mathrm{adv}_\mu(C, f), \\
\mathrm{suc}(C, f) &= \inf_\mu \mathrm{suc}_\mu(C, f), \\
\mathrm{adv}^r(C, f) &= \inf_\mu \mathrm{adv}^r_\mu(C, f), \\
\mathrm{suc}^r(C, f) &= \inf_\mu \mathrm{suc}^r_\mu(C, f). \qquad (4)
\end{aligned}$$

Given any distribution μ on $\mathcal{X} \times \mathcal{Y}$, define the n-fold product distribution μ^n on $\mathcal{X}^n \times \mathcal{Y}^n$ by $\mu^n(xy) = \prod_{j=1}^n \mu(x_j y_j)$. Theorem 1 is proved by proving this stronger bound:

Theorem 11. *There is a universal constant $\kappa > 0$ such that if $C > 1/\kappa$ and $\mathrm{adv}_\mu(C, f) \le \kappa$, then $\mathrm{adv}_{\mu^n}(\kappa C\sqrt{n}/\log(Cn), f^{\oplus n}) \le \exp(-\kappa n)$.*

To prove Theorem 11, suppose that there is a protocol p computing $f^{\oplus n}$ with advantage $\exp(-\kappa n)$ and communication $T = \kappa C \cdot \sqrt{n}/\log(Cn)$. If $T/n \ge 1$, we set $I = T/n$ and apply Theorem 5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \le O(T + \kappa I n) \le O(In)$. Next, apply Theorem 6 to find a protocol p' with $\mathsf{M}_I(p', f) \le O(I)$. Finally, apply Theorem 7 to obtain a protocol computing f with advantage $\Omega(1)$ and communication proportional to

$$\begin{aligned}
\frac{T}{n} + 2\sqrt{IT}\log(T) &\le \frac{T}{n} + 2\frac{T\log T}{\sqrt{n}} \\
&\lesssim \frac{\kappa C}{\log nC} \cdot \log T \lesssim \kappa C.
\end{aligned}$$

If $T/n < 1$, we set $I = 1$ and apply Theorem 5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \le O(In)$. Next, apply Theorem 6 to find a protocol p' with $\mathsf{M}_I(p', f) \le O(I) = O(1)$. Finally, we apply Theorem 9 to obtain a protocol computing f with advantage $\Omega(1)$ and communication $O(1)$. Setting κ sufficiently small, we obtain a contradiction in either case, which proves that there is no protocol p as above. Theorem 1 can be obtained from Theorem 11 using Equation (4) and the fact that the worst-case success probability of a communication protocol can be increased by taking the majority outcome of several runs of the protocol. We leave these details to the reader.

Theorems 1 and 11 yield bounds on the success probability for computing f^n as well:

Corollary 12. *There is a universal constant $\kappa > 0$ such that if $C > 1/\kappa$ and $\mathrm{adv}(C, f) < \kappa$, then $\mathrm{suc}(\kappa C\sqrt{n}/\log(Cn)), f^n) < \exp(-\kappa n)$.*

Corollary 13. *There is a universal constant $\kappa > 0$ such that if $C > 1/\kappa$ and $\mathrm{adv}_\mu(C, f) < \kappa$, then $\mathrm{suc}_{\mu^n}(\kappa C\sqrt{n}/\log(Cn)), f^n) < \exp(-\kappa n)$.*

This matches the result proved by [8] mentioned earlier. These corollaries are obtained by observing that if $S \subseteq \{1, 2, \ldots, n\}$ is chosen uniformly at random, and xy are sampled according to μ^n, then

$$\mathbb{E}\left[(-1)^{\sum_{j \in S} \pi(xy)_j + f(x_j y_j)}\right] = \mathbf{Pr}[\pi(xy) = f^n(xy)],$$

so a protocol computing f^n with success probability $\exp(-n/2)$ yields a set of $n' = \Omega(n)$ coordinates where the protocol computes $f^{\oplus n'}$ with advantage $\exp(-\Omega(n))$. Again, we leave the details to the reader. When the distribution $\mu(xy) = \mu(x) \cdot \mu(y)$ is a product distribution, we obtain stronger bounds:

Theorem 14. *There is a universal constant $\kappa > 0$ such that for every product distribution μ, if $C > 1/\kappa$ and $\mathrm{adv}_\mu(C, f) < \kappa$, then $\mathrm{adv}_{\mu^n}(\kappa Cn/\log^2(Cn), f^{\oplus n}) < \exp(-\kappa n)$.*

To prove Theorem 14, suppose we are given a protocol p computing $f^{\oplus n}$ with advantage $\exp(-\kappa n)$ and communication $T = \kappa Cn/\log^2(Cn)$. If $T/n \ge 1$, we set $I = T/n$ and apply Theorem 5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \le O(nI)$. Next, apply Theorem 6 to find a protocol p' with $\mathsf{M}_I(p', f) \le O(I)$. Finally, using the fact that for product distributions, $\mathsf{M}_I^{\mathrm{ext}}(p, f) = \mathsf{M}_I(p, f)$, we can apply Theorem 8 to obtain a protocol computing f with advantage $\Omega(1)$ and communication $O(I\log^2(Cn)) \le O(\kappa C)$. Otherwise, if $T/n < 1$, set $I = 1$ and apply Theorem 5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \le O(n)$. Then, apply Theorem 6 to find a protocol p' with $\mathsf{M}_I(p', f) \le O(I) = O(1)$. Lastly, we apply Theorem 9 to obtain a protocol computing f with advantage $\Omega(1)$ and communication $O(1)$. Setting κ to be small enough gives a contradiction in either case.

As before, this yields a corollary for computing f^n:

Corollary 15. *There is a universal constant $\kappa > 0$ such that for every product distribution μ, if $C > 1/\kappa$ and $\mathrm{adv}_\mu(C, f) < \kappa$, then $\mathrm{suc}_{\mu^n}(\kappa Cn/\log^2(Cn), f^n) < \exp(-\kappa n)$.*

Again, this is identical to a bound proved by [8] using a different approach. For the bounded-round setting, we prove:

Theorem 16. *There is a universal constant $\kappa > 0$ such that if $C > (r(\log r) + 1)/\kappa$, and $\mathrm{adv}^r_\mu(C, f) < \kappa$, then $\mathrm{adv}^r_{\mu^n}((\kappa C/r - \log r)n, f^{\oplus n}) < \exp(-\kappa n)$.*

Yu [31] proves the same bound on the advantage with a communication budget that grows like $\Omega((C/r^r - O(1))n)$. Our bound eliminates the exponential dependence on r. To prove Theorem 14, set $T = (\kappa C/r - \log r)n$, and suppose there is a protocol computing f with r rounds, communication T and advantage $\exp(-\kappa n)$. Set $I = T/n \geq 1$. Then, M_I can be bounded by $O(T + \kappa I n)$ by Theorem 5. Applying Theorem 6 gives an r-round protocol with M_I bounded by $O(I)$, and applying Theorem 10 gives an r-round protocol with communication complexity $O(r(I + \log r)) = O(\kappa C)$ computing f with advantage $\Omega(1)$. Setting κ to be small enough proves the result. As usual, we obtain the following corollaries:

Corollary 17. *There is a universal constant $\kappa > 0$ such that if $C > 7(r \log r)/\kappa$ and $\mathrm{adv}^r_\mu(C, f) < \kappa$, then $\mathrm{suc}^r_{\mu^n}((\kappa C/r - \log r)n, f^n) < \exp(-\kappa n)$.*

Corollary 18. *There is a universal constant $\kappa > 0$ such that if $C > 7(r \log r)/\kappa$, and $\mathrm{adv}^r(C, f) < \kappa$, then $\mathrm{suc}^r((\kappa C/r - \log r)n, f^n) < \exp(-\kappa n)$.*

ACKNOWLEDGMENTS

Thanks to Paul Beame, Makrand Sinha, Oscar Sprumont, Michael Whitmeyer and Amir Yehudayoff for helpful conversations.

REFERENCES

[1] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. 2002. An Information Statistics Approach to Data Stream and Communication Complexity. In *43rd Symposium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*. IEEE Computer Society, 209–218. https://doi.org/10.1109/SFCS.2002.1181944

[2] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. 2010. How to Compress Interactive Communication. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing* (Cambridge, Massachusetts, USA) *(STOC '10)*. Association for Computing Machinery, New York, NY, USA, 67–76. https://doi.org/10.1145/1806689.1806701

[3] Gábor Braun and Sebastian Pokutta. 2016. Common Information and Unique Disjointness. *Algorithmica* 76, 3 (2016), 597–629. https://doi.org/10.1007/S00453-016-0132-0

[4] Mark Braverman. 2015. Interactive Information Complexity. *SIAM J. Comput.* 44, 6 (2015), 1698–1739. https://doi.org/10.1137/130938517 arXiv:https://doi.org/10.1137/130938517

[5] Mark Braverman and Gillat Kol. 2018. Interactive Compression to External Information. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* (Los Angeles, CA, USA) *(STOC 2018)*. Association for Computing Machinery, New York, NY, USA, 964–977. https://doi.org/10.1145/3188745.3188956

[6] Mark Braverman and Anup Rao. 2011. Information Equals Amortized Communication. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, Los Alamitos, CA, USA, 748–757. https://doi.org/10.1109/FOCS.2011.86

[7] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. 2013. Direct Product via Round-Preserving Compression. In *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 7965)*, Fedor V. Fomin, Rusins Freivalds, Marta Z. Kwiatkowska, and David Peleg (Eds.). Springer, 232–243. https://doi.org/10.1007/978-3-642-39206-1_20

[8] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. 2013. Direct Products in Communication Complexity. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*. IEEE Computer Society, 746–755. https://doi.org/10.1109/FOCS.2013.85

[9] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. 2001. Informational Complexity and the Direct Sum Problem for Simultaneous Message Complexity. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*. IEEE Computer Society, 270–278. https://doi.org/10.1109/SFCS.2001.959901

[10] Tomás Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. 1995. Amortized Communication Complexity. *SIAM J. Comput.* 24, 4 (1995), 736–750. https://doi.org/10.1137/S0097539792235864

[11] Anat Ganor, Gillat Kol, and Ran Raz. 2016. Exponential Separation of Information and Communication for Boolean Functions. *J. ACM* 63, 5, Article 46 (nov 2016), 31 pages. https://doi.org/10.1145/2907939

[12] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. 2010. The Communication Complexity of Correlation. *IEEE Transactions on Information Theory* 56, 1 (2010), 438–449. https://doi.org/10.1109/TIT.2009.2034824

[13] Thomas Holenstein. 2007. Parallel Repetition: Simplifications and the No-Signaling Case. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing* (San Diego, California, USA) *(STOC '07)*. Association for Computing Machinery, New York, NY, USA, 411–419. https://doi.org/10.1145/1250790.1250852

[14] Rahul Jain, Attila Pereszlényi, and Penghui Yao. 2012. A Direct Product Theorem for the Two-Party Bounded-Round Public-Coin Communication Complexity. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*. IEEE Computer Society, 167–176. https://doi.org/10.1109/FOCS.2012.42

[15] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. 2003. A Direct Sum Theorem in Communication Complexity via Message Compression. In *Automata, Languages and Programming*, Jos C. M. Baeten, Jan Karel Lenstra, Joachim Parrow, and Gerhard J. Woeginger (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 300–315.

[16] Xinrui Jia, Ola Svensson, and Weiqiang Yuan. 2023. The Exact Bipartite Matching Polytope Has Exponential Extension Complexity. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 1635–1654. https://doi.org/10.1137/1.9781611977554.ch61

[17] Mauricio Karchmer, Ran Raz, and Avi Wigderson. 1995. Super-Logarithmic Depth Lower Bounds Via the Direct Sum in Communication Complexity. *Comput. Complex.* 5, 3/4 (1995), 191–204. https://doi.org/10.1007/BF01206317

[18] Gillat Kol. 2016. Interactive Compression for Product Distributions. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* (Cambridge, MA, USA) *(STOC '16)*. Association for Computing Machinery, New York, NY, USA, 987–998. https://doi.org/10.1145/2897518.2897537

[19] Mihai Pătraşcu. 2011. Unifying the Landscape of Cell-Probe Lower Bounds. *SIAM J. Comput.* 40, 3 (jun 2011), 827–847. https://doi.org/10.1137/09075336X

[20] Sivaramakrishnan Natarajan Ramamoorthy and Anup Rao. 2015. How to Compress Asymmetric Communication. In *Proceedings of the 30th Conference on Computational Complexity* (Portland, Oregon) *(CCC '15)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, DEU, 102–123.

[21] Anup Rao. 2011. Parallel Repetition in Projection Games and a Concentration Bound. *SIAM J. Comput.* 40, 6 (2011), 1871–1891. https://doi.org/10.1137/080734042 arXiv:https://doi.org/10.1137/080734042

[22] Anup Rao and Makrand Sinha. 2018. Simplified Separation of Information and Communication. *Theory of Computing* 14, 20 (2018), 1–29. https://doi.org/10.4086/toc.2018.v014a020

[23] Anup Rao and Amir Yehudayoff. 2020. *Communication Complexity: and Applications*. Cambridge University Press. https://doi.org/10.1017/9781108671644

[24] Ran Raz. 1995. A Parallel Repetition Theorem. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing* (Las Vegas, Nevada, USA) *(STOC '95)*. Association for Computing Machinery, New York, NY, USA, 447–456. https://doi.org/10.1145/225058.225181

[25] Alexander A. Razborov. 1992. On the Distributional Complexity of Disjointness. *Theor. Comput. Sci.* 106, 2 (1992), 385–390. https://doi.org/10.1016/0304-3975(92)90260-M

[26] Thomas Rothvoss. 2017. The Matching Polytope Has Exponential Extension Complexity. *J. ACM* 64, 6, Article 41 (sep 2017), 19 pages. https://doi.org/10.1145/3127497

[27] Georg Schnitger and Bala Kalyanasundaram. 1987. The probabilistic communication complexity of set intersection. In *Proceedings of the Second Annual Conference on Structure in Complexity Theory, Cornell University, Ithaca, New York, USA, June 16-19, 1987*. IEEE Computer Society, 41–47. https://ieeexplore.ieee.org/document/10319253

[28] Claude Elwood Shannon. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27 (1948), 379–423. http://plan9.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

[29] Alexander A. Sherstov. 2018. Compressing Interactive Communication Under Product Distributions. *SIAM J. Comput.* 47, 2 (2018), 367–419. https://doi.org/10.1137/16M109380X arXiv:https://doi.org/10.1137/16M109380X

[30] Makrand Sinha. 2018. Lower Bounds for Approximating the Matching Polytope. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, Artur Czumaj (Ed.). SIAM, 1585–1604. https://doi.org/10.1137/1.9781611975031.104

[31] Huacheng Yu. 2022. Strong XOR Lemma for Communication with Bounded Rounds : (extended abstract). In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*. IEEE, 1186–1192. https://doi.org/10.1109/FOCS54457.2022.00114

Received 13-NOV-2023; accepted 2024-02-11