Lifting for Arbitrary Gadgets

Siddharth Iyer*

siyer@cs.washington.edu

March 29, 2025

Abstract

We prove a sensitivity-to-communication lifting theorem for arbitrary gadgets. Given functions $f : \{0,1\}^n \to \{0,1\}$ and $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, denote $f \circ g(x,y) := f(g(x_1,y_1), \ldots, g(x_n,y_n))$. We show that for any f with sensitivity s and any g,

$$D(f \circ g) \geq s \cdot \left(\frac{\Omega(D(g))}{\log \mathsf{rk}(g)} - \log \mathsf{rk}(g) \right),$$

where $D(\cdot)$ denotes the deterministic communication complexity and $\mathsf{rk}(g)$ is the rank of the matrix associated with g. As a corollary, we get that if $D(g)$ is a sufficiently large constant, $D(f \circ g) = \Omega(\min\{s,d\} \cdot \sqrt{D(g)})$, where s and d denote the sensitivity and degree of f. In particular, computing the OR of n copies of g requires $\Omega(n \cdot \sqrt{D(g)})$ bits.

1 Introduction

Given two functions f and g, how much harder is it to compute their composition than it is to compute each of the functions? In this work, we study this question in the model of deterministic communication complexity (see [RY20, KN96] for a detailed reference on this topic). For a function $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, let $D(g)$ denote the deterministic communication complexity of g. Given a function $g(x,y)$ and a function $f : \{0,1\}^n \to \{0,1\}$ we will be interested in the deterministic communication complexity of

$$f \circ g(x,y) := f(g(x_1,y_1), \ldots, g(x_n,y_n)).$$

Besides being a natural problem to study, understanding function compositions especially in the context of deterministic communication, has connections to circuit lower bounds [KRW95]. We begin by recalling some prior work on related questions.

Connections to Direct Sums and XOR Lemmas. Feder, Kushilevitz, Naor and Nisan [FKNN95] were the first to consider the direct sum problem for deterministic communication, which asks for lower bounds on the communication required to compute n copies of g, denoted $g^n(x,y) := g(x_1,y_1), \ldots, g(x_n,y_n)$. Before we state their result, we define the cover number, $C(g)$, to be the smallest number of rectangles needed to cover $\mathcal{X} \times \mathcal{Y}$ such that each rectangle is constant for g. Feder et al. showed the following.

Theorem 1 ([FKNN95]). $D(g^n) \geq \log C(g^n) \geq n \cdot (\sqrt{D(g)} - \log \log(|\mathcal{X}| \cdot |\mathcal{Y}|) - 1)$.

This was the best known bound for the deterministic communication complexity of g^n until our recent work with Rao [IR24a], where we gave an XOR lemma: a lower bound for the deterministic communication complexity of $\oplus \circ g := g(x_1, y_1) \oplus \ldots \oplus g(x_n, y_n)$. To state that result, we need the concept of rank in the context of communication complexity. For a function $g(x, y)$, let M_g denote the matrix with a row for each $x \in \mathcal{X}$, a column for each $y \in \mathcal{Y}$, and whose (x, y)-th entry is simply $g(x, y)$. We use the notation $\mathsf{rk}(g)$ to denote the rank of the matrix M_g.

One might be tempted to guess that $D(\oplus \circ g) = \Omega(n \cdot D(g))$, i.e. one cannot do much better than computing g in each coordinate followed by taking the parity of all the bits. However, this is false in general; when g is itself the parity function on two bits, it is possible to compute $\oplus \circ g$ by simply exchanging 2 bits: $x_1 \oplus \ldots \oplus x_n$ and $y_1 \oplus \ldots \oplus y_n$. In [IR24a], we showed that if g requires sufficiently large communication, then computing $\oplus \circ g$ requires much larger communication.

Theorem 2 ([IR24a]). *There exists $c_0 \geq 1$ such that for any function g with $D(g) \geq c_0$, $D(\oplus \circ g) \geq \log C(\oplus \circ g) \geq n \cdot \left(\frac{\Omega(D(g))}{\log \mathsf{rk}(g)} - \log \mathsf{rk}(g) \right)$.*

Rank and communication are believed to be closely related. For any function g, we know that $D(g) \geq \log \mathsf{rk}(g)$ and the log-rank conjecture, due to Lovász and Saks [LS88] asserts that $D(g) \leq (\log \mathsf{rk}(g))^{O(1)}$. Sudakov and Tomon [ST24] (see also [Lov14] who gave a slightly weaker bound) recently showed that $D(g) \leq O(\sqrt{\mathsf{rk}(g)})$. When $D(g) = \Omega(\log^2 \mathsf{rk}(g))$, Theorem 2 implies $D(\oplus \circ g) \geq \Omega(n \cdot \sqrt{D(g)})$. Yang [Yan24] observed that $\mathsf{rk}(\oplus \circ g) \geq (\mathsf{rk}(g) - 1)^n - 1$ since rank tensorizes. He then concluded that if $D(g) \ll \log^2 \mathsf{rk}(g)$,

$$D(\oplus \circ g) \geq \log \mathsf{rk}(\oplus \circ g) \geq \Omega(n \cdot \log \mathsf{rk}(g)) \geq \Omega(n \cdot \sqrt{D(g)}).$$

Therefore, for all g with sufficiently large communication complexity, $D(\oplus \circ g) = \Omega(n \cdot \sqrt{D(g)})$.

In this work, we generalize Theorem 2 to all functions f with large sensitivity, which is defined as follows. The sensitivity of f at $z \in \{0, 1\}^n$ is

$$\mathsf{s}_z(f) := |\{i : f(z) \neq f(z_1, \ldots, z_{i-1}, 1 - z_i, z_{i+1}, \ldots, z_n)\}|,$$

and the sensitivity of f, is given by $\mathsf{s}(f) := \max_z \mathsf{s}_z(f)$. Our main result is the following.

Theorem 3 (Main Theorem). *There exists $c_0 \geq 1$ such that for any function g with $D(g) \geq c_0$ and any function $f : \{0, 1\}^n \to \{0, 1\}$,*

$$D(f \circ g) \geq \log C(f \circ g) \geq \mathsf{s}(f) \cdot \left(\frac{\Omega(D(g))}{\log \mathsf{rk}(g)} - \log \mathsf{rk}(g) \right).$$

Several basic functions, such as parity, the AND function, the OR function etc. all have sensitivity n, and the above result gives a statement analogous to Theorem 2 for such functions. Understanding the complexity of computing $f \circ g$ has also received significant attention in the area of query-to-communication lifting. Here, one typically fixes an inner function g, and asks for the communication complexity of computing $f \circ g$ for all f. We discuss the connection between our work and lifting theorems in more detail.

Connections to Lifting Theorems. We begin by noting that when f only depends on a small number of coordinates, $f \circ g$ can easily be computed by computing g in the appropriate coordinates. More generally, if one has a short decision tree (see Section 2 for a definition) for

f, then $D(f \circ g)$ can be computed by tracing out a root-to-leaf path in the decision tree and communicating to only compute g in the coordinates corresponding to the path. Let the *decision tree complexity* of f, denoted $\mathsf{DT}(f)$, be the least depth decision tree computing f. Using the best decision tree for f, we get $D(f \circ g) \leq \mathsf{DT}(f) \cdot D(g)$.

A natural question is whether or not the above bound is optimal. There is a large body of literature showing that the above bound is indeed optimal for several *gadgets* g. The earliest such result is due to Raz and McKenzie [RM99] who considered the index function gadget, $\mathsf{Ind}_m :$ $[m] \times \{0, 1\}^m \to \{0, 1\}$ given by $\mathsf{Ind}_m(x, y) = y_x$. They showed that for $m = n^{O(1)}$, and any function $f : \{0, 1\}^n \to \{0, 1\}$, $D(f \circ \mathsf{Ind}_m) = \Theta(\mathsf{DT}(f) \cdot D(\mathsf{Ind}_m))$. Their proof was simplified by Göös, Pitassi, and Watson [GPW18], who also used the lifting paradigm to exhibit functions with rank r and communication complexity $\widetilde{\Omega}(\log^2 r)$. The result of [RM99] was recently improved by [LMM+22], they showed the same result as that of [RM99] for $m = O(n \log n)$. Chattopadhyay, Koucký, Loff, and Mukhopadhyay [CKLM19] showed a similar result for the inner product gadget (as well as for any gadget with a certain pseudorandom property) $\mathsf{IP}_m : \{0, 1\}^m \times \{0, 1\}^m \to \{0, 1\}$ given by $\mathsf{IP}_m(x, y) = x_1 y_1 \oplus \ldots \oplus x_m y_m$. They showed that for $m = \Omega(\log n)$ and any function f, $D(f \circ \mathsf{IP}_m) = \Theta(\mathsf{DT}(f) \cdot D(\mathsf{IP}_m))$. Manor and Meir [MM22] proved that $D(f \circ g) = \Omega(\mathsf{DT}(f) \cdot D(g))$ for all functions g with discrepancy at most $n^{-O(1)}$. Lifting theorems have also been studied for various other complexity measures, such as decision tree complexity to randomized communication [GPW20, CFK+21, MM22], resolution width to DAG-like communication [GGKS20, LMM+22], approximate degree to approximate rank [She11, PR18] etc.

In light of the preceding discussion, Theorem 3 can be seen as a statement that lifts sensitivity to deterministic communication complexity for any gadget g whose communication complexity is a large enough constant. We also note that ours is not the first work to lift a Boolean function complexity measure other than decision tree complexity to communication. The work of Sherstov [She11] yields randomized communication lower bounds for compositions with the index gadget using approximate degree as the complexity measure for the outer function.

Related Boolean Function Complexity Measures. Sensitivity and decision tree complexity are two among a few well-studied Boolean function complexity measures, such as certificate complexity, block-sensitivity, and degree; we refer the reader to the survey of Buhrman and de Wolf [BdW02] for a detailed reference on this topic. We use the connections between one of these complexity measures, namely the degree, to derive a lower bound on $D(f \circ g)$ solely in terms of $D(g)$ and $\mathsf{DT}(f)$. The degree of f, denoted $\mathsf{deg}(f)$ is the degree of the unique real, multilinear polynomial computing f. The above complexity measures are known to be related to each other up to polynomial factors. In particular, we know that for any f,

$$\mathsf{deg}(f), \mathsf{s}(f) \leq \mathsf{DT}(f), \tag{1}$$

$$\sqrt{\mathsf{deg}(f)} \leq \mathsf{s}(f) \leq 2 \cdot \mathsf{deg}(f)^2, \text{ and} \tag{2}$$

$$\mathsf{DT}(f) \leq 2 \cdot \mathsf{deg}(f)^3. \tag{3}$$

In the preceding facts, Equation (1) is due to Nisan and Szegedy [NS92] (see also [Nis89]). The lower bound in Equation (2) is due to Huang [Hua19], and the upper bound is again due to Nisan and Szegedy [NS92]. Lastly, Equation (3) was shown by Midrijānis [Mid04]. Next, we discuss our results and use the connections between the above complexity measure to derive lower bounds on the communication complexity of computing $f \circ g$.

Theorem 3 together with Yang's observation yields the following lower bound on $D(f \circ g)$ in terms of the sensitivity and degree.

Corollary 4. *There exists $c_0 \geq 1$ such that for any function g with $D(g) \geq c_0$ and any function $f : \{0,1\}^n \to \{0,1\}$,*

$$D(f \circ g) \geq \frac{\mathsf{s}(f) \cdot \mathsf{deg}(f)}{2\mathsf{s}(f) + \mathsf{deg}(f)} \cdot \Omega\left(\frac{D(g)}{\log \mathsf{rk}(g)} + \log \mathsf{rk}(g)\right).$$

Using the fact that $2\mathsf{s}(f) + \mathsf{deg}(f) \leq 3 \max\{\mathsf{s}(f), \mathsf{deg}(f)\}$ and the AM-GM inequality, we get the following bound on $D(f \circ g)$ from Corollary 4

$$D(f \circ g) \geq \Omega\big(\min\{\mathsf{s}(f), \mathsf{deg}(f)\} \cdot \sqrt{D(g)}\big). \tag{4}$$

Combining Equation (4) with Equations (2) and (3) we get that for any g with sufficiently large communication complexity and any f,

$$D(f \circ g) \geq \Omega(\mathsf{DT}(f)^{1/6} \cdot \sqrt{D(g)}).$$

Other Related Work. Several works have studied the complexity of repeated computation. In communication complexity, this began with the influential proof of the randomized communication lower bound for the disjointness function [SK87, Raz92], which is the composition of the OR function on n bits with the AND gadget. Their techniques were modified to prove direct sum type statements [CSWY01, JRS03, HJMR10] and streaming lower bounds [BJKS02]. Around the same time Raz [Raz95] proved the parallel repetition theorem, which was simplified by Holenstein [Hol07] and Rao [Rao11]. Barak, Braverman, Chen and Rao [BBCR10] adapted the techniques developed for the parallel repetition theorem to prove general direct sum statements for randomized communication complexity. Several aspects of this were improved over the last decade, for example [BRWY13] gave a direct product theorem and [Kol16, She18] gave near-optimal distributional direct-sum statements for product distributions. Recently, Yu [Yu22] gave an XOR lemma for bounded-round randomized communication, which we improved in our work with Rao [IR24b] to obtain an XOR lemma for general randomized communication.

Techniques. The following lemma is key to the proof of Theorem 3.

Lemma 5. *If $f \circ g$ can be covered with 2^T monochromatic rectangles, then g contains a monochromatic rectangle of density $2^{-2T/\mathsf{s}(f)} \cdot (4 \cdot \mathsf{rk}(g))^{-2}$.*

A similar statement was also derived in [IR24a], where f was assumed to be the parity function on n bits. We explain the high-level idea of the proof in a few lines and how it differs from the proof in [IR24a]. Let $u(x,y)$ denote the uniform distribution on $\mathcal{X}^n \times \mathcal{Y}^n$. A key observation made in [IR24a] is that there exists a rectangle R that is constant for $\oplus \circ g$, a coordinate i, inputs $x_1, \ldots, x_{i-1}$ and $y_{i+1}, \ldots, y_n$ such that

1. $|\mathsf{supp}(u(x_i, y_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n, R))| \geq \Omega(|\mathcal{X}| \cdot |\mathcal{Y}| \cdot 2^{-T/n})$, where supp denotes the support of a distribution, and

2. there exists a bit $b \in \{0,1\}$ so that for every $(x', y') \in R$ consistent with $x_1, \ldots, x_{i-1}$ and $y_{i+1}, \ldots, y_n$ we have $\oplus_{j \neq i} g(x'_j, y'_j) = b$.

Since R is constant for $\oplus \circ g$, Items 1 and 2 together imply that $g(x_i, y_i)$ is fixed for every (x_i, y_i) in the support of $u(x_i, y_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n, R)$. Moreover, since x, y are independent in u, the support is a rectangle.

To prove lower bounds for general function composition, we need an appropriate generalization of the parity constraint in Item 2. We enforce such a constraint by switching from the uniform distribution to a correlated distribution.

For simplicity, assume $\mathsf{s}(f) = n$ and that g is balanced. By definition, there exists $z \in \{0,1\}^n$ with $f(z) \neq f(z_1, \ldots, z_{i-1}, 1-z_i, z_{i+1}, \ldots, z_n)$, for all i. Consider the distribution $p(x,y)$ on $\mathcal{X}^n \times \mathcal{Y}^n$ obtained by sampling each (x_i, y_i) uniformly conditioned on $g(x_i, y_i) = z_i$. Now, we obtain a weaker statement than Item 1, which suffices for our purposes. We show that there exists a rectangle R that is constant for $f \circ g$, a coordinate i and inputs $x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n$ such that the sets

$$A := \mathsf{supp}(p(x_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n, R)) \text{ and}$$
$$B := \mathsf{supp}(p(y_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n, R)),$$

satisfy $|A| \geq \Omega(|\mathcal{X}| \cdot 2^{-T/n})$ and $|B| \geq \Omega(|\mathcal{Y}| \cdot 2^{-T/n})$.

Similar to the proof of Item 1, this uses the sub-additivity of entropy, we refer the reader to Claim 13 for more details. Next, we show that $A \times B$ is a monochromatic rectangle for g. We note that

$$\mathsf{supp}(p(x_i, y_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n, R)) \subseteq A \times B,$$

and although g is constant on the former set (by the definition of p), it is not obvious that the same holds for the $A \times B$. Nevertheless, we show that this is indeed the case.

The intuition for this is as follows. For any $(x_i, y_i) \in A \times B$, one can use the definitions of A, B and $p(x,y)$, to show that there exist $(x', y') \in R$ satisfying $x'_i = x_i$, $y'_i = y_i$ and $g(x'_j, y'_j) = z_j$, for all $j \neq i$. Therefore, if $g(x_i, y_i) \neq z_i$, then $f \circ g(x', y') \neq f(z)$, by the sensitivity of f. This is a contradiction because for every $(x, y) \in R \supseteq \mathsf{supp}(p(x, y | R))$, $f \circ g$ evaluates to $f(z)$.

The main difference between the above high-level description and the proof of Theorem 3 is that g need not be balanced. To address this, we consider two cases. First, we assume that g is extremely biased, say $\mathbf{Pr}[g(x, y) = 1] > 1 - 1/(4 \cdot \mathsf{rk}(g))$. In this case, we obtain a monochromatic rectangle for g using an observation of Gavinsky and Lovett [GL14], ignoring the cover for $f \circ g$. Otherwise, g is not too biased and we can apply the above discussion, albeit with a loss of $1/(4 \cdot \mathsf{rk}(g))$ in the final bound.

Organization. We recall relevant definitions in Section 2 and prove Lemma 5 in Section 3. Theorem 3 and Corollary 4 are proved in Section 4.

2 Preliminaries

For shorthand, we denote by $[n]$ the set $\{1, \ldots, n\}$. For an element $x \in \mathcal{X}^n$, we will refer to $x_1, \ldots, x_{i-1}$ by $x_{<i}$ and similarly, we will refer to $x_i, \ldots, x_n$ by $x_{\geq i}$. Similarly, for a random variable X taking values in $\mathcal{X}^n$, we use $X_{<i}$ to denote $X_1, \ldots, X_{i-1}$. All logarithms will be taken base 2. Given a distribution $p(x)$ we use $\mathsf{supp}(p(x))$ to denote the set of points in the support of $p(x)$. We recall some relevant mathematical facts, which we will use later.

Definition 6 (Entropy). *Given a random variable A distributed according $p(a)$ the entropy of X is given by*

$$H(A) := \mathop{\mathbf{E}}_{p(a)} \left[\log \frac{1}{p(a)} \right].$$

Fact 7. *If A has finite support, then $H(A) \leq \log |\mathsf{supp}(p(a))|$, with equality if $p(a)$ is the uniform distribution.*

Given two jointly distributed random variables A and B distributed according to $p(ab)$, the conditional entropy of B given A is defined as

$$H(B|A) := \mathop{\mathbf{E}}_{p(ab)} \left[\log \frac{1}{p(b|a)} \right].$$

It is well-known that $H(B|A) \leq H(B)$. We also recall the chain rule for entropy

$$H(AB) = H(A) + H(B|A). \tag{5}$$

Next, we recall the notion of KL-divergence.

Definition 8 (KL-divergence). *Given two probability distributions $p(a)$ and $q(a)$, the* KL-divergence *between p and q is defined as*

$$\mathbf{D}(p||q) := \mathop{\mathbf{E}}_{p(a)} \left[\log \frac{p(a)}{q(a)} \right].$$

Fact 9. *For any two distributions $p(a)$ and $q(a)$, it holds that $\mathbf{D}(p||q) \geq 0$.*

Next, we recall two complexity measures associated with Boolean functions: decision tree complexity and degree. A decision tree of depth d is an adaptive (deterministic) query algorithm, making at most d queries to compute a given function. The algorithm queries variables $x_{i_1}, \ldots, x_{i_d}$ adaptively and outputs a bit based on the values of the variables. We say that a decision tree computes a function f, if on every input x, the algorithm outputs $f(x)$. The decision tree complexity of f, denoted $\mathsf{DT}(f)$ is the least depth of a decision tree among those that compute f. We also recall that for every function $f : \{0,1\}^n \to \{0,1\}$, there exists a unique real, multilinear polynomial

$$q(x) = \sum_{S \subseteq [n]} c_S \cdot \prod_{i \in S} x_i,$$

such that $q(x) = f(x)$ for all $x \in \{0,1\}^n$. The degree of f, denoted $\mathsf{deg}(f)$, is the degree of q.

We conclude this section with some facts regarding communication complexity and its connections to the rank of matrices, whose proofs can be found in [RY20]. First, we recall that rank is sub-additive.

Fact 10. *For two matrices A_1 and A_2, we have $\mathsf{rk}(A_1 + A_2) \leq \mathsf{rk}(A_1) + \mathsf{rk}(A_2)$.*

Next, we note that the communication complexity of a function is at least the logarithm of rank of the corresponding matrix.

Fact 11. *For any function $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, we have $D(g) \geq \lceil \log \mathsf{rk}(g) \rceil$.*

Lastly, we need the fact that a protocol with a small number of leaves can be simulated by a short protocol (see [RY20] Theorem 1.7).

Fact 12. *Given a protocol π with ℓ leaves, there exists a protocol with communication at most $\lceil 2 \log_{3/2} \ell \rceil$ that outputs $\pi(x,y)$ on inputs x and y.*

6

3   Proof of Lemma 5

First, assume that $|\mathbf{E}_{x,y}[g(x,y)] - 1/2| > 1/2 - 1/(4 \cdot \mathsf{rk}(f))$. In this case, we claim g contains a monochromatic rectangle of constant density. Indeed,

$$\left| \mathbf{E}_{x,y}[g(x,y)] - \frac{1}{2} \right| = \max \left\{ \mathbf{Pr}_{x,y}[g(x,y) = 1] - \frac{1}{2}, \mathbf{Pr}_{x,y}[g(x,y) = 0] - \frac{1}{2} \right\},$$

and we can assume without loss of generality that

$$\mathbf{Pr}_{x,y}[g(x,y) = 1] > 1 - \frac{1}{4 \cdot \mathsf{rk}(g)}.$$

Let E be the set of all x such that $\mathbf{Pr}_y[g(x,y) = 1] \leq 1 - 1/(2 \cdot \mathsf{rk}(g))$. We have,

$$\mathbf{Pr}_{x,y}[g(x,y) = 1] \leq \mathbf{Pr}_x[x \in E] \cdot \left(1 - \frac{1}{2 \cdot \mathsf{rk}(g)} \right) + \mathbf{Pr}_x[x \notin E]$$

$$= 1 - \frac{\mathbf{Pr}_x[x \in E]}{2 \cdot \mathsf{rk}(g)},$$

which implies that $\mathbf{Pr}_x[x \in E] \leq 1/2$. Let $x_1, \ldots, x_r \in E^c$ be such that the corresponding rows are maximally linearly independent in M_g. Moreover, let $G = \{y : f(x_i, y) = 1, \forall i \in [r]\}$. By a union bound, we have

$$\mathbf{Pr}_y[y \notin G] \leq r \cdot \frac{1}{2\mathsf{rk}(g)} \leq \frac{1}{2}.$$

We observe that $E^c \times G$ is a monochromatic rectangle of density at least $1/4$. Since $2^{-2T/s} \cdot (4 \cdot \mathsf{rk}(g))^{-2} \leq 1/16 < 1/4$ we have found a monochromatic rectangle of the desired density.

Next, suppose that $|\mathbf{E}_{x,y}[g(x,y)] - 1/2| < 1/2 - 1/(4 \cdot \mathsf{rk}(f))$. We have

$$\frac{1}{4 \cdot \mathsf{rk}(g)} \leq \mathbf{Pr}_{x,y}[g(x,y) = 0], \mathbf{Pr}_{x,y}[g(x,y) = 1] < 1 - \frac{1}{4 \cdot \mathsf{rk}(g)}. \tag{6}$$

For shorthand let s be the sensitivity of g. By definition, there exists an input $z \in \{0,1\}^n$ and a set $S \subseteq [n]$ such that

$$f(z) \neq f(z_{<i}, 1 - z_i, z_{>i}).$$

We may assume without loss of generality that $S \supseteq [s]$, for otherwise, this can be ensured by renaming the coordinates. Let $u(x,y)$ denote the uniform distribution over all inputs $(x,y) \in \mathcal{X}^n \times \mathcal{Y}^n$, and let $p(x,y)$ be a distribution obtained by sampling each (x_i, y_i) randomly and independently subject to $g(x_i, y_i) = z_i$. By Equation (6) we have the following inequality relating the two distributions:

$$\max_{x,y} \frac{p(x_{\leq s}, y_{\leq s})}{u(x_{\leq s}, y_{\leq s})} = \max_{xy} \prod_{i \leq s} \frac{u(x_i, y_i | g(x_i, y_i) = z_i)}{u(x_i, y_i)}$$

$$= \prod_{i \leq s} \frac{1}{\mathbf{Pr}_{u(x_i, y_i)}[g(x_i, y_i) = z_i]} \leq (4 \cdot \mathsf{rk}(g))^s. \tag{7}$$

Recall that $f \circ g$ can be covered with at most 2^T monochromatic rectangles, say $R_1, \ldots, R_{2^T}$. Thus, there exists a rectangle R in the cover with $p(R) \geq 2^{-T}$. Let X and Y be random variables denoting rows and columns of $\mathcal{X}^n$ and $\mathcal{Y}^n$ respectively, where XY is distributed according to $p(x, y | R)$.

Claim 13.

$$\sum_{i \in [s]} H(X_i | X_{<i}, X_{>s}, Y_{>i}) + H(Y_i | X_{<i}, X_{>s}, Y_{>i}) \geq s \log \frac{|\mathcal{X}| \cdot |\mathcal{Y}|}{(4 \cdot \mathsf{rk}(g))^2} - 2T.$$

Proof. Applying the chain rule for entropy we get

$$\sum_{i \in [s]} H(X_i | X_{<i}, X_{>s}, Y_{>i}) + H(Y_i | X_{<i}, X_{>s}, Y_{>i})$$

$$\geq \sum_{i \in [s]} H(X_i | X_{<i}, X_{>s}, Y) + H(Y_i | X, Y_{>i}) = H(X | X_{>s}, Y) + H(Y | X, Y_{>s}).$$

Let $p'(x, y)$ be the distribution obtained by sampling (x_j, y_j) uniformly at random, for each $j \in [s]$, and according to $p(x_j, y_j)$ for each $j \notin S$. Using this notation, we bound the term $H(X | X_{>s}, Y)$ above as follows

$$H(X | Y, X_{>s})$$

$$= \operatorname*{\mathbf{E}}_{p(x,y|R)} \left[\log \frac{1}{p(x | y, x_{>s}, R)} \right]$$

$$= \operatorname*{\mathbf{E}}_{p(x,y|R)} \left[\log \frac{p(R) \cdot p(x_{>s}, y | R)}{p(x, y)} \right]$$

$$\geq \operatorname*{\mathbf{E}}_{p(x,y|R)} \left[\log \frac{p(x_{>s}, y | R)}{(4 \cdot \mathsf{rk}(g))^s \cdot u(x_{\leq s}, y_{\leq s}) \cdot p(x_{>s}, y_{>s})} \right] - T \qquad \text{(Equation (7) and } p(R) \geq 2^{-T})$$

$$= \operatorname*{\mathbf{E}}_{p(x,y|R)} \left[\log \frac{|\mathcal{X}|^s \cdot p(x_{>s}, y | R)}{p'(x_{>s}, y)} \right] - T - s \log(4 \cdot \mathsf{rk}(g))$$

$$= s \log \frac{|\mathcal{X}|}{4 \cdot \mathsf{rk}(g)} + \mathbf{D}(p(x_{>s}, y | R) || p'(x_{>s}, y)) - T \geq s \log \frac{|\mathcal{X}|}{4 \cdot \mathsf{rk}(g)} - T,$$

which follows by Fact 9. A similar calculation shows that $H(Y | X, Y_{>s}) \geq s \log \frac{|\mathcal{Y}|}{4 \cdot \mathsf{rk}(g)} - T$, yielding the desired bound. □

By an averaging argument, we obtain an index $i \in [s]$ and $x_{<i}, x_{>s}, y_{>i}$ such that

$$H(X_i | x_{<i}, x_{>s}, y_{>i}) + H(Y_i | x_{<i}, x_{>s}, y_{>i}) \geq \log \frac{|\mathcal{X}| \cdot |\mathcal{Y}|}{(4 \cdot \mathsf{rk}(g))^2} - \frac{2T}{s}.$$

For shorthand, let

$$A := \mathsf{supp}(p(x_i | x_{<i}, x_{>s}, y_{>i}, R)) \text{ and } B := \mathsf{supp}(p(y_i | x_{<i}, x_{>s}, y_{>i}, R)).$$

Using Fact 7 we conclude that the rectangle given by $A \times B$ has size at least

$$\frac{|\mathcal{X}| \cdot |\mathcal{Y}|}{16 \cdot \mathsf{rk}(g)^2 \cdot 2^{2T/s}}.$$

Moreover, we claim that $A \times B$ is monochromatic for g. Indeed, for any $x_i \in A$, there exists a row $x' \in \mathsf{supp}(p(x | x_{<i}, x_{>s}, y_{>i}, R))$ such that $x'_i = x_i$. Similarly, for any $y_i \in B$, there exists a column $y' \in \mathsf{supp}(p(y | x_{<i}, x_{>s}, y_{>i}, R))$ such that $y'_i = y_i$. In particular, $(x', y') \in R$ and in addition, $x'_j = x_j$ for all $j < i$ and $y'_j = y_j$ for all $j > i$.

Since $y' \in \mathsf{supp}(p(y | x_{<i}, x_{>s}, y_{>i}, R))$, we get $g(x'_t, y'_t) = g(x_t, y'_t) = z_t$ for all $t < i$. Similarly, we have $g(x'_t, y'_t) = g(x'_t, y_t) = z_t$ for all $t > i$. Since $i \in [s]$, if $g(x'_i, y'_i) \neq z_i$, then $f \circ g(x', y') = f(z_{<i}, 1 - z_i, z_{>i}) \neq f(z)$. However, this contradicts the fact that R is monochromatic for $f \circ g$ because for every $x, y \in \mathsf{supp}(p(x, y | R))$, we know that $f \circ g(x, y) = f(z)$.

4   Proof of Theorem 3

At a high level, we apply Lemma 5 repeatedly to find dense monochromatic rectangles and combine this with the arguments of [NW95] to obtain a protocol for g.

Fix some two functions f and g and consider any cover of $f \circ g$ with some 2^T monochromatic rectangles. For shorthand, let s denote the sensitivity of f. Applying Lemma 5, we obtain a monochromatic rectangle R in M_g with density $2^{-2T/s} \cdot (4 \cdot \mathsf{rk}(g))^{-2}$.

By renaming the rows and columns of M_g appropriately, we can rewrite it as

$$\begin{bmatrix} R & A \\ B & Z \end{bmatrix},$$

for some matrices A, B and Z. Now, we observe that

$$\min\left\{ \mathsf{rk}\left(\begin{bmatrix} R & A \end{bmatrix} \right), \mathsf{rk}\left(\begin{bmatrix} R \\ B \end{bmatrix} \right) \right\} \leq \frac{\mathsf{rk}(g) + 3}{2}. \tag{8}$$

Since R has rank one we get

$$\mathsf{rk}\left(\begin{bmatrix} R & A \end{bmatrix} \right) + \mathsf{rk}\left(\begin{bmatrix} R \\ B \end{bmatrix} \right) \leq \mathsf{rk}(A) + \mathsf{rk}(B) + 2 \qquad \text{(by Fact 10)}$$

$$\leq \mathsf{rk}\left(\begin{bmatrix} 0 & A \\ B & Z \end{bmatrix} \right) + 2 \qquad \text{(Gaussian Elimination)}$$

$$\leq \mathsf{rk}\left(\begin{bmatrix} R & A \\ B & Z \end{bmatrix} \right) + 3 \qquad \text{(by Fact 10)}$$

$$= \mathsf{rk}(g) + 3,$$

and Equation (8) follows.

If $\mathsf{rk}\left(\begin{bmatrix} R & A \end{bmatrix} \right) \leq (\mathsf{rk}(g) + 3)/2$, Alice sends a bit to Bob indicating whether or not her input is consistent with the rows of R. Otherwise, Bob sends a bit to Alice indicating whether or not his input is consistent with the columns of R. We can assume without loss of generality that $\mathsf{rk}\left(\begin{bmatrix} R & A \end{bmatrix} \right) \leq (\mathsf{rk}(g) + 3)/2$ as the proof is symmetric.

Let g' and g'' denote the functions encoded by the matrices $\begin{bmatrix} R & A \end{bmatrix}$ and $\begin{bmatrix} B & Z \end{bmatrix}$ respectively. We note that a cover of $M_{f \circ g}$ also gives a cover of both $M_{f \circ g'}$ and $M_{f \circ g''}$. If Alice's input is consistent with the rows of R, the players repeat the above argument using the rectangle cover for $M_{f \circ g'}$. Otherwise, they repeat the argument using the rectangle cover for $M_{f \circ g''}$. In the former case, we have $\mathsf{rk}(g') \leq (\mathsf{rk}(g) + 3)/2$ and in the latter case, the size of $\mathcal{X} \times \mathcal{Y}$ shrinks by a factor of $1 - 2^{-2T/s} \cdot (4 \cdot \mathsf{rk}(g))^{-2}$.

We claim that after $(4 \cdot \mathsf{rk}(g))^3 \cdot 2^{2T/s} + O(\log \mathsf{rk}(g))$ recursive steps either the rank is at most 5 or the size of the matrix is at most 1. Indeed, as long as the $\mathsf{rk}(g) \geq 5$, we have $\mathsf{rk}(g') \leq (\mathsf{rk}(g) + 3)/2 \leq 4 \cdot \mathsf{rk}(g)/5$. Hence, there can only be after $\log_{5/4} \mathsf{rk}(g)$ many steps where the rank reduces by a factor of 4/5. Similarly, there can be only $k = (4 \cdot \mathsf{rk}(g))^3 \cdot 2^{2T/s}$ many steps where the size of the

matrix reduces, since

$$\left(1 - \frac{1}{2^{2T/s}(4 \cdot \mathsf{rk}(g))^2}\right)^k \leq \exp\left(-\frac{k}{2^{2T/s}(4 \cdot \mathsf{rk}(g))^2}\right) = e^{-4 \cdot \mathsf{rk}(g)} \leq \frac{1}{|\mathcal{X}| \cdot |\mathcal{Y}|},$$

where we used the fact that $|\mathcal{X}|$ and $|\mathcal{Y}|$ are both at most $2^{\mathsf{rk}(g)}$ in the last step.

Every leaf of this protocol either corresponds to a size 1 matrix or a matrix of rank at most 5. Thus, with constantly more bits of communication, we get a protocol for g with the following upper bound on the number of leaves:

$$\binom{(4 \cdot \mathsf{rk}(g))^3 \cdot 2^{2T/s} + O(\log \mathsf{rk}(g))}{O(\log \mathsf{rk}(g))} \cdot O(1) \leq O(\mathsf{rk}(g)^3 \cdot 2^{2T/s})^{O(\log \mathsf{rk}(g))}$$

$$\leq 2^{O((T/s + \log \mathsf{rk}(g)) \cdot \log \mathsf{rk}(g))},$$

where all the inequalities hold for c_0 large enough.

By Fact 12 the above protocol can be rebalanced to have communication at most

$$O\left(\left(\frac{T}{s} + \log \mathsf{rk}(g)\right) \cdot \log \mathsf{rk}(g)\right).$$

Since g requires communication at least $D(g)$, we have

$$\left(\frac{T}{s} + \log \mathsf{rk}(g)\right) \cdot \log \mathsf{rk}(g) \geq \Omega(D(g)),$$

and the theorem follows by rearranging.

4.1   Proof of Corollary 4

We start with the following claim relating $\mathsf{rk}(g)$ and $\mathsf{rk}(f \circ g)$.

Lemma 14. *For any two functions $f : \{0,1\}^n \to \{0,1\}$ and $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, it holds that $\mathsf{rk}(f \circ g) \geq (\mathsf{rk}(g) - 1)^{\deg(f)}$.*

Proof. For shorthand, denote by d, the degree of f. By definition, there exists a subset of size d whose corresponding coefficient in the polynomial expansion of f is non-zero. We can assume without loss of generality that this set is $[d]$, otherwise, we can rename the variables to ensure this. Let $u_1, \ldots, u_r$ be a maximal set of linearly independent rows of M_g, and let $x_1, \ldots, x_r$ be the corresponding inputs. Further, define the vectors $\widetilde{u}_1, \ldots, \widetilde{u}_r$, where $\widetilde{u}_i$ is the projection of u_i onto the space orthogonal to the all-ones vector, $\mathbf{1}$. We note that the dimension of $\mathsf{span}(\widetilde{u}_1, \ldots, \widetilde{u}_r)$ is at least $r - 1$.

In what follows, we adopt the following notation for the tensor product of 2 (or more vectors). Given two vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^k$, we denote the tensor product of u with v by $u \otimes v \in \mathbb{R}^{mk}$ where $u \otimes v[i, j] = u(i) \cdot v(j)$.

The key observation is that the projection of the rows of $M_{f \circ g}$ to the space

$$\mathcal{V} := \mathsf{span}(\{v_1 \otimes \ldots \otimes v_d \otimes \underbrace{\mathbf{1} \otimes \ldots \otimes \mathbf{1}}_{n-d \text{ times}} : v_i \in \{\widetilde{u}_1, \ldots, \widetilde{u}_r\}\})$$

10

has full rank. Indeed, consider any function $h : [n] \to [r]$ and let u_h be the row corresponding to the inputs $x_{h(1)}, \ldots, x_{h(n)}$. For any $y_1, \ldots, y_n$, using the multilinear polynomial for f we can write

$$
\begin{aligned}
u_h(y_1, \ldots, y_n) &= f(g(x_1, y_1), \ldots, f(x_n, y_n)) \\
&= \sum_{S \subseteq [n]} \alpha_S \cdot \prod_{i \in S} g(x_i, y_i) \\
&= \sum_{S \subseteq [n]} \alpha_S \cdot \prod_{i \in S} u_{h(i)}(y_i).
\end{aligned}
$$

For any set S, the last quantity above can be written as tensor product. For example, if we let $S = [t]$ then

$$
\begin{aligned}
\prod_{i \in [t]} u_{h(i)}(y_i) &= u_{h(1)} \otimes \ldots u_{h(t)}[y_1, \ldots, y_t] \\
&= u_{h(1)} \otimes \ldots u_{h(t)} \otimes \underbrace{\mathbf{1} \otimes \ldots \otimes \mathbf{1}}_{n-t \text{ times}}[y_1, \ldots, y_n].
\end{aligned}
$$

Applying this to a general set S, we can write

$$
\prod_{i \in S} u_{h(i)}(y_i) = \otimes_{i \in S} u_{h(i)} \otimes_{i \notin S} \mathbf{1}[y_1, \ldots, y_n],
$$

where the subscript is used to denote the vector in the i-th coordinate of the tensor product depending on whether or not $i \in S$.

For any set $S \neq [d]$ of size at most d, the projection of $\otimes_{i \in S} u_{h(i)} \otimes_{i \notin S} \mathbf{1}$ onto $\mathcal{V}$ is zero, since there exists $i \in [d] \setminus S$ such that the vector in the i-th coordinate of the tensor product is $\mathbf{1}$. Moreover, by the definition of degree, $\alpha_S = 0$ for sets S of size larger than d. Lastly, the projection of $\alpha_{[d]} \cdot \otimes_{i \in [d]} u_{h(i)} \otimes_{i=d+1}^n \mathbf{1}$ is exactly $\alpha_{[d]} \cdot \otimes_{i \in [d]} \widetilde{u}_{h(i)} \otimes_{i=d+1}^n \mathbf{1}$.

This establishes that the projection of the rows of $M_{f \circ g}$ to $\mathcal{V}$ has full rank. It follows the rank of $M_{f \circ g}$ is at least the dimension of $\mathcal{V}$, which is at least $(r-1)^d = (\mathsf{rk}(g) - 1)^{\deg(f)}$. $\qquad\square$

Proof of Corollary 4. Recalling Fact 11, we have $D(f \circ g) \geq \lceil \log \mathsf{rk}(f \circ g) \rceil$. Therefore, we can put together Theorem 3 and Lemma 14 to conclude

$$
D(f \circ g) \geq \max \left\{ \mathsf{s}(f) \cdot \left(\frac{\Omega(D(g))}{\log \mathsf{rk}(g)} - \log \mathsf{rk}(g) \right), \deg(f) \cdot \log(\mathsf{rk}(g) - 1) \right\}.
$$

Using the fact that $\max\{a, b\} \geq \lambda \cdot a + (1 - \lambda) \cdot b$, for any $\lambda \in [0, 1]$, we can set $\lambda = \deg(f)/(2\mathsf{s}(f) + \deg(f))$ to get

$$
\begin{aligned}
D(f \circ g) &\geq \frac{\mathsf{s}(f) \cdot \deg(f)}{2\mathsf{s}(f) + \deg(f)} \cdot \left(\frac{\Omega(D(g))}{\log \mathsf{rk}(g)} - \log \mathsf{rk}(g) + 2\log(\mathsf{rk}(g) - 1) \right) \\
&\geq \frac{\mathsf{s}(f) \cdot \deg(f)}{2\mathsf{s}(f) + \deg(f)} \cdot \Omega\left(\frac{D(g)}{\log \mathsf{rk}(g)} + \log \mathsf{rk}(g) \right),
\end{aligned}
$$

where we used the fact that for c_0 large enough, $(\mathsf{rk}(g) - 1)^2 \geq \mathsf{rk}(g)^{3/2}$. $\qquad\square$

5   Concluding Remarks

At a high level, we try to understand the communication complexity of computing $f \circ g$ for arbitrary f and g. We expect that this is $\Omega(\mathsf{DT}(f) \cdot D(g))$, for any sufficiently complex gadget g. Corollary 4 gives the lower bound $\Omega(\min\{\mathsf{s}(f), \mathsf{deg}(f)\} \cdot \sqrt{D(g)}) \geq \Omega(\mathsf{DT}(f)^{1/6} \cdot \sqrt{D(g)})$ which can be seen as progress towards this. Below, we show that for some gadgets this can be further improved.

1. For certain gadgets g, we can obtain $D(f \circ g) \geq \Omega(\mathsf{DT}(f)^{1/3} \cdot \sqrt{D(g)})$ using the notion of *block-sensitivity*, a well-studied [NS92] Boolean function complexity measure. The block-sensitivity of f at $z \in \{0,1\}^n$ is the maximum number of disjoint sets $S_1, \ldots, S_t$ such that for all $i \in [t]$,

$$f(z) \neq f(z^{\oplus S_i}), \text{ where } z_j^{\oplus S_i} = \begin{cases} 1 - z_j, & \text{if } j \in S_i \text{ and} \\ z_j, & \text{otherwise.} \end{cases}$$

   The block-sensitivity of f, denoted $\mathsf{bs}(f)$, is the maximum across all z, of the block-sensitivity of f at z. By definition, $\mathsf{s}(f) \leq \mathsf{bs}(f)$. Moreover, the block-sensitivity is known to give a better upper bound for the decision-tree complexity than the sensitivity. In particular, Midrijānis [Mid04] showed that $\mathsf{DT}(f) \leq \mathsf{bs}(f) \cdot \mathsf{deg}(f)$. By the lower bound in Equation (2) we know that $\mathsf{deg}(f) \leq \mathsf{s}(f)^2 \leq \mathsf{bs}(f)^2$. Hence, $\mathsf{DT}(f) \leq \mathsf{bs}(f)^3$.

   For any function g with the following symmetry property, one can replace $\mathsf{s}(f)$ in Theorem 3 with $\mathsf{bs}(f)$, which together with Corollary 4 implies that $D(f \circ g) \geq \Omega(\mathsf{DT}(f)^{1/3} \cdot \sqrt{D(g)})$. The symmetry property in question is that for any x, there exists $\overline{x} \neq x$ such that $g(x, y) = 1 - g(\overline{x}, y)$, for every y. An example of such a function is the index function, Ind_m. We give a sketch of this claim. Suppose f has block-sensitivity b, achieved at a point $\widetilde{z}$ by sets $S_1, \ldots, S_b$. Consider the function $f' : \{0,1\}^b \to \{0,1\}$ given by

$$f'(z) = f(z'), \text{ where } z_j' = \begin{cases} |\widetilde{z}_j - z_j|, & \text{if } j \in S_1 \cup \ldots \cup S_b \text{ and} \\ \widetilde{z}_j, & \text{otherwise.} \end{cases}$$

   We note that f' has sensitivity b, since $f'(0) = f(\widetilde{z})$, and $f'(0^{\oplus\{i\}}) = f(\widetilde{z}^{\oplus S_i})$. Moreover, any protocol that computes $f \circ g$ can also be used to compute $f' \circ g$ in the following way. Suppose Alice and Bob gets inputs $x_1, \ldots, x_b$ and $y_1, \ldots, y_b$. For each set S_i and coordinate $j \in S_i$, Alice sets $x_j = x_i$ if $\widetilde{z}_j = 0$, and otherwise, sets $x_j = \overline{x}_i$ (from the symmetry property). Bob sets $y_j = y_i$ for each $j \in S_i$. For every coordinate $j \notin S_1 \cup \ldots \cup S_b$, the players arbitrarily fix inputs such that $g(x_j, y_j) = \widetilde{z}_j$. They can now run the protocol for $f \circ g$ to compute $f' \circ g$ and it follows that $D(f' \circ g) \leq D(f \circ g)$. However, Theorem 3 shows that

$$D(f' \circ g) \geq \mathsf{s}(f') \cdot \left(\frac{\Omega(D(g))}{\log \mathsf{rk}(g)} - \log \mathsf{rk}(g) \right) = \mathsf{bs}(f) \cdot \left(\frac{\Omega(D(g))}{\log \mathsf{rk}(g)} - \log \mathsf{rk}(g) \right).$$

2. Anup Rao observed that for certain other gadgets g, such as the inner product gadget IP_m, one can improve Theorem 3 to obtain $D(f \circ g) = \Omega(\mathsf{s}(f) \cdot D(g)/\log \mathsf{rk}_2(g))$, where $\mathsf{rk}_2(g)$ is the rank of M_g over $\mathbb{F}_2$. This can be seen by modifying the proof of Theorem 3 to keep track of $\mathsf{rk}_2(g)$ instead of $\mathsf{rk}(g)$. Each time we find a monochromatic rectangle R for g using Lemma 5, we can recurse on a sub-matrix where either $\mathsf{rk}_2(g)$ goes down by a factor of $4/5$ or the size of the matrix shrinks by the appropriate amount. If the rank over $\mathbb{F}_2$ is at most 5, one can just

use 6 bits of communication to compute the function since $D(g) \leq \mathsf{rk}_2(g) + 1$. This calcuation yields

$$D(f \circ g) = \mathsf{s}(f) \cdot \left(\frac{\Omega(D(g))}{\log \mathsf{rk}_2(g)} - \log \mathsf{rk}_2(g) \right).$$

Furthermore, for any gadget satisfying $D(g) = \Omega(\log^2 \mathsf{rk}(g))$, we get that $D(f \circ g) = \Omega(\mathsf{s}(f) \cdot D(g)/\log \mathsf{rk}_2(g))$. In particular, for the innner product gadget, we know that $\mathsf{rk}_2(\mathrm{IP}_m) \leq m$ and $D(\mathrm{IP}_m) = \Omega(m)$.

Acknowledgments. We are grateful to Anup Rao for several helpful discussions and his encouragement to work on the problem. We would like to thank Guangxu Yang for letting us know that Theorem 2 can be strengthened; the proof of Corollary 4 uses his idea. We are also grateful to Paul Beame, Oscar Sprumont and Michael Whitmeyer for helpful conversations and feedback on a draft of this work.

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, 2010.

[BdW02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theor. Comput. Sci.*, 2002.

[BJKS02] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *43rd Symposium on Foundations of Computer Science (FOCS 2002)*. IEEE Computer Society, 2002.

[BRWY13] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS*. IEEE Computer Society, 2013.

[CFK+21] Arkadev Chattopadhyay, Yuval Filmus, Sajin Koroth, Or Meir, and Toniann Pitassi. Query-to-communication lifting using low-discrepancy gadgets. *SIAM Journal on Computing*, 2021.

[CKLM19] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation theorems via pseudo-random properties. *Comput. Complex.*, 2019.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *42nd Annual Symposium on Foundations of Computer Science, FOCS*. IEEE Computer Society, 2001.

[FKNN95] Tomás Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized communication complexity. *SIAM Journal on Computing*, 1995.

[GGKS20] Ankit Garg, Mika Göös, Pritish Kamath, and Dmitry Sokolov. Monotone circuit lower bounds from resolution. *Theory of Computing*, 2020.

[GL14]   Dmitry Gavinsky and Shachar Lovett. En route to the log-rank conjecture: New reductions and equivalent formulations. In *Automata, Languages, and Programming*. Springer Berlin Heidelberg, 2014.

[GPW18]   Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. *SIAM Journal on Computing*, 2018.

[GPW20]   Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. *SIAM J. Comput.*, 2020.

[HJMR10]   Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 2010.

[Hol07]   Thomas Holenstein. Parallel repetition: Simplifications and the no-signaling case. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '07. Association for Computing Machinery, 2007.

[Hua19]   Hao Huang. Induced subgraphs of hypercubes and a proof of the sensitivity conjecture. *Annals of Mathematics*, 2019.

[IR24a]   Siddharth Iyer and Anup Rao. An XOR Lemma for Deterministic Communication Complexity . In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, 2024.

[IR24b]   Siddharth Iyer and Anup Rao. Xor lemmas for communication via marginal information. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC, 2024.

[JRS03]   Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *Automata, Languages and Programming*, 2003.

[KN96]   Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1996.

[Kol16]   Gillat Kol. Interactive compression for product distributions. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 987–998. Association for Computing Machinery, 2016.

[KRW95]   Mauricio Karchmer, Ran Raz, and Avi Wigderson. Super-logarithmic depth lower bounds via the direct sum in communication complexity. *Comput. Complex.*, 1995.

[LMM+22]   Shachar Lovett, Raghu Meka, Ian Mertz, Toniann Pitassi, and Jiapeng Zhang. Lifting with Sunflowers. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

[Lov14]   Shachar Lovett. Communication is bounded by root of rank. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, STOC '14. Association for Computing Machinery, 2014.

[LS88]   László Miklós Lovász and Michael E. Saks. Lattices, mobius functions and communications complexity. *[Proceedings 1988] 29th Annual Symposium on Foundations of Computer Science*, 1988.

[Mid04]   Gatis Midrijanis. Exact quantum query complexity for total boolean functions, 2004.

[MM22]   Yahel Manor and Or Meir. Lifting with inner functions of polynomial discrepancy. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2022*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.

[Nis89]   N. Nisan. Crew prams and decision trees. In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery, 1989.

[NS92]   Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery, 1992.

[NW95]   Noam Nisan and Avi Wigderson. On rank vs. communication complexity. *Comb.*, 1995.

[PR18]   Toniann Pitassi and Robert Robere. Lifting nullstellensatz to monotone span programs over any field. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. Association for Computing Machinery, 2018.

[Rao11]   Anup Rao. Parallel repetition in projection games and a concentration bound. *SIAM Journal on Computing*, 2011.

[Raz92]   Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 1992.

[Raz95]   Ran Raz. A parallel repetition theorem. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '95. Association for Computing Machinery, 1995.

[RM99]   Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Comb.*, 1999.

[RY20]   Anup Rao and Amir Yehudayoff. *Communication Complexity: and Applications*. Cambridge University Press, 2020.

[She11]   Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 2011.

[She18]   Alexander A. Sherstov. Compressing interactive communication under product distributions. *SIAM Journal on Computing*, 2018.

[SK87]   Georg Schnitger and Bala Kalyanasundaram. The probabilistic communication complexity of set intersection. In *Proceedings of the Second Annual Conference on Structure in Complexity Theory*. IEEE Computer Society, 1987.

[ST24]   Benny Sudakov and István Tomon. Matrix discrepancy and the log-rank conjecture. *Mathematical Programming*, 2024.

[Yan24]   Guangxu Yang. Private Communication, November 2024.

[Yu22]   Huacheng Yu. Strong XOR lemma for communication with bounded rounds : (extended abstract). In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS*. IEEE, 2022.