# Log-Rank and Lifting for AND-Functions

Alexander Knop
aaknop@gmail.com
Mathematics, University of California, San Diego
La Jolla, CA, USA

Shachar Lovett
slovett@ucsd.edu
Computer Science and Engineering, University of California, San Diego
La Jolla, CA, USA

Sam McGuire
shmcguir@eng.uscd.edu
Computer Science and Engineering, University of California, San Diego
La Jolla, CA, USA

Weiqiang Yuan
yuanwq17@mails.tsinghua.edu.cn
Institute for Interdisciplinary Information Sciences, Tsinghua University
Beijing, China

## ABSTRACT

Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function, and let $f_\wedge(x,y) = f(x \wedge y)$ denote the *AND-function* of $f$, where $x \wedge y$ denotes bit-wise AND. We study the deterministic communication complexity of $f_\wedge$ and show that, up to a $\log n$ factor, it is bounded by a polynomial in the logarithm of the real rank of the communication matrix of $f_\wedge$. This comes within a $\log n$ factor of establishing the log-rank conjecture for AND-functions with *no assumptions* on $f$. Our result stands in contrast with previous results on special cases of the log-rank conjecture, which needed significant restrictions on $f$ such as monotonicity or low $\mathbb{F}_2$-degree. Our techniques can also be used to prove (within a $\log n$ factor) a *lifting theorem* for AND-functions, stating that the deterministic communication complexity of $f_\wedge$ is polynomially related to the *AND-decision tree complexity* of $f$.

The results rely on a new structural result regarding boolean functions $f : \{0,1\}^n \to \{0,1\}$ with a sparse polynomial representation, which may be of independent interest. We show that if the polynomial computing $f$ has few monomials then the set system of the monomials has a small hitting set, of size poly-logarithmic in its sparsity. We also establish extensions of this result to multi-linear polynomials $f : \{0,1\}^n \to \mathbb{R}$ with a larger range.

## CCS CONCEPTS

• **Theory of computation → Communication complexity**; **Oracles and decision trees**.

## KEYWORDS

AND-functions, Lifting, Log-rank conjecture

**ACM Reference Format:**
Alexander Knop, Shachar Lovett, Sam McGuire, and Weiqiang Yuan. 2021. Log-Rank and Lifting for AND-Functions. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC '21), June 21–25, 2021, Virtual, Italy*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3406325.3450999


*STOC '21, June 21–25, 2021, Virtual, Italy*

ACM ISBN 978-1-4503-8053-9/21/06…$15.00
https://doi.org/10.1145/3406325.3450999

## 1 INTRODUCTION

Communication complexity has seen rapid development in the last couple of decades. However, most of the celebrated results in the field are about the communication complexity of important *concrete functions*, such as set disjointness [29] and gap Hamming distance [4]. Unfortunately, the understanding of communication complexity of *arbitrary functions* is still lacking.

Probably the most famous problem of this type is the log-rank conjecture [21]. It speculates that given any total boolean communication problem, its deterministic communication complexity is polynomially related to the logarithm of the real rank of its associated communication matrix. Currently, there is an exponential gap between the lower and upper bounds relating to the log-rank conjecture. The best known upper bound [22] states that the communication complexity of a boolean function $F$ is at most $O(\sqrt{\operatorname{rank}(F)} \log \operatorname{rank}(F))$, where $\operatorname{rank}(F)$ denotes the real rank of the communication matrix of $F$. On the other hand, the best known lower bound [15] states that there exist a boolean function $F$ with communication complexity $\Omega(\log^2(\operatorname{rank}(F)))$.

Given this exponential gap and lack of progress for general communication problems, many works [5, 7, 9, 11, 12, 14, 15, 17, 18, 20, 23–25, 27, 30, 34, 35, 37, 38] focused on the communication complexity of functions with some restricted structure. In particular, the study of composed functions was especially successful, and produced the so-called lifting method, which connects query complexity measures of boolean functions with communication complexity measures of their corresponding communication problems.

Concretely, given a boolean function $f : \{0,1\}^n \to \{0,1\}$ and a *gadget* $g : \{0,1\}^\ell \times \{0,1\}^m \to \{0,1\}$, the corresponding lifted function is the following communication problem: Alice gets as input $x \in (\{0,1\}^\ell)^n$, Bob gets as input $y \in (\{0,1\}^m)^n$, and their goal is to compute the composed function $f \circ g^n$, defined as $(f \circ g^n)(x,y) = f(g(x_1,y_1),\ldots,g(x_n,y_n))$. Lifting theorems allow to connect query complexity measures of the underlying boolean function $f$ with communication complexity measures of the composed function. Figure 1 lists some notable examples.

| Gadget | QM | CM | Total Functions | Reference |
|---|---|---|---|---|
| $\mathrm{Ind}_m$ | $\mathrm{P}^{\mathrm{dt}}$ | $\mathrm{P}^{\mathrm{cc}}$ | No | [28] |
| | $\mathrm{BPP}^{\mathrm{dt}}$ | $\mathrm{BPP}^{\mathrm{cc}}$ | No | [14] |
| $\mathrm{IP}_{\log m}$ | $\mathrm{P}^{\mathrm{dt}}$ | $\mathrm{P}^{\mathrm{cc}}$ | No | [6] |
| | $\mathrm{BPP}^{\mathrm{dt}}$ | $\mathrm{BPP}^{\mathrm{cc}}$ | No | [5] |
| $\mathrm{EQ}_{\log m}$ | $\mathrm{P}^{\wedge\text{-dt}}$ | $\mathrm{P}^{\mathrm{cc}}$ | No | [25] |
| $\oplus$ | $\mathrm{P}^{\oplus\text{-dt}}$ | $\mathrm{P}^{\mathrm{cc}}$ | Yes | [17] |
| $g$ | deg | rank | Yes | [31] |

**Figure 1: Query-to-communication lifting theorems. QM stands for query model and CM stands for communication model. The parameter $m$ is polynomial in $n$; $g$ in the last line is any function that has as sub-functions both an AND and an OR. $\mathrm{P}^{\mathrm{cc}}$ denotes determenistic communication complexity, $\mathrm{P}^{\mathrm{dt}}$ denotes decision tree complexity, $\mathrm{BPP}^{\mathrm{dt}}$ denotes the probabilistic decision tree complexity with bounded error, $\mathrm{BPP}^{\mathrm{cc}}$ denotes the probabilistic communication complexity with bounded error, $\mathrm{P}^{\wedge\text{-dt}}$ denotes AND-decision tree complexity, deg denotes the real degree, and rank denotes the real rank. The "total functions" column contains "Yes" if the corresponding theorem holds only for total functions, and "No" if it holds for partial functions and search relations.**

Of particular interest to us are lifting theorems with very simple gadgets. The reason for that is twofold. First, using complex gadgets (such as inner product or indexing) yields sub-optimal bounds in applications. A second and perhaps more important reason is that the study of composed functions with complex gadgets does not bring us any closer towards the understanding of general communication problems. This is because the corresponding lifting theorems connect the communication complexity of the lifted function to well-studied query measures of the underlying boolean function (such as decision tree complexity, or degree as a real polynomial), and hence does not shed new light on general communication problems.

Thus, in this paper we consider gadgets which are as simple as they could be — one-bit gadgets. In fact, there are only two non-equivalent one-bit gadgets: one-bit XOR, which yields XOR-functions; and one-bit AND, which yields AND-functions. As we shortly discuss, they naturally correspond to query models which extend the standard ones: parity-decision trees and AND-decision trees.

*XOR-functions.* These functions have been studied in several works [17, 18, 20, 23, 24, 30, 34, 35, 37, 38]. Given a boolean function $f : \{0,1\}^n \to \{0,1\}$, its corresponding XOR-function is $f_\oplus = f \circ \oplus^n$, defined as $f_\oplus(x,y) = f(x \oplus y)$. A natural query measure corresponding to the communication complexity of XOR-functions is the *Parity-Decision Tree* (PDT) model. This model is an extension of the standard decision tree model, where nodes can query an arbitrary parity of the bits. To see the connection, note that if $f$ can be computed by a PDT of depth $d$ (denoted by $\mathrm{P}^{\oplus\text{-dt}}(f) = d$), then $f_\oplus$ has a communication protocol of complexity $2d$. This is by simulating the computation in the PDT: whenever the PDT needs to compute the parity of $x \oplus y$ on some set $S$ of coordinates, each player computes the corresponding parity on their input, and then they exchange the answers, which allows to compute the corresponding parity on $x \oplus y$ as well, and continue to traverse the tree. Thus we have $\mathrm{P}^{\mathrm{cc}}(f_\oplus) \le 2\mathrm{P}^{\oplus\text{-dt}}(f)$.

In the other direction, [17] proved that $\mathrm{P}^{\oplus\text{-dt}}(f)$ is at most a polynomial in the communication complexity of $f_\oplus$. That is, $\mathrm{P}^{\oplus\text{-dt}}(f) \le \mathrm{poly}\,(\mathrm{P}^{\mathrm{cc}}(f_\oplus))$. Thus, the two measures are equivalent, up to polynomial factors.

If one considers the log-rank conjecture for XOR-functions, then a simple observation [34] is that the rank of the communication matrix of $f_\oplus$ is equal to the Fourier sparsity of $f$. Thus, in order to prove the log-rank conjecture for XOR-functions it is sufficient to show that $\mathrm{P}^{\oplus\text{-dt}}(f)$ is at most a polynomial in the log of the Fourier sparsity of $f$. Unfortunately, the latter relation is currently unknown.

*AND-functions.* The goal of this paper is to develop an analogous theory of AND-functions. Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. Its corresponding AND-function is $f_\wedge = f \circ \wedge^n$, defined as $f_\wedge(x,y) = f(x \wedge y)$. Similar to the case of XOR-functions, there is a corresponding natural query model, *AND-Decision Tree* (ADT), where each node in the decision tree can query an arbitrary AND of the input bits. We denote by $\mathrm{P}^{\wedge\text{-dt}}(f)$ the minimal depth of an ADT computing $f$. Also here, efficient ADTs for $f$ imply efficient communication protocols for $f_\wedge$, where $\mathrm{P}^{\mathrm{cc}}(f_\wedge) \le 2\mathrm{P}^{\wedge\text{-dt}}(f)$. Our main focus in this work is

(i) lifting theorems for AND-functions, and
(ii) the log-rank conjecture for AND-functions.

Concretely, we will show that assuming that $f_\wedge$ has either (i) efficient deterministic communication protocol or (ii) low rank, then $f$ has an efficient ADT. As we will shortly see, understanding both questions is directly related to understanding the monomial structure of polynomials computing boolean functions.

## 1.1 Main Results

Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. It is computed by a unique multi-linear polynomial over the reals. That is, $f(x) = \sum_s f_s \prod_{i \in s} x_i$, where $s \subseteq [n]$ and $f_s \in \mathbb{R}$ are real-valued coefficients; note that this is not the Fourier representation of $f$, as we are working with $\{0,1\}^n$ instead of $\{-1,1\}^n$. The sparsity of $f$, denoted $\mathrm{spar}(f)$, is the number of nonzero coefficients in the decomposition. This is related to AND-functions, as a simple observation (Claim 4.1) is that this also equals the rank of its communication matrix, namely $\mathrm{rank}(f_\wedge) = \mathrm{spar}(f)$.

Before describing our results, we need one more definition. Let $\mathcal{F}$ be a set system (family of sets). A set $H$ is a *hitting set* for $\mathcal{F}$ if it intersects all the sets in $\mathcal{F}$. Of particular interest to us are set systems that correspond to the monomials of boolean functions. Given a boolean function $f$, define $\mathcal{M}(f) = \{s : f_s \neq 0, s \neq 0\}$ to be the set system of the non-constant monomials of $f$. We exclude the constant term as it is irrelevant for the purpose of constructing hitting sets, and it simplifies some of the later arguments. Note that $|\mathcal{M}(f)| \in \{\mathrm{spar}(f), \mathrm{spar}(f) - 1\}$.

Our main combinatorial result is that set systems corresponding to the monomials of boolean functions have small hitting sets. (Note that an upper bound on the hitting set in terms of degree was known before [8].)

**Theorem 1.1.** Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function with sparsity $\mathrm{spar}(f) = r$. Then there exists a hitting set $H$ for $\mathcal{M}(f)$ of size $|H| = O((\log r)^5)$.

This result can be seen as an analog of a similar result for union-closed families. A set system $\mathcal{F}$ is union-closed if it is closed under taking unions; namely, if $S_1, S_2 \in \mathcal{F}$ then also $S_1 \cup S_2 \in \mathcal{F}$. A famous conjecture of Frankl [10] is that in any union-closed family $\mathcal{F}$ there is an element which belongs to at least half the sets in the set system. Assume $|\mathcal{F}| = r$; the best known result in this direction is that $\mathcal{F}$ has a hitting set of size $\log(r)$ [19], which implies that one of its elements belongs to a $1/\log(r)$ fraction of sets in the set system. We view Theorem 1.1 as a qualitative extension of this result to more general set systems.

Our main application of Theorem 1.1 is a near-resolution of the log-rank conjecture for AND-functions. Our bounds nearly match the conjectured bounds (poly-log in the rank), except for an extra $\log(n)$ factor that we are currently unable to eliminate.

**Theorem 1.2** (Log-rank Theorem for AND-functions)**.** Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. Let $r = \mathrm{spar}(f) = \mathrm{rank}(f_\wedge)$. Then $f$ can be computed by an AND-decision tree of depth

$$\mathrm{P}^{\wedge\text{-dt}}(f) = O((\log r)^5 \cdot \log n).$$

In particular, the deterministic communication complexity of $f_\wedge$ is bounded by

$$\mathrm{P}^{\mathrm{cc}}(f_\wedge) = O((\log r)^5 \cdot \log n).$$

Note that if $f : \{0,1\}^n \to \{0,1\}$ is a function of sparsity at least $n^{0.1}$, say, then Theorem 1.2 proves the log-rank conjecture for its corresponding AND-function. Thus, the only remaining obstacle is to extend the result to very sparse functions.

Observe that Theorem 1.2 implies a lifting theorem for AND-functions. Assume that $f_\wedge$ has deterministic communication complexity $C$. The rank of the associated communication matrix is then at most $2^C$, which by Theorem 1.2 gives an ADT for $f$ of depth $O(C^5 \log n)$. We can improve the exponent 5 to 3 by directly exploiting the existence of a communication protocol.

**Theorem 1.3** (Lifting Theorem for AND-functions)**.** Let $f$ be a boolean function from $\{0,1\}^n$ to $\{0,1\}$, and let $C = \mathrm{P}^{\mathrm{cc}}(f_\wedge)$ denote the deterministic communication complexity of its corresponding AND-function. Then $f$ can be computed by an AND-decision tree of depth

$$\mathrm{P}^{\wedge\text{-dt}}(f) = O(C^3 \cdot \log n).$$

## 1.2 Proof Overview

We first discuss how our combinatorial theorem (Theorem 1.1) implies the log-rank theorem (Theorem 1.2). It relies on showing that sparse boolean functions have efficient AND-decision trees (ADTs).

Let $f$ be a boolean function with $\mathrm{spar}(f) = r$. Our goal is to construct an ADT for $f$ of depth $\mathrm{poly}(\log r) \cdot \log(n)$. This directly implies Theorem 1.2, as the sparsity of $f$ equals the rank of its AND-function $f_\wedge$, and an ADT for $f$ of depth $d$ implies a protocol for $f_\wedge$ which sends $2d$ bits.

It will be convenient to first consider another model of decision trees, called *zero decision trees*. A (standard) decision tree computing $f$ has zero decision tree complexity $d$, if any path from root to leaf in it queries at most $d$ variables which evaluate to 0. We denote by $\mathrm{P}^{0\text{-dt}}(f)$ the minimal such $d$ over all decision trees that compute $f$. It is shown in [25] (see also Claim 4.4) that ADT complexity and zero DT complexity are tightly connected. Concretely, for any boolean function $f$ they show that

$$\mathrm{P}^{0\text{-dt}}(f) \le \mathrm{P}^{\wedge\text{-dt}}(f) \le \mathrm{P}^{0\text{-dt}}(f) \cdot \lceil \log(n+1) \rceil.$$

Thus, we will show that $\mathrm{P}^{0\text{-dt}}(f) \le \mathrm{poly}(\log r)$, which implies our target bound of $\mathrm{P}^{\wedge\text{-dt}}(f)$.

Theorem 1.1 gives that there is a hitting set size $h = \mathrm{poly}(\log r)$ which intersects all the monomials of $f$. In particular, there is a variable $x_i$ that intersects at least a $1/h$ fraction of the monomials of $f$. The decision tree will first query $x_i$, and then branch depending on whether $x_i = 0$ or $x_i = 1$. We use the simple fact that the sparsity of $f$ cannot increase when variables are fixed, and continue this process, until the value of the function is determined. Observe that every time that we query a variable and get 0, we eliminates a $1/h$ fraction of the monomials. If we get a 1 the number of monomials can either stay the same or decrease, but it cannot increase. So, as $f$ starts with $r$ monomials, we get that the maximal number of 0s queried before all monomials are eliminated is at most $h \cdot \log(r)$. Hence $\mathrm{P}^{0\text{-dt}}(f) \le h \cdot \log(r) = \mathrm{poly}(\log r)$, as claimed.

Thus, from now on we focus on proving Theorem 1.1. Let $f$ be a boolean function of sparsity $r$, and let $\mathcal{M}(f)$ denote the set system of its monomials. We consider four complexity measures associated with it:

(1) The hitting set complexity (HSC) is the minimal size of a hitting set for it. This is what we are trying to bound, and can be phrased as an covering integer program.
(2) The fractional hitting set complexity (FHSC) is the fractional relaxation for HSC. Here, we want a distribution over variables that hits every monomial with high probability, which can be phrased as a fractional covering linear program.
(3) The fractional monotone block sensitivity (FMBS) is the dual linear program. The reason for the name would become clear soon. It can be phrased as a fractional packing linear program.
(4) The monotone block sensitivity (MBS) is the integral version of FMBS. It equals the maximal number of pairwise disjoint monomials in $f$. Equivalently, it is block sensitivity of $f$ at $0^n$. It can be phrased as a packing integer program.

More generally, given $s \subseteq [n]$, let $f_s$ denote the restriction of $f$ given by setting $x_i = 1$ for all $i \in s$. It will be convenient to identify $s$ with its indicator vector $1_s \in \{0,1\}^n$. Thus, for $z \in \{0,1\}^n$, we denote by $f_z$ the restriction of $f$ to the 1s in $z$. Define $\mathrm{HSC}(f,z)$, $\mathrm{FHSC}(f,z)$, $\mathrm{FMBS}(f,z)$, $\mathrm{MBS}(f,z)$ to be the above four measures for the monomials of $f_z$. It is simple to observe (see Claim 2.16) that for each $z$ we have:

$$\mathrm{MBS}(f,z) \le \mathrm{FMBS}(f,z) = \mathrm{FHSC}(f,z) \le \mathrm{HSC}(f,z).$$

Note that hitting set complexity is an analogue of certificate complexity, monotone block sensitivity is an analogue of block sensitivity; and fractional hitting set complexity and fractional monotone block sensitivity are analogues of fractional certificate complexity and fractional block sensitivity defined by [1, 32].

As a first step, we use existing techniques in boolean function analysis techniques to bound $\mathrm{MBS}(f,z)$ in terms of the sparsity of $f$. Lemma 3.1 shows that

$$\mathrm{MBS}(f,z) \le O((\log \mathrm{spar}(f_z))^2) \le O((\log r)^2).$$

Thus, to complete the picture, we need to show that if $\mathrm{MBS}(f,z)$ is low then so is $\mathrm{HSC}(f,z)$. This however is false, if one compares them point wise (for a single $z$). However, we show that the measures are equivalent (up to polynomial factors) if instead we consider their maximal value over all $z$. Define

$$\mathrm{MBS}(f) = \max_{z\in\{0,1\}^n} \mathrm{MBS}(f,z)$$

and similarly define $\mathrm{FMBS}(f), \mathrm{FHSC}(f), \mathrm{HSC}(f)$. Lemma 3.2 shows that

$$\mathrm{FMBS}(f) = O(\mathrm{MBS}(f)^2),$$

linear programming duality gives $\mathrm{FHSC}(f) = \mathrm{FMBS}(f)$, and we show in Lemma 3.3 that

$$\mathrm{HSC}(f) = O(\mathrm{FHSC}(f)\cdot \log r).$$

This completes the proof of Theorem 1.1.

We also briefly discuss Theorem 1.3. The improved exponent is obtained by using the bound $\mathrm{MBS}(f) = O(\mathrm{P}^{\mathrm{cc}}(f_\wedge))$, which we prove in Corollary 4.9. Its proof is based on the observation that if $\mathrm{MBS}(f) = b$ then $f_\wedge$ embeds as a sub-function unique disjointness on $b$ bits, and combine it with known lower bounds on the communication complexity of unique disjointness.

### 1.3 Generalizations

Several of our definitions and techniques readily extend to non-boolean functions, namely to functions $f : \{0,1\}^n \to \mathbb{R}$. We refer the reader to Section 2 for the relevant definitions and Section 5 for a detailed discussion of the generalized results. Here, we briefly state some of the results.

**Theorem 1.4.** Let $f : \{0,1\}^n \to \mathbb{R}$ be a multlinear polynomial with sparsity $r$. Suppose $\mathrm{MBS}(f) = m$. Then the hitting set complexity of $f$ is bounded by

$$\mathrm{HSC}(f) = O(m^2 \log r).$$

**Theorem 1.5.** Let $f : \{0,1\}^n \to S$ for $S \subset \mathbb{R}$. Assume that $\mathrm{spar}(f) = r$ and $|S| = s$. Then the hitting set complexity of $f$ is bounded by

$$\mathrm{HSC}(f) = O(s^4 (\log r)^5).$$

*Acknowledgements.* S.L. thanks Kaave Hosseini, who was involved in early stages of this work. S.M. thanks Russell Impagliazzo for useful discussions throughout the course of this work.

## 2 PRELIMINARIES

This section introduces a number of complexity measures used in the proofs of our main results. We start by collecting some simple definitions, proceed to define the complexity measures, and then provide some examples which clarify some aspects of these definitions.

Throughout this section, fix a boolean function $f : \{0,1\}^n \to \mathbb{R}$. We identify elements of $\{0,1\}^n$ with subsets of $[n]$. Namely, we identify $z \in \{0,1\}^n$ with the set $\{i : z_i = 1\}$, and shorthand $[n]\setminus z = \{i : z_i = 0\}$. Given two inputs $z, w \in \{0,1\}^n$ we denote by $z\vee w$ their union and by $z\wedge w$ their intersection. The partial order on $\{0,1\}^n$ is defined by the relation $z \le w$, satisfied precisely when $z$ is a subset of $w$. Define $f_z : \{0,1\}^{[n]\setminus z} \to \mathbb{R}$ to be the restriction of $f$ to inputs which are consistent with the 1s in $z$; namely $f_z(w) = f(z\vee w)$. Define $\mathcal{W}(f,z) = \left\{w \in \{0,1\}^{[n]\setminus z} : f(z) \ne f(z\vee w)\right\}$ which is equivalent to $\mathcal{W}(f,z) = \left\{w \in \{0,1\}^{[n]\setminus z} : f_z(w) \ne f_z(0)\right\}$.

Recall also the notation from the proof overview. Any function $f : \{0,1\}^n \to \mathbb{R}$ can be written uniquely as a multilinear real-valued polynomial $f(x) = \sum_{s\subseteq[n]} \alpha_s \prod_{i\in s} x_i$ (note that this is *not* the Fourier expansion, as here the variables are $\{0,1\}$-valued whereas in the Fourier expansion they are $\{\pm1\}$-valued). The *sparsity* of $f$, denoted $\mathrm{spar}(f)$, is the number of nonzero coefficients in the polynomial expansion of $f$. Next, let $\mathcal{M}(f) = \{s \subseteq [n] : \alpha_s \ne 0,\ s \ne 0^n\}$ denote the set system of non-zero, non-constant monomials in $f$ when written as a multilinear polynomial. We emphasize that the coefficient $\alpha_\emptyset$ is not included in $\mathcal{M}(f)$; $\alpha_\emptyset$ is inessential, since we are interested in hitting sets for monomials and $\emptyset$ is trivially hit by any set. Observe that $|\mathcal{M}(f)| \in \{\mathrm{spar}(f), \mathrm{spar}(f)-1\}$.

For any set system $\mathcal{F}$ over $[n]$, an element $z \in \mathcal{F}$ is *minimal* if there does not exist $w \in \mathcal{F}$ with $w < z$.

**Claim 2.1.** Fix $f : \{0,1\}^n \to \mathbb{R}$, $z \in \{0,1\}^n$ and $\mathcal{W}(f,z)$, $\mathcal{M}(f_z)$ as above. Then, for any $w \in \{0,1\}^n$, $w$ is a minimal element in $\mathcal{W}(f,z)$ if and only if $w$ is a minimal element in $\mathcal{M}(f_z)$.

Proof. We assume for simplicity that $z = \emptyset$ so that $f_z(w) = f(w)$, $f(\emptyset) = \alpha_\emptyset$ and write $\mathcal{W} = \mathcal{W}(f,\emptyset)$. Suppose $w \in \mathcal{M}(f)$ is a minimal element. Writing $f$ as a multilinear polynomial, we get $f(w) = \sum_{u\le w} \alpha_u$. Since $\alpha_w$ is minimal, $f(w) = \alpha_\emptyset + \alpha_u$ and so $f(w) \ne f(\emptyset)$ and $w \in \mathcal{W}$. Additionally, $w$ is minimal in $\mathcal{W}$ because if $w' < w$ then the non-constant terms of $f(w') = \sum_{u\le w'} \alpha_u$ are all 0, hence $f(w') = f(0)$ and $w' \notin \mathcal{W}$.

In the other direction, suppose $w \in \mathcal{W}$ is a minimal element. Assume there is $w' < w$ in $\mathcal{M}(f)$; choosing such a minimal $w'$, we would get $f(w') \ne f(0)$ which violates the minimality of $w$. Similarly, if $w \notin \mathcal{M}(f)$ then we get $f(w) = \sum_{u\le w} \alpha_u = f(0)$, which violates the assumption that $w \in \mathcal{W}$. Thus $w$ is a minimal element in $\mathcal{M}(f)$. □

### 2.1 Monotone Block Sensitivity

First, we consider *monotone block sensitivity*, a variant of the standard notion of *block sensitivity* due to Nisan and Szegedy [26]. In a nutshell, this is a "directed" restriction of block sensitivity, where we can only change an input by flipping 0's to 1's. We also define MBS (and all other complexity measures introduced later in this section) with respect to real-valued functions over $\{0,1\}^n$. This differs from

block sensitivity, which is usually (though not always [13]) studied in the context of boolean-valued functions. The generalization to real-valued $f$ will be immaterial to some of our proofs, permitting us to draw more general conclusions regarding the monomial structure of multilinear polynomials; see Section 5 for more details.

Say that two inputs $z, w$ are disjoint if $z \wedge w = 0^n$; namely, their corresponding sets are disjoint.

**Definition 2.2** (Monotone block sensitivity)**.** For $f : \{0,1\}^n \to \mathbb{R}$ and $z \in \{0,1\}^n$, the *monotone block sensitivity of $f$ at $z$*, denoted $\mathrm{MBS}(f,z)$, is the largest integer $k$ such that there exist $k$ pairwise disjoint inputs $w_1, \ldots, w_k \in \mathcal{W}(f,z)$. We denote $\mathrm{MBS}(f) = \max_z \mathrm{MBS}(f,z)$.

For two motivating examples, observe that for the $n$-bit AND and OR functions we have $\mathrm{MBS}(\mathrm{AND}) = 1$ and $\mathrm{MBS}(\mathrm{OR}) = n$, respectively.

**Remark 2.3.** We emphasize that $\mathcal{W}(f,z) \subseteq \{0,1\}^{[n]\setminus z}$, so each $w_i$ is disjoint from $z$. This corresponds to the standard definition of block sensitivity where we restrict each block $w_i$ to be disjoint from the support of $z$.

**Remark 2.4.** Suppose $w_1, \ldots, w_k$ are minimal witnesses that $\mathrm{MBS}(f,z) = k$ in the sense that for any $i \in [k]$ there is no $w_i' < w_i$ so that $w_i' \in \mathcal{W}(f,z)$. Then by Claim 2.1, each $w_i$ is a minimal element in $\mathcal{M}(f_z)$.

As alluded to in the proof overview, $\mathrm{MBS}(f,z)$ can be phrased as the value of a particular set packing linear program (LP). Fixing $z$, write $\mathcal{W} = \mathcal{W}(f,z)$. The program optimizes over variables $a_w$ for each $w \in \mathcal{W}$.

$$\begin{aligned} \text{maximize} \quad & \sum_{w \in \mathcal{W}} a_w \\ \text{subject to} \quad & \sum_{w \in \mathcal{W} : w_i = 1} a_w \le 1 \quad \text{for all } i \in [n] \\ & a_w \in \{0,1\} \quad \text{for all } w \in \mathcal{W} \end{aligned}$$

*Fractional monotone block sensitivity* (FMBS) is obtained by relaxing the constraints in the above LP, allowing variables $a_w$ to assume non-integral values in $[0,1]$. We use an alternative formulation of FMBS whose equivalence to the LP formulation is simple to verify.

**Definition 2.5** (Smooth distribution)**.** A distribution $\mathcal{D}$ over $\{0,1\}^n$ is said to be $p$-smooth if for any $i \in [n]$ it holds that

$$\Pr_{w \sim \mathcal{D}}[w_i = 1] \le p.$$

**Definition 2.6** (Fractional monotone block sensitivity)**.** The fractional monotone block sensitivity of a function $f : \{0,1\}^n \to \mathbb{R}$ at an input $z \in \{0,1\}^n$, denoted $\mathrm{FMBS}(f,z)$, is equal to $1/p$, where $p > 0$ is the smallest number for which there exists a $p$-smooth distribution $\mathcal{D}$ supported on a subset of $\mathcal{W}(f,z)$. We denote $\mathrm{FMBS}(f) = \max_z \mathrm{FMBS}(f,z)$.

**Remark 2.7.** To see the equivalence between this definition of FMBS and the LP formulation, notice that a solution

$$\{a_w \;:\; w \in \mathcal{W}(f,z)\}$$

to the LP with $s = \sum a_w$ gives rise to a $1/s$-smooth distribution $\mathcal{D}$ over $\mathcal{W}(f,z)$ via $\mathcal{D}(w) = a_w/s$.

**Remark 2.8.** Clearly, any solution to the fractional program for FMBS is a solution to the integral program for MBS. Hence, both being *maximization* problems, FMBS upper bounds MBS. Later, we prove in Lemma 3.2 that the converse of this inequality holds in the sense that $\mathrm{FMBS}(f)$ is upper bounded by a polynomial in $\mathrm{MBS}(f)$.

**Remark 2.9.** Fractional block sensitivity (the non-monotone variant) was considered by Tal in [33]. Tal mentions explicitly the problem of finding separations between fractional block sensitivity and sensitivity.

## 2.2 Hitting Set Complexity

Next, we consider *hitting set complexity*. This can be viewed as a variant of *certificate complexity*, a commonly-studied quantity in standard query complexity.

**Definition 2.10** (Hitting set complexity)**.** The hitting set complexity of a function $f : \{0,1\}^n \to \mathbb{R}$ at an input $z \in \{0,1\}^n$, denoted $\mathrm{HSC}(f,z)$, is the minimal size of a set $H \subseteq [n]$ which intersects all sets in $\mathcal{W}(f,z)$. In other words, for every $w \in \mathcal{W}(f,z)$ there is some $i \in H$ so that $w_i = 1$. We denote $\mathrm{HSC}(f) = \max_z \mathrm{HSC}(f,z)$.

Similarly to MBS, it is simple to see that the $n$-bit AND and OR functions have $\mathrm{HSC}(\mathrm{AND}_n) = 1$ and $\mathrm{HSC}(\mathrm{OR}_n) = n$, respectively.

**Remark 2.11.** Note that $\mathrm{HSC}(f,z)$ is equal to the certificate complexity of $f_z$ at the all 0s input.

**Remark 2.12.** It suffices to consider $H \subseteq [n]$ which have non-empty intersection with any *minimal* element of $\mathcal{W}(f,z)$. This is simply because if $H$ hits an element $w$ then it also hits every superset of $w$.

**Remark 2.13.** Let $H \subseteq [n]$ be a set such that $|H| = b$ and it witnesses $\mathrm{HSC}(f, 0^n) = b$. By the previous remark and Claim 2.1, one can see that $H$ is hitting set of $\mathcal{M}(f)$.

We can also phrase $\mathrm{HSC}(f,z)$ as the value of a certain set covering LP. Putting $\mathcal{W} = \mathcal{W}(f,z)$, the LP optimizes over the variables $\{b_i \;:\; i \in [n]\}$ as follows:

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in [n]} b_i \\ \text{subject to} \quad & \sum_{i \in [n] : w_i = 1} b_i \ge 1 \quad \text{for all } w \in \mathcal{W} \\ & b_i \in \{0,1\} \quad \text{for all } i \in [n] \end{aligned}$$

One can easily verify that this LP is dual to the LP defining monotone block sensitivity. Fractional hitting set complexity is obtained from hitting set complexity by relaxing each constraint $b_i \in \{0,1\}$

to $b_i \in [0,1]$. We give an alternative definition, equivalent to the LP formulation:

**Definition 2.14** (Fractional hitting set complexity)**.** The fractional hitting set complexity of a function $f : \{0,1\}^n \to \mathbb{R}$ at an input $z \in \{0,1\}^n$, denoted $\mathrm{FHSC}(f,z)$, is $1/p$, where $p > 0$ is the smallest number for which there exists a distribution $\mathcal{D}$ of indices $i \in [n]$ with the property that $\Pr_{i\sim\mathcal{D}}[w_i = 1] \geq p$ for each $w \in \mathcal{W}(f,z)$. We denote $\mathrm{FHSC}(f) = \max_z \mathrm{FHSC}(f,z)$.

**Remark 2.15.** The same reasoning as the FMBS case can be used to show that this definition is equivalent to the LP definition. Also by analogous reasoning, $\mathrm{FHSC}(f,z) \leq \mathrm{HSC}(f,z)$ (recalling that FHSC is a minimization problem).

The LPs defining FHSC and FMBS are dual, so linear programming duality yields $\mathrm{FHSC}(f,z) = \mathrm{FMBS}(f,z)$. Combined with the remarked-upon relationships between MBS/FMBS and HSC/FHSC, we conclude the following:

**Claim 2.16.** For any function $f : \{0,1\}^n \to \mathbb{R}$ and input $z \in \{0,1\}^n$,

$$\mathrm{MBS}(f,z) \leq \mathrm{FMBS}(f,z) = \mathrm{FHSC}(f,z) \leq \mathrm{HSC}(f,z).$$

## 2.3 Some Informative Examples

To digest the definitions, some examples are in order. We start by noting that there are large gaps in the inequalities from Claim 2.16 for *fixed* $z$. These correspond to integrality gaps for the set cover and hitting set linear programs (of which $\mathrm{FHSC}(f,z)$ and $\mathrm{FMBS}(f,z)$ are a special case), which are central to combinatorial optimization.

The first example gives a separation between $\mathrm{FMBS}(f,z)$ and $\mathrm{MBS}(f,z)$.

**Example 2.17** (Projective plane)**.** For a prime power $m$, let $P$ be the set of 1-dimensional subspaces of $\mathbb{F}_m^3$ and $L$ the set of 2-dimensional subspaces of $\mathbb{F}_m^3$. $P$ is the set of *points* and $L$ is the set of *lines*. Note that $|P| = |L| = m^2 + m + 1$.

It is well-known that $P$ and $L$ form a *projective plane*, in that they satisfy the following properties:

(1) Any two points in $P$ are contained in exactly one line in $L$. Moreover, each point is contained in $m+1$ lines.
(2) Any two lines in $L$ intersect at exactly one point in $P$. Moreover, each line contains $m+1$ points.
(3) There are 4 points, no 3 of which lie on the same line.

For more background on finite geometry, see, for example, [2].

Let $n = m^2+m+1$, thinking of each $i \in [n]$ as corresponding to a point $p_i \in P$. For lines $\ell \in L$, let $S_\ell = \{i \in [n] : p_i \text{ is contained in } \ell\}$ be the set of (indices of) points incident to $\ell$ and define a function $f : \{0,1\}^n \to \{0,1\}$ as

$$f(z) = \bigvee_{\ell \in L} \Big( \bigwedge_{i \in S_\ell} z_i \Big).$$

Since any two lines intersect at a point, any $\ell_1, \ell_2 \in L$ have $S_{\ell_1} \cap S_{\ell_2} \neq \emptyset$. This implies $\mathrm{MBS}(f, 0^n) = 1$. On the other hand, because each line contains $m+1$ points, $\Pr_{i\in[n]}[i \in S_\ell] = (m+1)/(m^2+m+1)$ when $i$ is uniform and therefore $\mathrm{FMBS}(f, 0^n) \approx m \approx \sqrt{n}$.

The next example gives a similar separation between $\mathrm{FHSC}(f,z)$ and $\mathrm{HSC}(f,z)$.

**Example 2.18** (Majority)**.** For $n$ even, let $f(z) = \mathbf{1}[\sum_i z_i \geq n/2]$ be the Majority function. The minimal elements of $\mathcal{M}(f)$ consist of sets $s$ with $n/2$ members.

Any set $s$ of size at most $n/2$ will fail to hit $[n] \setminus s \in \mathcal{M}(f)$. Therefore any hitting set for the monomials of $f$, namely for $\mathcal{M}(f)$, has size more than $n/2$. In particular, $\mathrm{HSC}(f, 0^n) = n/2 + 1$ (clearly $n/2+1$ suffices). On the other hand, the uniform distribution over $[n]$ satisfies $\Pr[i \in s] = 1/2$ for any minimal monomial $s \in \mathcal{M}(f)$. Hence $\mathrm{FHSC}(f, 0^n) = 2$.

These two examples show that it will be necessary to utilize the fact that MBS and HSC are defined as the maximum over all inputs.

The next examples shows that $\mathrm{HSC}(f)$ and $\mathrm{MBS}(f)$ can be constant while $\mathrm{spar}(f)$ grows exponentially.

**Example 2.19** (AND-OR)**.** Consider a string $z \in \{0,1\}^{2n}$ written as $z = xy$ for $x, y \in \{0,1\}^n$. Define

$$f(x,y) = \bigwedge_{j\in[n]} \left(x_j \vee y_j\right).$$

One can verify that $\mathrm{HSC}(f) = \mathrm{MBS}(f) = 2$. On the other hand, writing $f$ as a multilinear polynomial yields

$$f(x,y) = \prod_{j\in[n]} (x_i + y_j - x_j \cdot y_j),$$

which clearly has sparsity exponential in $n$.

Note that this holds for the *global* (i.e. maximizing over $\{0,1\}^n$) definitions of MBS and HSC. To see the significance of this example, recall from the proof overview that we are interested in eventually showing $\mathrm{MBS}(f) \leq O((\log \mathrm{spar}(f))^2)$. This example shows that this latter inequality can be very far from the truth; we are able to make up for this discrepancy by using the low-sparsity assumption multiple times.

Finally, we include an example which will become relevant to our applications to communication complexity in Section 4.

**Example 2.20** (Redundant indexing)**.** Let $k \geq 1$, and consider two sets of variables $\{x_S\}_{S\subseteq[k]}$ and $\{y_i\}_{i\in[k]}$ of sizes $2^k$ and $k$, respectively. Let $n = 2^k + k$ and define

$$f(x,y) = \sum_{i\in[k]} \left( \prod_{S\,:\,i\in S} x_S \right) (1-y_i) \left( \prod_{j\neq i} y_j \right).$$

In words, $f(x,y) = 1$ when $y$ has weight exactly $k-1$ with $y_i = 0$ and $x_S = 1$ for every $S$ containing $i$.

By the mutilinear representation, one can see that the sparsity of $f$ is $2k \sim \log n$. Moreover, $\mathrm{HSC}(f) \leq 2$. To see why, consider an input $z = (a,b)$ and note that $f$ restricted to inputs $w = (x,y) \geq z$ becomes

$$f'(x,y) = \sum_{i:b_i=0} \left( \prod_{S:i\in S, a_S=0} x_S \right) (1-y_i) \left( \prod_{j:j\neq i, b_j=0} y_j \right).$$

In particular, if $a \neq 1^{[2^k]}$ then the variable $x_{[2^k]\setminus a}$ hits all the monomials, and if $a = 1^{[2^k]}$ then any two $y_i, y_j$ hit all the monomials.

We view this as an important example in understanding the $\log n$ factor currently present in the statements of Theorem 1.2 and Theorem 1.3. This connection will be discussed in more detail in Section 6.

## 3 PROOF OF THEOREM 1.1

We recall the statement of Theorem 1.1.

**Theorem 1.1.** Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function with sparsity $\mathrm{spar}(f) = r$. Then there exists a hitting set $H$ for $\mathcal{M}(f)$ of size $|H| = O((\log r)^5)$.

The proof relies on three lemmas which provide various relationships between $\mathrm{spar}(f)$, $\mathrm{MBS}(f)$ and $\mathrm{HSC}(f)$, as well as their fractional variants. In this subsection, we will state the lemmas and show how Theorem 1.1 follows as a consequence. Then, in the following subsections, we prove the lemmas.

The first gives an upper bound on the monotone block sensitivity of a *boolean-valued* $f$ in terms of its sparsity.

**Lemma 3.1.** For any $f : \{0,1\}^n \to \{0,1\}$,

$$\mathrm{MBS}(f) = O(\log(\mathrm{spar}(f))^2).$$

We stress that this only holds for boolean-valued functions. To some extent, we will be able to relax this condition when we consider generalizations in Section 5. Additionally, we note that this inequality can be very far from tight: Example 2.19 gives a function with constant MBS but exponential sparsity.

The second lemma shows that FMBS and MBS are equivalent up to a polynomial factor. Unlike Lemma 3.1, this holds for any real-valued function.

**Lemma 3.2.** For any $f : \{0,1\}^n \to \mathbb{R}$,

$$\mathrm{FMBS}(f) = O(\mathrm{MBS}(f)^2).$$

The third lemma, which also holds for any real-valued function, upper bounds the hitting set complexity of $f$ in terms of $\mathrm{FMBS}(f)$ and $\mathrm{spar}(f)$.

**Lemma 3.3.** For any $f : \{0,1\}^n \to \mathbb{R}$,

$$\mathrm{HSC}(f) \le \mathrm{FMBS}(f) \cdot \log(\mathrm{spar}(f)).$$

Theorem 1.1 now follows quite readily from the three lemmas.

Proof of Theorem 1.1. Fix a boolean function $f$ with sparsity $r$ as in the theorem statement. By Lemma 3.1, $\mathrm{MBS}(f) = O((\log r)^2)$. By Lemma 3.2, $\mathrm{FMBS}(f) = O((\log r)^4)$. Finally, by Lemma 3.3, $\mathrm{HSC}(f) \le O((\log r)^5)$, as desired. □

### 3.1 MBS from Sparsity

We begin by proving Lemma 3.1.

**Lemma 3.1.** For any $f : \{0,1\}^n \to \{0,1\}$,

$$\mathrm{MBS}(f) = O(\log(\mathrm{spar}(f))^2).$$

The proof uses a well-known relationship between the degree and the sensitivity of boolean functions [26]. The *sensitivity* $S(f)$ of a boolean function $f$ is the largest $s$ so that there exists an input $z$ and $s$ coordinates $\{i_1, \ldots, i_s\}$ so that $f(z) \neq f(z \oplus e_{i_j})$ for all $j \in [s]$.

**Claim 3.4** (Nisan-Szegedy, [26])**.** For any $f : \{0,1\}^n \to \{0,1\}$,

$$S(f) = O(\deg(f)^2).$$

Proof of Lemma 3.1. Suppose $\mathrm{MBS}(f) = k$, witnessed by pairwise disjoint $z, w_1, \ldots, w_k \subseteq [n]$. Namely, $f(z) \neq f(z \vee w_i)$ for $i \in [k]$. Let $g : \{0,1\}^k \to \{0,1\}$ denote the function obtained from $f$ by identifying variables in each $w_i$ and setting all variables not occurring in any $w_i$ to the corresponding bit in $z$. That is, $g(x) = f(z + \sum x_i w_i)$. Note that $S(g) = k$, since $g(0) \neq g(e_i)$ for $i \in [k]$, and $\mathrm{spar}(g) \le \mathrm{spar}(f)$.

Let $r = \mathrm{spar}(f)$. We will reduce the degree of $g$ to $d = O(\log r)$ by repeating the following process $k/2$ times: set to zero the coordinate which appears in the largest number of monomials of degree at least $d$.

Let $M_i$ denote the number of monomials of degree at least $d$ remaining after the $i$-th step. Initially $M_0 \le r$. Next, note that if $M_i > 0$, then there is a variable that occurs in at least a $d/k$ fraction of the monomials of degree $\ge d$. Indeed, the total length of the monomials of degree $\ge d$ is at least $M_i d$; however, under the assumption that each variable occur in less than $d/k$ fraction of these monomials, we can see that the total length is less than $(d/k)M_i d = dM_i$, which is a contradiction. We therefore obtain the recurrence $M_{i+1} \le (1 - d/k)M_i$. After $k/2$ steps, $M_{k/2} \le (1 - d/k)^{k/2} r \le \exp(-d/2)r < 1$ for $d = O(\log r)$. As $M_{k/2}$ is an integer, we obtain that $M_{k/2}$ is zero.

Let $h$ denote the function obtained by this restriction process. Since $M_{k/2} = 0$ we have $\deg(h) < d$. Moreover, since $g$ had full sensitivity at $0^k$ and we restricted only $k/2$ coordinates, $S(h) \ge k/2$. Finishing up, we have $k/2 \le S(h) = O(\deg(h)^2) = O((\log r)^2)$, completing the proof. □

### 3.2 Fractional vs. Integral Solutions for MBS

This subsection proves Lemma 3.2, restated here:

**Lemma 3.2.** For any $f : \{0,1\}^n \to \mathbb{R}$,

$$\mathrm{FMBS}(f) = O(\mathrm{MBS}(f)^2).$$

We first need the following claim, which states that any function $f : \{0,1\}^n \to \mathbb{R}$ is not too sensitive to noise which is $q$-smooth for $q \ll 1/\mathrm{FMBS}(f)$.

**Claim 3.5.** Let $f : \{0,1\}^n \to \mathbb{R}$, $z \in \{0,1\}^n$ and $\mathcal{D}$ a distribution on $\{0,1\}^{[n]\setminus z}$. Assume that $\mathcal{D}$ is $q$-smooth for some $q \in (0,1]$. Then

$$\Pr_{w \sim D}[f(z) \neq f(z \vee w)] \le q \cdot \mathrm{FMBS}(f, z).$$

Proof. Assume $\mathrm{FMBS}(f, z) = 1/p$. We may assume $q < p$ as otherwise the claim is trivial. Let $\delta = \Pr_{w \sim \mathcal{D}}[f(z) \neq f(z \vee w)]$. Let $\mathcal{D}'$ be the distribution $\mathcal{D}$ restricted to inputs $w$ such that $f(z) \neq f(z \vee w)$. Observe that $\mathcal{D}'$ is $(q/\delta)$-smooth, and is supported on inputs $w$ such that $f(z) \neq f(z \vee w)$. As $\mathrm{FMBS}(f, z) = 1/p$ we have $q/\delta \ge p$ which implies the claim. □

Proof of Lemma 3.2. Let $\mathrm{FMBS}(f) = 1/p$. Let $z \in \{0,1\}^n$ such that $\mathrm{FMBS}(f,z) = 1/p$, and let $\mathcal{D}$ be a $p$-smooth distribution supported on $\mathcal{W}(f,z)$.

Fix $k$ to be determined later, and sample inputs $w_1, \ldots, w_k \sim \mathcal{D}$ independently. Let $u$ denote all the elements that appear at least in two of the $w_i$, namely

$$u = \bigvee_{i \neq j} \left( w_i \bigwedge w_j \right).$$

The main observation is that $u$ is $q$-smooth for $q = (pk)^2$. This holds since for every $\ell \in [n]$ we have

$$\Pr[u_\ell = 1] \leq \sum_{i \neq j} \Pr[(w_i)_\ell = 1, (w_j)_\ell = 1] \leq k^2 p^2.$$

Define the following "bad" events:

$$E_0 = [f(z) \neq f(z \vee u)],$$
$$E_t = [f(z \vee w_t) \neq f(z \vee w_t \vee u)] \text{ for } t \in [k].$$

We claim that $\Pr[E_t] \leq q/p = pk^2$ for all $t = 0, \ldots, k$. The proof for $E_0$ follows directly from Claim 3.5. To see why it holds for $E_t$ for $t = 1, \ldots, k$, define $u_t$ to be the elements that appear in two sets $w_i$, excluding $w_t$, namely

$$u_t = \bigvee_{i \neq j,\ i,j \neq t} \left( w_i \bigwedge w_j \right).$$

Observe that $w_t, u_t$ are independent since $u_t$ is a subset of $\bigvee_{i \neq t} w_i$, that $u_t$ is $(pk)^2$-smooth and that $w_t \vee u = w_t \vee u_t$. Thus Claim 3.5 gives that, for any fixing of $w_t$, we have

$$\Pr_{u_t}[f(z \vee w_t) \neq f(z \vee w_t \vee u_t) \mid w_t] \leq$$
$$q \cdot \mathrm{FMBS}(f, z \vee w_t) \leq q \cdot \mathrm{FMBS}(f) = q/p = pk^2.$$

The claim for $E_t$ follows by averaging over $w_t$.

Pick $k = 1/(2\sqrt{p})$, meaning $E_t$ occurs with probability at most $1/4$ for each $0 \leq t \leq k$. Then conditioning on $\neg E_0$ will increase the probability of any event by a factor of at most $1/(1-1/4) = 4/3$. In particular, because $\Pr[E_t] \leq pk^2 = 1/4$ for any $t$, we have $\Pr[E_t | \neg E_0] \leq 1/3$ for any $t \neq 0$. This means that we can sample the $w_t$'s conditioned on $\neg E_0$, and still be sure that every $\neg E_t$ occurs with probability at least $2/3$. Averaging, some setting of the $\{w_t\}$ satisfies $\neg E_0$ and at least $2/3$ of $\neg E_t$ for $1 \leq t \leq k$. Fix these $\{w_t\}$.

Define $z' = z \vee u$ and $w'_t = w_t \setminus u$. For every $1 \leq t \leq k$ for which $\neg E_t$ holds, we have

$$f(z') = f(z), \qquad f(z' \vee w'_t) = f(z \vee w_t).$$

Thus $f(z') \neq f(z' \vee w'_t)$ for at least $2k/3$ choices of $w'_t$. Moreover, $z', w'_1, \ldots, w'_k$ are pairwise disjoint. Hence $\mathrm{MBS}(f) \geq 2k/3$. This completes the proof, by recalling that $k = 1/(2\sqrt{p})$ with $\mathrm{FMBS}(f) = 1/p$. □

A notable feature of this proof is that we need to employ upper bounds on the fractional block sensitivity for more than one choice of input. This is actually necessary; there is a function $f$ based on the projective plane for which $\mathrm{MBS}(f,z) = 1$ and $\mathrm{FMBS}(f,z) \sim \sqrt{n}$ at a point $z$. See Example 2.17 for details.

## 3.3 Hitting Sets from Sparsity

Our final lemma is an upper bound on the hitting set complexity of any $f : \{0,1\}^n \to \mathbb{R}$ in terms of $\mathrm{FMBS}(f)$ and $\log(\mathrm{spar}(f))$. Recall that FMBS and FHSC are equal, so such an upper bound implies that FHSC and HSC are polynomially related for sparse boolean functions.

**Lemma 3.3.** For any $f : \{0,1\}^n \to \mathbb{R}$,

$$\mathrm{HSC}(f) \leq \mathrm{FMBS}(f) \cdot \log(\mathrm{spar}(f)).$$

We note that the lemma is essentially the integrality gap of set cover. Before proving it, we need two straightforward claims which we will use again later on. The first allows us to find (non-uniformly) indices $i \in [n]$ which hit a large fraction of $\mathcal{M}(f)$, given that $f$ has small FMBS/FHSC at $0^n$.

**Claim 3.6.** Suppose $\mathrm{FMBS}(f, 0^n) = \mathrm{FHSC}(f, 0^n) = k$ and this is witnessed by a distribution $\mathcal{D}$ over $[n]$. Then

(1) $\Pr_{i \sim \mathcal{D}}[i \in w] \geq 1/k$ for every $w \in \mathcal{M}(f)$. That is, $\mathcal{D}$ is also a fractional hitting set for the *monomials* of $f$.
(2) There is some $i$ in the support of $\mathcal{D}$ which hits a $1/k$-fraction of $\mathcal{M}(f)$.

Proof. Note that the second part of the claim follows from the first by an averaging argument, so we are contented to prove the first part of the claim.

Let $\mathcal{D}, \mathrm{FHSC}(f, 0^n) = k$ be as stated, so that $\Pr_{i \sim \mathcal{D}}[i \in w] \geq 1/k$ for all $w \in \mathcal{W}(f, 0^n)$. By Claim 2.1, it is the case that $\Pr_{i \sim \mathcal{D}}[i \in w] \geq 1/k$ for any minimal monomial $w$. The measure of $\mathcal{D}$ on some $w$ is non-decreasing with respect to taking supersets, meaning $\Pr_{i \sim \mathcal{D}}[i \in w] \geq 1/k$ for every monomial $w \in \mathcal{M}(f)$. □

The second claim says that $\mathrm{FHSC}(f)$ is non-increasing under restrictions. For simplicity, we only consider reductions which set a single bit (which can be extended to more bits by induction).

**Claim 3.7.** Let $f : \{0,1\}^n \to \mathbb{R}$ be a function, $i \in [n]$ and $b \in \{0,1\}$. Let $f' : \{0,1\}^{[n] \setminus \{i\}} \to \mathbb{R}$ be the function obtained by restricting to inputs with $x_i = b$. Then

$$\mathrm{FHSC}(f') \leq \mathrm{FHSC}(f).$$

Proof. Fix $z \in \{0,1\}^{[n] \setminus \{i\}}$. We will show that $\mathrm{FHSC}(f', z) \leq \mathrm{FHSC}(f, z^*)$ where $z^* = z$ if $b = 0$ and $z^* = z \cup \{i\}$ if $b = 1$. In either case, $\mathrm{FHSC}(f', z) \leq \mathrm{FHSC}(f)$ and hence $\mathrm{FHSC}(f') \leq \mathrm{FHSC}(f)$.

Consider first the case of $b = 0$, and assume that $\mathrm{FHSC}(f, z) = 1/p$. Recall that $f_z$ is the restriction of $f$ to inputs $x \geq z$, and that $\mathrm{FHSC}(f, z) = \mathrm{FHSC}(f_z, 0)$. By definition, there is a distribution $\mathcal{D}$ over $[n]$ such that for every $w \in \mathcal{M}(f_z)$ we have $\Pr_{i \sim \mathcal{D}}[w_i = 1] \geq p$. Observe that $\mathcal{M}(f'_z) \subset \mathcal{M}(f_z)$ since setting a variable to 0 can only remove monomials. Thus we get $\mathrm{FHSC}(f', z) \leq \mathrm{FHSC}(f, z)$.

Next, consider the case of $b = 1$. Note that $f'_z = f_{z \cup \{i\}}$ and hence $\mathrm{FHSC}(f', z) = \mathrm{FHSC}(f, z \cup \{i\})$. □

Proof of Lemma 3.3. Let $k = \mathrm{FHSC}(f, 0) \leq \mathrm{FHSC}(f)$, $S_0 = \emptyset$, $f_0 = f$ and perform the following iterative process. At time $t \geq 1$, let $S_t = S_{t-1} \cup \{i_t\}$ where $i_t \in [n]$ is the index which hits a $1/k$-fraction of $\mathcal{M}(f_{t-1})$, guaranteed to exist by Claim 3.6. Let $f_t = f_{t-1}|_{z_{i_t} = 0}$. At each step, the restriction $z_{i_t} = 0$ sets every monomial containing $i_t$

to zero, causing the sparsity of $f_{t-1}$ to decrease by a multiplicative factor $(1-1/k)$. Let $r_t = |\mathcal{M}(f_t)|$. Since $S_t$ is a hitting set for $\mathcal{M}(f)$ when $f_t$ has no non-zero monomials, this process terminates with a hitting set when

$$r_t = (1-1/k)^t r_0 \le e^{-t/k} r_0 < 1.$$

Therefore, taking $t = k \log r_0$ suffices. □

# 4 COROLLARIES IN COMMUNICATION COMPLEXITY

## 4.1 Preliminaries

Fix a *boolean* function $f : \{0,1\}^n \to \{0,1\}$. Let $f_\wedge = f \circ \wedge$ denote the AND function corresponding to $f$, given by $f_\wedge(x,y) = f(x \wedge y)$. The sparsity of $f$ characterizes the rank of $f_\wedge$.

**Claim 4.1** ([3]). For any $f : \{0,1\}^n \to \{0,1\}$,

$$\mathrm{spar}(f) = \mathrm{rank}(f_\wedge).$$

We assume familiarity with the standard notion of a *decision tree*. Our primary interest is in a variant of decision trees called *AND decision trees*, which strengthens decision trees by allowing queries of the conjunction of an arbitrary subset of the variables, namely queries of the form $\wedge_{i\in S} z_i$ for arbitrary $S \subseteq [n]$. Let $\mathrm{P}^{\wedge\text{-dt}}(f)$ denote the smallest depth of an AND decision tree computing $f$. The following simple connection to the communication complexity of $f_\wedge$ motivates our interest in this model:

**Claim 4.2.** For any $f : \{0,1\}^n \to \{0,1\}$,

$$\mathrm{P}^{\mathrm{cc}}(f_\wedge) \le 2\mathrm{P}^{\wedge\text{-dt}}(f).$$

Proof. Whenever the AND-decision tree queries a set $S \subseteq [n]$, Alice and Bob privately evaluate $a = \wedge_{i\in S} x_i$ and $b = \wedge_{j\in S} y_j$, exchange them and continue the evaluation on the sub-tree obtained by following the edge labelled $a \wedge b$. If the decision tree height is $d$, this protocol uses $2d$ bits of communication. Correctness follows from the observation that $\bigwedge_{i\in S}(x_i \wedge y_i) = (\bigwedge_{i\in S} x_i) \wedge (\bigwedge_{j\in S} y_j)$. □

There is also a simple connection between AND-decision trees and sparsity:

**Claim 4.3.** Let $f : \{0,1\}^n \to \{0,1\}$ with $d = \mathrm{P}^{\wedge\text{-dt}}(f)$. Then $\mathrm{spar}(f) \le 3^d$.

Proof. Assume that $f$ is computed by a depth-$d$ AND decision tree, where the first query is $\wedge_{i\in S} z_i$, and where $f_1, f_2$ are the functions computed by the left and right subtrees, respectively. Note that both are computed by AND decision trees of depth $d-1$. We have

$$f(z) = \prod_{i\in S} z_i \cdot f_1(z) + \left(1 - \prod_{i\in S} z_i\right) f_2(z).$$

Thus

$$\mathrm{spar}(f) \le \mathrm{spar}(f_1) + 2 \cdot \mathrm{spar}(f_2).$$

The claim follows, since in the base case, functions computed by an AND-decision tree of depth 1 has sparsity at most 2. □

A related complexity measure introduced in [25], called the 0-decision tree complexity of $f$, is defined as follows. The 0-*depth* of a (standard) decision tree $\mathcal{T}$ is largest number of 0-edges encountered on a root-to-leaf path in $\mathcal{T}$. The 0-decision tree complexity of $f$, denoted $\mathrm{P}^{0\text{-dt}}(f)$, is the smallest 0-depth over all trees $\mathcal{T}$ computing $f$. The following relationship between AND decision trees and 0-decision tree complexity is from [25]:

**Claim 4.4** ([25]). For any $f : \{0,1\}^n \to \{0,1\}$,

$$\mathrm{P}^{0\text{-dt}}(f) \le \mathrm{P}^{\wedge\text{-dt}}(f) \le \mathrm{P}^{0\text{-dt}}(f)\,\lceil \log(n+1) \rceil.$$

For completeness, we include the short proof.

Proof. The first inequality follows since an AND query can be simulated by querying the bits in it one at a time, until the first 0 is queried, or until they are all queried to be 1. In particular, at most a single 0 query is made. This implies that an AND decision tree of depth $d$ can be simulated by a standard decision tree of 0-depth $d$.

For the second inequality, let $\mathcal{T}$ be a decision tree computing $f$ with 0-depth $d$. Consider the subtree which is truncated after the first 0 is read. We can compute which leaf in the subtree is reached by doing a binary search on the at most $n+1$ options, which can be implemented using $\lceil \log(n+1) \rceil$ computations of ANDs. Then, the same process continues on the tree rooted at the node reached, which has 0-depth at most $d-1$. □

The following example shows that this gap of $\log n$ cannot be avoided.

**Example 4.5.** For $z \in \{0,1\}^n$, let $\mathrm{ind}(z) \in [n]$ denote the first index $i$ for which $z_i = 0$. Let

$$f(z) = \begin{cases} 1 & \text{if } z = 1^n \text{ or } z = 1^{n-1}0 \\ z_{\mathrm{ind}(z)+1} & \text{otherwise} \end{cases}$$

A simple decision tree for $f$ that queries bits of $z$ one after another, will query at most two zeroes, corresponding to $z_{\mathrm{ind}(z)}$ and $x_{\mathrm{ind}(z)+1}$, and hence $\mathrm{P}^{0\text{-dt}}(f) \le 2$. However, a direct calculation shows that $\mathrm{spar}(f) = \Omega(n)$ and therefore, by Claim 4.3, $\mathrm{P}^{\wedge\text{-dt}}(f) = \Omega(\log n)$.

We also use a lemma closely related to Lemma 3.3.

**Lemma 4.6.** For any $f : \{0,1\}^n \to \{0,1\}$,

$$\mathrm{P}^{0\text{-dt}}(f) = O(\mathrm{FMBS}(f) \cdot \log \mathrm{spar}(f)).$$

Proof. Let $k = \mathrm{FHSC}(f,0) \le \mathrm{FHSC}(f)$. By Claim 3.6, there is an $i \in [n]$ that belongs to at least a $(1/k)$-fraction of $\mathcal{M}(f)$. Query the variable $x_i$ and let $b_i \in \{0,1\}$ be the outcome. Let $f' : \{0,1\}^n \to \{0,1\}$ be the function $f$ restricted to $x_i = b_i$. Consider the sparsity of $f'$:

- If $x_i = 0$ then $|\mathcal{M}(f')| \le (1-1/k)|\mathcal{M}(f)|$, as setting $x_i = 0$ kills a $(1/k)$-fraction of the non-constant monomials. Thus, as long as $f$ is not a constant function, $|\mathcal{M}(f)| \ge 1$ and we have

  $$\mathrm{spar}(f') \le \mathrm{spar}(f) - |\mathcal{M}(f)|/k \le \mathrm{spar}(f)(1-1/2k).$$

- If $x_i = 1$ then $\mathrm{spar}(f') \le \mathrm{spar}(f)$, since fixing variables to constants cannot increase the number monomials.

Let $t$ the maximum number of 0's queried along any path in the obtained decision tree. The sparsity of the subfunction $f'$ corresponding to a leaf must be 0 or else $f'$ is non-constant. By, Claim 3.7 $f'$ is constant when $(1-1/2k)^t \mathrm{spar}(f) \le e^{-t/2k}\mathrm{spar}(f) < 1$, which occurs when $t \ge 2k \cdot \log \mathrm{spar}(f)$. □

## 4.2 The Log-Rank Conjecture

A weak version of the log-rank conjecture for AND-functions, which includes an additional $\log n$ factor, now follows quite readily from the tools we have developed.

**Theorem 1.2** (Log-rank Theorem for AND-functions)**.** Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. Let $r = \mathrm{spar}(f) = \mathrm{rank}(f_\wedge)$. Then $f$ can be computed by an AND-decision tree of depth

$$\mathrm{P}^{\wedge\text{-dt}}(f) = O((\log r)^5 \cdot \log n).$$

In particular, the deterministic communication complexity of $f_\wedge$ is bounded by

$$\mathrm{P}^{\mathrm{cc}}(f_\wedge) = O((\log r)^5 \cdot \log n).$$

Proof. By Lemma 3.1, $\mathrm{MBS}(f) = O((\log r)^2)$. By Lemma 3.2, $\mathrm{FMBS}(f) = O((\log r)^4)$. By Lemma 4.6, $\mathrm{P}^{0\text{-dt}}(f) = O((\log r)^5)$. By Claim 4.4 this gives us an AND-decision tree of height $O((\log r)^5 \cdot \log n)$. Finally, we convert the AND-decision tree for $f$ into a protocol for $f_\wedge$ using Claim 4.2 with complexity $O((\log r)^5 \cdot \log n)$. □

In particular, the log-rank conjecture for AND-functions is true for any $f$ with $\mathrm{spar}(f) \ge n^c$ for any constant $c > 0$. In some sense this is an extremely mild condition, which random $f$ will satisfy with exceedingly high probability. On the other hand, the log-rank conjecture is about *structured* functions; rank and communication complexity are both maximal for random functions, whereas we are interested in low-complexity functions/low-rank matrices. It could very well be the case that the ultra-sparse regime of $\mathrm{spar}(f) = n^{o(1)}$ is precisely where the log-rank conjecture *fails*. We therefore see removing the $\log n$ factor as an essential problem left open by this work. See Section 6 for additional discussion.

## 4.3 Lifting AND-Functions

Since $\log(\mathrm{spar}(f))$ lower bounds the deterministic communication of $f_\wedge$, the log-rank result from the previous section immediately implies a new upper bound on the AND decision tree complexity of $f$. We can prove a better upper bound by making use of our stronger assumption: instead of only assuming $\log(\mathrm{spar}(f))$ is small, we assume that $\mathrm{P}^{\mathrm{cc}}(f_\wedge)$ is small.

If $f$ has large monotone block sensitivity, then its AND-function embeds unique disjointness as a sub-function. The unique disjointness function on $k$ bits, denoted $\mathrm{UDISJ}_k$, takes two inputs $a, b \in \{0,1\}^k$, and is defined as the partial function:

$$\mathrm{UDISJ}_k(a,b) = \begin{cases} 0 & \text{if } |a \wedge b| = 1 \\ 1 & \text{if } |a \wedge b| = 0 \\ \text{undefined} & \text{otherwise} \end{cases},$$

where $|\cdot|$ is the Hamming weight.

**Claim 4.7** (c.f. [36])**.** Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function with $\mathrm{MBS}(f) = k$. Then $f_\wedge$ contains as a sub-matrix $\mathrm{UDISJ}_k$. That is, there are maps $\mathbf{x}, \mathbf{y} : \{0,1\}^k \to \{0,1\}^n$ and $c \in \{0,1\}$ such that the following holds. For any $a, b \in \{0,1\}^k$ which satisfy that $|a \wedge b| \in \{0,1\}$, it holds that

$$\mathrm{UDISJ}_k(a,b) = f_\wedge(\mathbf{x}(a), \mathbf{y}(b)) \oplus c.$$

Proof. Let $z, w_1, \ldots, w_k \in \{0,1\}^n$ be pairwise disjoint such that $f(z) \ne f(z \vee w_i)$ for all $i \in [k]$. We may assume without loss of generality that $f(z) = 1$, otherwise replace $f$ with its negation, and set $c = 1$.

Assume that Alice and Bob want to solve unique-disjointness on inputs $a, b \in \{0,1\}^k$, which we identify with subsets of $[k]$. Define

$$\mathbf{x}(a) = z \vee \bigvee_{i \in a} w_i, \qquad \mathbf{y}(b) = z \vee \bigvee_{j \in b} w_j.$$

Observe that

$$\mathbf{x}(a) \wedge \mathbf{y}(b) = \begin{cases} z & \text{if } a \wedge b = \emptyset \\ z \vee w_i & \text{if } a \wedge b = \{i\}. \end{cases}$$

Thus we get that $\mathrm{UDISJ}_k(a,b) = f(\mathbf{x}(a) \wedge \mathbf{y}(b))$ for all $a, b$. □

It is well known that $\mathrm{UDISJ}_k$ is hard with respect to several communication complexity measures such as deterministic, randomized and nondeterministic.

**Theorem 4.8** ([16, 29])**.** For any communication complexity measure $\Delta \in \{\mathrm{P}^{\mathrm{cc}}, \mathrm{BPP}^{\mathrm{cc}}, \mathrm{NP}^{\mathrm{cc}}\}$,

$$\Delta(\mathrm{UDISJ}_k) = \Omega(k).$$

We immediately get the following corollary:

**Corollary 4.9.** Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function and $\Delta \in \{\mathrm{P}^{\mathrm{cc}}, \mathrm{BPP}^{\mathrm{cc}}, \mathrm{NP}^{\mathrm{cc}}\}$ be a communication complexity measure. Then $\mathrm{MBS}(f) = O(\Delta(f_\wedge))$.

Proof. Assume that $\mathrm{MBS}(f) = k$. Claim 4.7 shows that any protocol for $f_\wedge$ also solves $\mathrm{UDISJ}_k$. Hence by Theorem 4.8 we have $k = O(\Delta(f_\wedge))$. □

Taking $\Delta = \mathrm{P}^{\mathrm{cc}}$, we obtain the main theorem of this section:

**Theorem 1.3** (Lifting Theorem for AND-functions)**.** Let $f$ be a boolean function from $\{0,1\}^n$ to $\{0,1\}$, and let $C = \mathrm{P}^{\mathrm{cc}}(f_\wedge)$ denote the deterministic communication complexity of its corresponding AND-function. Then $f$ can be computed by an AND-decision tree of depth

$$\mathrm{P}^{\wedge\text{-dt}}(f) = O(C^3 \cdot \log n).$$

Proof. Claim 4.1 gives that $\log \mathrm{spar}(f) = \log \mathrm{rank}(f_\wedge) \le C$. By Claim 4.7, $\mathrm{MBS}(f) = O(C)$. By Lemma 3.2, $\mathrm{FMBS}(f) = O(C^2)$. Combining this with the fact that $\log \mathrm{spar}(f) \le C$, we see, by Lemma 4.6, that $\mathrm{P}^{0\text{-dt}}(f) = O(C^3)$. Finally, by Claim 4.4, we get that $\mathrm{P}^{\wedge\text{-dt}}(f) = O(C^3 \cdot \log n)$. □

# 5 GENERALIZATIONS TO NON-BOOLEAN FUNCTIONS

In this section, we extend our conclusion to general multilinear polynomials and set systems. The main observation is that all measures introduced in Section 2 are defined for general real-valued

functions. In addition, both Lemma 3.2 and Lemma 3.3 are established for real-valued functions. The following theorem holds true as the joint result of these two lemmas.

**Theorem 1.4.** Let $f : \{0,1\}^n \to \mathbb{R}$ be a multilinear polynomial with sparsity $r$. Suppose $\mathrm{MBS}(f) = m$. Then the hitting set complexity of $f$ is bounded by

$$\mathrm{HSC}(f) = O(m^2 \log r).$$

Proof. By Lemma 3.2, $\mathrm{FHSC}(f) = O(m^2)$. Then by Lemma 3.3, we obtain the claimed bound. □

## 5.1 Finite-Range Functions

Lemma 3.1 is not true for general multilinear polynomials. Nevertheless, if we make the assumption that the multilinear polynomial's range is finite, denote its size by $s$, then we can bound the monotone block sensitivity by a polynomial of log-sparsity and $s$.

**Lemma 5.1.** Let $f : \{0,1\}^n \to S$ be a multilinear polynomial where $\mathrm{spar}(f) = r$ and $|S| = s$. Then $\mathrm{MBS}(f) = O(s^2 \log^2 r)$.

Proof. Suppose $\mathrm{MBS}(f) = \mathrm{MBS}(f, z) = k$ for $z \in \{0,1\}^n$, and let $a = f(z) \in S$. Define a polynomial $p : \mathbb{R} \to \{0,1\}$ such that $p(a) = 1$ and $p(b) = 0$ for $b \in S \setminus \{a\}$. There exist such a polynomial of degree $\deg(p) = |S| - 1$. Define a boolean function $g : \{0,1\}^n \to \{0,1\}$ by $g(z) = p(f(z))$. Note that $\mathrm{MBS}(g, z) = k$ and $\mathrm{spar}(g) \le r^{s-1}$. Then by Lemma 3.1, we have

$$k = O(\log^2(\mathrm{spar}(g))) = O(s^2 \log^2 r).$$

□

Combining it with Theorem 1.4, one can bound the hitting set complexity of finite-range functions.

**Theorem 1.5.** Let $f : \{0,1\}^n \to S$ for $S \subset \mathbb{R}$. Assume that $\mathrm{spar}(f) = r$ and $|S| = s$. Then the hitting set complexity of $f$ is bounded by

$$\mathrm{HSC}(f) = O(s^4 (\log r)^5).$$

The following example shows that a polynomial dependency on the range size is necessary in Theorem 1.5.

**Example 5.2.** Let $f(x) = x_1 + \cdots + x_s$. Then $\mathrm{spar}(f) = s$, the range of $f$ has size $s + 1$, and $\mathrm{HSC}(f) = s$.

# 6 DISCUSSION

## 6.1 Ruling out the $\log n$ Factor

Both results about communication complexities of AND-functions (Theorems 1.2 and 1.3) are not "tight" in the sense that both of them have a $\log n$ factor in the right side of the inequality. Unfortunately, $n$ can be exponential in sparsity (see Example 2.20).

It is easy to see that if the $\log n$ factor is truly necessary in these theorems we are very close to refuting the log-rank conjecture. Hence, we believe that a "tighter" version of the log-rank theorem (Theorem 1.2) is true.

**Conjecture 6.1.** Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function, where $\mathrm{spar}(f) = r$. Then

$$\mathrm{P}^{\wedge\text{-dt}}(f) \le \mathrm{poly}(\log r).$$

Note that this conjecture would imply a "tighter" version of the lifting theorem as well.

## 6.2 Randomized Complexity

The main results of this paper are concerned with the deterministic communication complexity of AND-functions. However, Corollary 4.9 says that the randomized communication complexity of an AND-function is lower bounded by its monotone block sensitivity. The relation between randomized communication complexity and sparsity remains unclear. We conjecture that the relation between these two measures is the same as the proved relation (Theorem 1.3) between sparsity and *deterministic* communication complexity.

**Conjecture 6.2.** Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function. Suppose that $\mathrm{BPP}^{\mathrm{cc}}(f_\wedge) = C$. Then

$$\log(\mathrm{spar}(f)) \le \mathrm{poly}(C) \cdot \log n.$$

In particular, $f$ can be computed by an AND-decision tree of depth

$$\mathrm{P}^{\mathrm{cc}}(f_\wedge) \le \mathrm{poly}(C) \cdot \log n.$$

Observe that Conjecture 6.2 implies that randomness does not significantly help to compute AND-functions. Concretely, it implies that

$$\mathrm{P}^{\mathrm{cc}}(f_\wedge) \le \mathrm{poly}\left(\mathrm{BPP}^{\mathrm{cc}}(f_\wedge)\right) \cdot \log n.$$

Interestingly, the $\log n$ factor in this conjecture is necessary as shown by the following example.

**Example 6.3** (Threshold Functions)**.** Let $f : \{0,1\}^n \to \{0,1\}$ be the threshold function such that

$$f(x) = 1 \iff |x| \ge n - 1.$$

It is clear that $\mathrm{spar}(f) = n + 1$; however, $\mathrm{BPP}^{\mathrm{cc}}(f) = O(1)$. Indeed, let us consider the following randomized AND-decision tree for $f$: it samples a subset $S \subseteq [n]$ uniformly at random, then output the value of

$$q_S(x) = \left(\bigwedge_{i \in S} x_i\right) \vee \left(\bigwedge_{i \notin S} x_i\right).$$

Note that if $|x| \ge n - 1$ then $q_S(x) = 1$ with probability 1. If $|x| \le n - 2$, let $i, j$ be such that $x_i = x_j = 0$. With probability $1/2$ we have $i \in S$, $j \notin S$ or $i \notin S$, $j \in S$, in both cases $q_S(x) = 0$. In order to reduce the error, repeat this for a few random sets $S$.

## 6.3 Sparsity vs. Coefficients Size

Let $f : \{0,1\}^n \to \{0,1\}$ and consider the multi-linear polynomial computing $f$, namely $f(x) = \sum_s f_s \prod_{i \in s} x_i$. It is well known that the coefficients $f_s$ take integer values. In particular, if we denote by $\|f\|_1 = \sum |f_s|$ the $L_1$ norm of the coefficients, then we get the obvious inequality

$$\mathrm{spar}(f) \le \|f\|_1.$$

We note the following corollary of Theorem 1.2, which shows that $\|f\|_1$ cannot be much larger than $\mathrm{spar}(f)$.

**Claim 6.4.** Let $f : \{0,1\}^n \to \{0,1\}$ and assume that $\mathrm{spar}(f) = r$. Then $\|f\|_1 \le n^{O(\log r)^5}$.

Proof. By Theorem 1.2, $\mathrm{P}^{\wedge\text{-dt}}(f) = d$ for $d = O((\log r)^5 \log n)$. By a similar proof to Claim 4.3, any function $f$ computed by an AND-decision tree of depth $d$ has $\|f\|_1 \le 3^d$. The claim follows. □

We conjecture that the gap between sparsity and $L_1$ is at most polynomial.

**Conjecture 6.5.** For any $f : \{0,1\}^n \to \{0,1\}$,

$$\|f\|_1 \le \mathrm{poly}(\mathrm{spar}(f)).$$

## ACKNOWLEDGMENTS

The second author was supported by NSF award 2006443. The third author was supported by NSF award 1909634 and a Simons Investigator award.

## REFERENCES


[1] Aaronson, S. Quantum certificate complexity. *J. Comput. Syst. Sci. 74*, 3 (2008), 313–322.

[2] Ball, S., and Weiner, Z. An introduction to finite geometry, 2011. lecture notes, https://web.mat.upc.edu/simeon.michael.ball/IFG.pdf.

[3] Buhrman, H., and de Wolf, R. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity, Chicago, Illinois, USA, June 18-21, 2001* (2001), IEEE Computer Society, pp. 120–130.

[4] Chakrabarti, A., and Regev, O. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM J. Comput. 41*, 5 (2012), 1299–1317.

[5] Chattopadhyay, A., Filmus, Y., Koroth, S., Meir, O., and Pitassi, T. Query-to-communication lifting for BPP using inner product. In *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece* (2019), C. Baier, I. Chatzigiannakis, P. Flocchini, and S. Leonardi, Eds., vol. 132 of *LIPIcs*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, pp. 35:1–35:15.

[6] Chattopadhyay, A., Filmus, Y., Koroth, S., Meir, O., and Pitassi, T. Query-to-communication lifting using low-discrepancy gadgets. *Electronic Colloquium on Computational Complexity (ECCC) 26* (2019), 103.

[7] Chattopadhyay, A., Mande, N. S., and Sherif, S. The log-approximate-rank conjecture is false. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019* (2019), M. Charikar and E. Cohen, Eds., ACM, pp. 42–53.

[8] Chiarelli, J., Hatami, P., and Saks, M. E. An asymptotically tight bound on the number of relevant variables in a bounded degree boolean function. *Comb. 40*, 2 (2020), 237–244.

[9] de Rezende, S. F., Meir, O., Nordström, J., Pitassi, T., Robere, R., and Vinyals, M. Lifting with simple gadgets and applications to circuit and proof complexity. *CoRR abs/2001.02144* (2020).

[10] Frankl, P. On the trace of finite sets. *J. Comb. Theory, Ser. A 34*, 1 (1983), 41–45.

[11] Göös, M., Kamath, P., Pitassi, T., and Watson, T. Query-to-communication lifting for $P^{NP}$. *Comput. Complex. 28*, 1 (2019), 113–144.

[12] Göös, M., Koroth, S., Mertz, I., and Pitassi, T. Automating cutting planes is NP-hard. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020* (2020), K. Makarychev, Y. Makarychev, M. Tulsiani, G. Kamath, and J. Chuzhoy, Eds., ACM, pp. 68–77.

[13] Göös, M., and Pitassi, T. Communication lower bounds via critical block sensitivity. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing* (New York, NY, USA, 2014), STOC '14, Association for Computing Machinery, p. 847–856.

[14] Göös, M., Pitassi, T., and Watson, T. Query-to-communication lifting for BPP. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017* (2017), C. Umans, Ed., IEEE Computer Society, pp. 132–143.

[15] Göös, M., Pitassi, T., and Watson, T. Deterministic communication vs. partition number. *SIAM J. Comput. 47*, 6 (2018), 2435–2450.

[16] Göös, M., Pitassi, T., and Watson, T. The landscape of communication complexity classes. *Comput. Complex. 27*, 2 (2018), 245–304.

[17] Hatami, H., Hosseini, K., and Lovett, S. Structure of protocols for XOR functions. *SIAM J. Comput. 47*, 1 (2018), 208–217.

[18] Hatami, H., and Qian, Y. The unbounded-error communication complexity of symmetric XOR functions. *arXiv preprint arXiv:1704.00777* (2017).

[19] Knill, E. Graph generated union-closed families of sets, 1994.

[20] Leung, M. L., Li, Y., and Zhang, S. Tight bounds on communication complexity of symmetric XOR functions in one-way and SMP models. In *International Conference on Theory and Applications of Models of Computation* (2011), Springer, pp. 403–408.

[21] Lovász, L., and Saks, M. E. Lattices, möbius functions and communication complexity. In *29th Annual Symposium on Foundations of Computer Science, White Plains, New York, USA, 24-26 October 1988* (1988), IEEE Computer Society, pp. 81–90.

[22] Lovett, S. Communication is bounded by root of rank. *J. ACM 63*, 1 (2016), 1:1–1:9.

[23] Mande, N. S., and Sanyal, S. On parity decision trees for fourier-sparse boolean functions. *Electronic Colloquium on Computational Complexity (ECCC) 27* (2020), 119.

[24] Montanaro, A., and Osborne, T. On the communication complexity of XOR functions. *arXiv preprint arXiv:0909.3392* (2009).

[25] Mukhopadhyay, S., and Loff, B. Lifting theorems for equality. In *STACS 2019* (2019).

[26] Nisan, N., and Szegedy, M. On the degree of boolean functions as real polynomials. *Computational complexity 4*, 4 (1994), 301–313.

[27] Pitassi, T., Shirley, M., and Watson, T. Nondeterministic and randomized boolean hierarchies in communication complexity. In *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)* (2020), A. Czumaj, A. Dawar, and E. Merelli, Eds., vol. 168 of *LIPIcs*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, pp. 92:1–92:19.

[28] Raz, R., and McKenzie, P. Separation of the monotone NC hierarchy. *Combinatorica 19*, 3 (1999), 403–435.

[29] Razborov, A. A. On the distributional complexity of disjointness. *Theoretical Computer Science 106*, 2 (1992), 385–390.

[30] Sanyal, S. Near-optimal upper bound on fourier dimension of boolean functions in terms of fourier sparsity. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I* (2015), M. M. Halldórsson, K. Iwama, N. Kobayashi, and B. Speckmann, Eds., vol. 9134 of *Lecture Notes in Computer Science*, Springer, pp. 1035–1045.

[31] Sherstov, A. A. On quantum-classical equivalence for composed communication problems. *Quantum Information & Computation 10*, 5&6 (2010), 435–455.

[32] Tal, A. Properties and applications of boolean function composition. In *Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, USA, January 9-12, 2013* (2013), R. D. Kleinberg, Ed., ACM, pp. 441–454.

[33] Tal, A. Properties and applications of boolean function composition. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science* (2013), pp. 441–454.

[34] Tsang, H. Y., Wong, C. H., Xie, N., and Zhang, S. Fourier sparsity, spectral norm, and the log-rank conjecture. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* (2013), IEEE, pp. 658–667.

[35] Yao, P. Parity decision tree complexity and 4-party communication complexity of XOR-functions are polynomially equivalent. *arXiv preprint arXiv:1506.02936* (2015).

[36] Zhang, S. On the tightness of the buhrman-cleve-wigderson simulation. In *Algorithms and Computation, 20th International Symposium, ISAAC 2009, Honolulu, Hawaii, USA, December 16-18, 2009. Proceedings* (2009), Y. Dong, D. Du, and O. H. Ibarra, Eds., vol. 5878 of *Lecture Notes in Computer Science*, Springer, pp. 434–440.

[37] Zhang, S. Efficient quantum protocols for XOR functions. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms* (2014), SIAM, pp. 1878–1885.

[38] Zhang, Z., and Shi, Y. Communication complexities of symmetric XOR functions. *Quantum Information & Computation 9*, 3 (2009), 255–263.