Improved Bounds on the Sign-Rank of AC^0*

Mark Bun[1] and Justin Thaler[2]

1 **Harvard University, Cambridge, MA, USA**
mbun@seas.harvard.edu

2 **Yahoo Labs, New York, NY, USA**
jthaler@fas.harvard.edu

Abstract

The *sign-rank* of a matrix A with entries in $\{-1,+1\}$ is the least rank of a real matrix B with $A_{ij} \cdot B_{ij} > 0$ for all i, j. Razborov and Sherstov (2008) gave the first exponential lower bounds on the sign-rank of a function in AC^0, answering an old question of Babai, Frankl, and Simon (1986). Specifically, they exhibited a matrix $A = [F(x,y)]_{x,y}$ for a specific function $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ in AC^0, such that A has sign-rank $\exp(\Omega(n^{1/3}))$.

We prove a generalization of Razborov and Sherstov's result, yielding exponential sign-rank lower bounds for a non-trivial class of functions (that includes the function used by Razborov and Sherstov). As a corollary of our general result, we improve Razborov and Sherstov's lower bound on the sign-rank of AC^0 from $\exp(\Omega(n^{1/3}))$ to $\exp(\tilde{\Omega}(n^{2/5}))$. We also describe several applications to communication complexity, learning theory, and circuit complexity.

1998 ACM Subject Classification F.2.0 Analysis of Algorithms and Problem Complexity – General

Keywords and phrases Sign-rank, circuit complexity, communication complexity, constant-depth circuits

Digital Object Identifier 10.4230/LIPIcs.ICALP.2016.37

1 Introduction

The *sign-rank* of a matrix A with entries in $\{-1,+1\}$ is the least rank of a real matrix B with $A_{ij} \cdot B_{ij} > 0$ for all i, j. This fundamental matrix-theoretic complexity measure has diverse applications in theoretical computer science. For example:

- Upper bounds on sign-rank underly many state of the art learning algorithms, including the fastest known algorithms for PAC learning DNF and read-once formulas. Algorithms based on sign-rank are additionally robust to random classification noise, a property not satisfied by the handful of known PAC learning algorithms that cannot be captured in the sign-rank framework (all of which are based on Gaussian Elimination) [12].
- In communication complexity, sign-rank is known to characterize *unbounded error communication*. Introduced by Paturi and Simon [16] and captured by the communication complexity class $\mathbf{UPP}^{cc}$, this is a powerful communication model that lies at the frontier of our understanding. It is essentially the most powerful communication model against which we know how to prove lower bounds. In fact, the only known communication models that $\mathbf{UPP}^{cc}$ cannot efficiently simulate are the communication analogues of the polynomial hierarchy introduced by Babai, Frankl, and Simon [4]. We direct the interested reader to the recent paper of Göös et al. [11] for a detailed overview of communication complexity classes and their relationships.

* A full version of the paper is available at http://eccc.hpi-web.de/report/2016/075/.

43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016).
Editors: Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi;
Article No. 37; pp. 37:1–37:14
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

- In circuit complexity, sign-rank lower bounds on a matrix $A = [F(x,y)]_{x,y}$ imply lower bounds on the size of threshold-of-majority circuits computing F.

Despite the importance of these applications, our understanding of sign-rank remains rather limited, and it is possible to summarize relevant prior work in a single paragraph. Alon et al. [2] proved lower bounds on the sign-rank of random matrices. The first nontrivial lower bounds for explicit matrix families was obtained in a breakthrough work of Forster [8], who proved strong lower bounds on the sign-rank of Hadamard matrices, and more generally of any sign matrix with small spectral norm. Several subsequent works improved and generalized Forster's method [9, 10, 14, 3]. Nearly tight estimates of the sign-rank were obtained by Sherstov in [20] for all symmetric predicates, i.e., matrices of the form $[D(\sum_i x_i \vee y_i)]_{x,y}$ where $D : \{0,1,\ldots,n\} \to \{0,1\}$ is a given predicate and x, y range over $\{0,1\}^n$. Razborov and Sherstov [17] answered an old question of Babai, Frankl, and Simon [4] by giving the first exponential sign-rank lower bounds on a function in AC^0. Specifically, they gave a matrix $A = [F(x,y)]_{x,y}$ for a function $F\colon \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ in AC^0, such that A has sign-rank $\exp(\Omega(n^{1/3}))$.

Our work strengthens and generalizes the results of Razborov and Sherstov on the sign-rank of AC^0.

1.1 Our Results

The *threshold degree* of a function $h\colon \{-1,1\}^n \to \{-1,1\}$, denoted $\deg_{\pm}(h)$, is the least degree of a real polynomial that agrees in sign with h at all inputs. Minsky and Papert [15] famously showed that the threshold degree of the DNF formula $\mathrm{MP}_n(x) = \vee_{i=1}^{n^{1/3}} \wedge_{j=1}^{n^{2/3}} x_{ij}$ – now known as the Minsky-Papert DNF – is $\Omega(n^{1/3})$. This is the same function that Razborov and Sherstov used to prove their sign-rank lower bounds, and their analysis is highly tailored to the Minsky-Papert DNF. We generalize their result as follows.

For any $d > 0$, we identify a class $\mathcal{C}_d$ of functions $f : \{-1,1\}^k \to \{-1,1\}$ such that any $f \in \mathcal{C}_d$ can be transformed into a function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$, where $n = O(dk)$, for which $A = [F(x,y)]_{x,y}$ has sign rank $\exp(\Omega(d))$. Crucially, this transformation is simple in the sense that if f is computed by a polynomial-size circuit of depth t, then F is computed by a polynomial-size circuit of depth at most $t+1$ (and in some cases, F may be shallower).

In particular, the k-variate AND_k function is in $\mathcal{C}_d$ for some $d = \Omega(k^{1/2})$. Our transformation of AND_k into a function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ for $n = O(k^{3/2})$ recovers Razborov and Sherstov's function, with the same sign-rank bound of $\exp(\Omega(k^{1/2})) = \exp(\Omega(n^{1/3}))$. We also identify a k-variate AC^0 function that is in $\mathcal{C}_d$ for some $d = \tilde{\Omega}(k^{2/3})$, which in turn yields new sign-rank lower bounds for AC^0.

The precise definition of $\mathcal{C}_d$ is rather technical, so for expository purposes, we restrict ourselves to an informal statement of this result in this introduction. We define $\mathcal{C}_d$ formally in Section 2.2.

Informal description of the class $\mathcal{C}_d$. Our class $\mathcal{C}_d$ consists of all functions of the form $f : \{-1,1\}^k \to \{-1,1\}$, where f satisfies the following (informally stated) property: there exists a "small" set $S \subseteq f^{-1}(+1)$ such that f cannot be uniformly approximated to error $1/2$ by degree d polynomials, even under the promise that the input x is in $f^{-1}(-1) \cup S$. The precise definition of $\mathcal{C}_d$ is based on a *dual* (in the sense of linear programming duality) interpretation of this property.

Transforming functions in $\mathcal{C}_d$ to functions with high sign-rank. For $g : \{-1,1\}^m \to \{-1,1\}$ and $f : \{-1,1\}^k \to \{-1,1\}$, the notation $g \circ f = g(f, \ldots, f)$ denotes the function on $n = mk$ bits obtained by block-composing g with f. Let OR_m and AND_m denote the logical OR and AND functions on m bits, respectively. Let C be a sufficiently large universal constant. Given a function $f : \{-1,1\}^k \to \{-1,1\}$, let $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ be defined by

$$F = \mathrm{OR}_{2d} \circ f \circ \mathrm{AND}_C \circ \mathrm{OR}_2,$$

and hence $n = 2Cdk = O(dk)$.

▶ **Theorem 1** (Informal). *For any $f \in \mathcal{C}_d$, the matrix $A = [F(x,y)]_{x,y}$ has sign-rank $\exp(\Omega(d))$.*

Examples of functions in $\mathcal{C}_d$. We consider two prominent examples of functions in $\mathcal{C}_d$. As mentioned above, the first is the function $\mathrm{AND}_k : \{-1,1\}^k \to \{-1,1\}$, which we show is in $\mathcal{C}_d$ for $d = \Omega(k^{1/2})$. Hence, we recover a new proof of Razborov and Sherstov's lower bound.

▶ **Corollary 2.** *Let $\mathrm{MP}_n = \mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$ be the Minsky-Papert DNF. Then $A = [\mathrm{MP}_n(x \vee y)]_{x,y}$ has sign-rank $\exp(\Omega(n^{1/3}))$.*

Let $\mathrm{ED}_k \colon \{-1,1\}^k \to \{-1,1\}$ denote the well-known ELEMENT DISTINCTNESS function (defined in Section 2.6). As we will show, the function ED_k is in $\mathcal{C}_d$ for some $d = \tilde{\Omega}(k^{2/3})$. Hence, we obtain the following corollary, which improves Razborov and Sherstov's lower bound on the sign-rank of AC^0 from $\exp(\Omega(n^{1/3}))$ to $\exp(\tilde{\Omega}(n^{2/5}))$.

▶ **Corollary 3.** *Let $F_n^{\mathrm{ED}} = \mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}} \circ \mathrm{AND}_C \circ \mathrm{OR}_2$. Then $A = [F_n^{\mathrm{ED}}(x,y)]_{x,y}$ has sign-rank $\exp(\tilde{\Omega}(n^{2/5}))$.*

As discussed in Section 2.6, the function F_n^{ED} is computed by a depth-3 AC^0 circuit with logarithmic bottom fan-in.

1.2 Applications

We describe applications of Corollary 3 to communication complexity, learning theory, and circuit complexity in detail in the full version of this work. Here, we briefly describe these applications.

- Razborov and Sherstov's result yielded a function in the communication complexity class $\mathbf{PH}^{\mathsf{cc}}$ (the communication analog of the polynomial hierarchy) that requires unbounded error communication complexity $\Omega(n^{1/3})$. This was the first separation between the communication complexity classes $\mathbf{PH}^{\mathsf{cc}}$ and $\mathbf{UPP}^{\mathsf{cc}}$, answering a longstanding open problem of Babai, Frankl, and Simon [4]. We improve this separation, giving a function in the communication complexity class $\mathbf{PH}^{\mathsf{cc}}$ (indeed, in $\mathbf{\Sigma}_2^{\mathsf{cc}}$) that requires unbounded error communication complexity $\tilde{\Omega}(n^{2/5})$.
- Razborov and Sherstov's result implied that learning algorithms in the sign-rank framework cannot PAC learn DNF formulae in time less than $\exp(O(n^{1/3}))$. This essentially matches the $\exp(\tilde{O}(n^{1/3}))$ runtime of the sign-rank based algorithm of Klivans and Servedio [13]. It is reasonable to ask whether the sign-rank framework can be used to learn depth-3 (or deeper) AC^0 circuits in the same $\exp(\tilde{O}(n^{1/3}))$ time bound. Our results rule this out, showing that sign-rank based learning algorithms require time $\exp(\tilde{\Omega}(n^{2/5}))$ to learn depth-3 AC^0 circuits, even when the bottom fan-in is $O(\log n)$.

- Razborov and Sherstov's result implied an exponential (specifically, $\exp(\Omega(n^{1/3}))$) lower bound on the size of threshold-of-majority circuits computing a function in AC^0. We improve their lower bound to $\exp\left(\tilde{\Omega}(n^{2/5})\right)$.

1.3 Our Techniques

It is well-known that the threshold degree of any function $h\colon \{-1,1\}^n \to \{-1,1\}$ is characterized by an (exponentially large) linear program. Using this formulation, if $\deg_\pm(h) = d$, then strong LP duality guarantees the existence of a dual solution μ that *witnesses* the fact that $\deg_\pm(h) \geq d$. Specifically, μ takes the form of a distribution on $\{-1,1\}^n$ such that h is uncorrelated under μ with all polynomials of degree at most d, i.e., $\sum_{x\in\{-1,1\}^n} \mu(x)\cdot h(x)\cdot p(x) = 0$ for all polynomials p of degree at most d. Razborov and Sherstov refer to μ as a *d-orthogonalizing distribution* for h (see Section 2.5 below for details).

In order to establish sign-rank lower bounds for the matrix $A = [(h \circ \mathrm{AND}_C)(x \vee y)]_{x,y}$, Razborov and Sherstov extended a lemma of Forster to show that it is enough to give an orthogonalizing distribution μ for h that additionally satisfies a *smoothness* property (cf. Theorem 18 in Section 4 for details). Specifically, a d-orthogonalizing distribution for h is said to be smooth if $\mu(x) = \exp(-O(d))$ for all but an $\exp(-\Omega(d))$ fraction of inputs $x \in \{-1,1\}^n$. Intuitively, this means that μ is smooth if it places "noticeable" mass on "almost all" inputs.

Razborov and Sherstov proved (non-constructively) that there exists a smooth d-orthogonalizing distribution for the Minsky-Papert DNF, for $d = n^{1/3}$. To generalize their result, for any $d > 0$ and any function $f \in \mathcal{C}_d$, we explicitly construct a smooth d-orthogonalizing distribution for the function $\mathrm{OR}_d \circ f$. Our construction combines new ideas with insights of Razborov and Sherstov, and ideas from prior works by the authors and Sherstov [7, 23] that constructed (non-smooth) orthogonalizing distributions for functions of the form $\mathrm{OR} \circ f$.

2 Preliminaries

2.1 Notation

We work with Boolean functions $f : \{-1,1\}^k \to \{-1,1\}$, where -1 corresponds to logical TRUE and $+1$ corresponds to logical FALSE. For $x \in \{-1,1\}^k$, let $|x| = \#\{i : x_i = -1\}$ denote the Hamming weight of x. Note that $|x|$ is computed by the linear function $|x| = \frac{k}{2} - \frac{1}{2}\sum_{i=1}^k x_i$.

2.2 Symmetrization

▶ **Definition 4.** Let $T : \{-1,1\}^k \to D$, where D is a finite subset of $\mathbb{R}^n$ for some $n \in \mathbb{N}$. The map T is *degree non-increasing* if for every polynomial $p : \{-1,1\}^k \to \mathbb{R}$, there exists a polynomial $q : D \to \mathbb{R}$ with $\deg q \leq \deg p$ such that

$$q(T(x)) = \mathop{\mathbb{E}}_{y \text{ s.t. } T(y)=T(x)} [p(y)]$$

for every $x \in \{-1,1\}^k$. We say that a degree non-increasing map T *symmetrizes* a function $f : \{-1,1\}^k \to \mathbb{R}$ if $f(x) = f(y)$ whenever $T(x) = T(y)$, and in this case we say that T is a *symmetrization* for f.

The canonical example of a degree non-increasing map is that which computes the Hamming weight.

▶ **Lemma 5** (Minsky and Papert [15]). *The map $T : \{-1,1\}^k \to \{0,1,\dots,k\}$ defined by $T(x) = |x|$ is degree non-increasing. Hence, T is a symmetrization for any symmetric Boolean function.*

For any function $\psi\colon \{-1,1\}^k \to \mathbb{R}$, a symmetrization $T\colon \{-1,1\}^k \to D$ for ψ induces a symmetrized function $\widetilde{\psi}\colon D \to \mathbb{R}$ defined via $\widetilde{\psi}(z) := \mathbb{E}_{x\in T^{-1}(z)}\psi(x)$. (If $T^{-1}(z)$ is empty, then we define $\widetilde{\psi}(z) = 0$). It will also be convenient to define an "unnormalized" version $\hat{\psi}$ of $\widetilde{\psi}$, defined via $\hat{\psi}(z) := \sum_{x\in T^{-1}(z)} \psi(x)$. Observe that if μ is a distribution on $\{-1,1\}^k$, then $\hat{\mu}$ is a distribution on D.

Similarly, let $T\colon \{-1,1\}^k \to D$ be a degree non-increasing map. A function $\hat{\psi} : D \to \mathbb{R}$ naturally induces an un-symmetrized function $\psi : \{-1,1\}^k \to \mathbb{R}$ by setting $\psi(x) = \frac{1}{|T^{-1}(z)|}\hat{\psi}(z)$ where $z = T(x)$. That is, ψ spreads the mass of $\hat{\psi}(z)$ out evenly over points $x \in T^{-1}(z)$. Observe that, for any $\hat{\psi}$ and any degree non-increasing map T, the induced function ψ is symmetrized by T.

We will often pass back and forth between a function ψ on $\{-1,1\}^k$ and its symmetrized versions $\widetilde{\psi}$ and $\hat{\psi}$ on D, when the underlying symmetrization $T\colon \{-1,1\}^k \to D$ is understood.

2.3 Norms and Inner Products

For a function $\psi : \{-1,1\}^k \to \mathbb{R}$, define the ℓ_1 norm of ψ by $\|\psi\|_1 = \sum_{x\in\{-1,1\}^k} |\psi(x)|$. For functions $\psi, \varphi : \{-1,1\}^k \to \mathbb{R}$, denote the inner product $\langle \psi, \varphi\rangle = \sum_{x\in\{-1,1\}^k} \psi(x)\varphi(x)$. We say a function $\psi : \{-1,1\}^k \to \mathbb{R}$ has *pure high degree* d if $\langle \psi, p\rangle = 0$ for every polynomial $p : \{-1,1\}^k \to \mathbb{R}$ of degree less than d.

2.4 Dual Objects and the Class $\mathcal{C}_d$

Central to our work is the following definition of a "dual object." We show that whenever a Boolean function f can be associated with such a dual object, then f can be transformed into a function F such that $[F(x,y)]_{x,y}$ has high sign-rank.

▶ **Definition 6.** Let $f : \{-1,1\}^k \to \{-1,1\}$, and let $T : \{-1,1\}^k \to D$ be a (degree non-increasing) symmetrization for f. Let $\hat{\psi}\colon D \to \mathbb{R}$ be any function, and let ψ be the associated function on $\{-1,1\}^k$ induced by T. We say that $\hat{\psi}$ is a (d,ε,η)-*dual object* for f (with respect to T) if:

$$\langle \psi, f\rangle \geq \varepsilon \tag{1}$$

$$\|\psi\|_1 = 1 \tag{2}$$

$$\langle \psi, p\rangle = 0 \text{ for every polynomial } p : \{-1,1\}^k \to \mathbb{R} \text{ with } \deg p < d \tag{3}$$

$$f(x) = -1 \implies \psi(x) < 0 \tag{4}$$

$$\hat{\psi}(z_+) \geq \eta \text{ for some } z_+ \in D \text{ satisfying } \widetilde{f}(z_+) = 1 \tag{5}$$

Definition 6 is motivated by a recent line of work establishing lower bounds for polynomial approximations via linear programming duality. We direct the reader to [21, 5, 25, 7, 23, 22, 19, 18, 6] for thorough discussions of this technique and its applications to longstanding open questions in complexity theory. In short, one can use linear programming duality to show that the existence of a $(d, 2\eta, \eta)$-dual object for a function f is implied by the *non-existence* of a degree d polynomial that approximates f in a certain precise sense. In a bit more detail (and still simplifying a little), a $(d, 2\eta, \eta)$-dual object for f always exists if f cannot be uniformly approximated to error 2η by any degree d polynomial, even under the promise that the input x is in $f^{-1}(-1) \cup T^{-1}(z_+)$. We will not use this *primal* interpretation of dual objects in our analysis, but we spell out this implication in the full version of this work for completeness and intuition.

Motivated by the study of uniform approximation of Boolean functions by polynomials, several works [24, 5, 6] have constructed dual objects directly. In particular, work of Špalek [24] and the authors [5] explicitly constructed an appropriate dual object for the AND function.

▶ **Lemma 7** (cf. [24, 5]). *Let $T : \{-1,1\}^k \to \{0,1,\ldots,k\}$ be the degree non-increasing map $T(x) = |x|$ that computes the Hamming weight. The function AND_k has a $(d, 1/2, 1/4)$-dual object with respect to T for $d = \Omega(\sqrt{k})$.*

We are now ready to define the class $\mathcal{C}_d$ of functions to which our techniques can be applied to yield sign-rank lower bounds.

▶ **Definition 8.** Let $f : \{-1,1\}^k \to \{-1,1\}$ be a Boolean function, and let $d > 0$. Then f is in the class $\mathcal{C}_d$ if there exists a symmetrization $T : \{-1,1\}^k \to D$ for f such that:
- there exists a $(d, 1/2, 1/4)$-dual object for f with respect to T, and
- the function f evaluates to TRUE (i.e. $f(x) = -1$) for at most a 2^{-d} fraction of inputs $x \in \{-1,1\}^k$.

2.5 Orthogonalizing Distributions

As indicated in Section 1.3, our analysis will make essential use of orthogonalizing distributions, which represent a dual formulation of the notion of threshold degree.

▶ **Definition 9.** A distribution $\mu : \{-1,1\}^n \to [0,1]$ is *d-orthogonalizing* for a function $h : \{-1,1\}^n \to \{-1,1\}$ if

$$\mathop{\mathbb{E}}_{x\sim\mu}[h(x)p(x)] = 0$$

for every polynomial $p : \{-1,1\}^n \to \mathbb{R}$ with $\deg p < d$. In other words, μ is d-orthogonalizing for h if the function $\mu(x)h(x)$ has pure high degree d.

2.6 The Element Distinctness Function

The Boolean function $\mathrm{ED}_k : \{-1,1\}^k \to \{-1,1\}$ is defined as follows. For simplicity, assume that $k = K \log_2 K$, where K is a power of 2. The function interprets its input x as blocks $x_1, \ldots, x_K$, where each $x_i \in \{-1,1\}^{\log_2 K}$. Each x_i is interpreted as the binary representation of $g_x(i)$ for a function $g_x : [K] \to [K]$. $\mathrm{ED}_k(x)$ is defined to equal -1 iff the function g_x is 1-to-1.

Observe that ED_k is symmetric with respect to permutations of the domain and range of g_x. That is, if $x, y \in \{-1,1\}^k$ are such that there exist permutations π, σ of $[K]$ with $g_x = \pi \circ g_y \circ \sigma$, then $\mathrm{ED}_k(x) = \mathrm{ED}_k(y)$.

In the full version of this work, we show that these symmetries imply the existence of a symmetrization T for ED_k and an associated dual object.

▶ **Lemma 10.** *There exists a symmetrization $T : \{-1,1\}^k \to [K]^K$ for the* Element Distinctness *function $\mathrm{ED}_k : \{-1,1\}^k \to \{-1,1\}$ such that ED_k has a $(d, 1/2, 1/4)$-dual object (with respect to the map T), for some $d = \Omega(K^{2/3}/\log K)$.*

▶ Remark. In fact, an explicit dual object for ED_k was constructed in our prior work [6]. In the full version of this work, we give an alternative *primal*-based proof of the existence of a dual object for ED_k. The proof is based on Aaronson and Shi's [1] influential lower bound of $\Omega(K^{2/3})$ on the *approximate degree*[1] of ED_k.

The ELEMENT DISTINCTNESS function is computed by a natural CNF formula:

$$\mathrm{ED}_k(x_1, \dots, x_K) = \bigwedge_{r=1}^{K} \bigwedge_{i \neq j} ((x_i \neq r) \vee (x_j \neq r)).$$

Notice that the fan-in of each bottom OR gate is only $2K \leq 2\log_2 k$. Recall (cf. Corollary 3) that our aim is to prove a sign-rank lower bound for the function $F_n^{\mathrm{ED}}(x, y) = (\mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}} \circ \mathrm{AND}_C)(x \vee y)$. Using the CNF for ELEMENT DISTINCTNESS described above, the function F_n^{ED} is naturally computed by an AC^0 circuit Γ of depth 5, with an OR gate at the top. However, as we now explain, F_n^{ED} is actually computable by a *depth-3* AC^0 circuit with logarithmic bottom fan-in.

Number the layers of Γ from 1 to 5, with layer 1 corresponding to the OR gate at the top. Since each OR gate at layer 3 of Γ has fan-in $O(\log n)$ (and the gates at layers 4 and 5 have constant fan-in), the sub-circuits rooted at each gate at layer 3 of Γ are functions of only $O(\log n)$ bits of x. Since any function on $O(\log n)$ inputs can be computed by a $\mathrm{poly}(n)$ size CNF with logarithmic bottom fan-in, we can replace each sub-tree rooted at layer 3 of Γ with such a CNF, to obtain a circuit Γ' of depth 4, in which layers 2 and 3 of Γ' both consist of AND gates. Collapsing layers 3 and 4 into a single layer yields a polynomial size depth 3 circuit with logarithmic bottom fan-in that computes F_n^{ED}.

3 Constructing a Smooth Orthogonalizing Distribution

Sherstov [23] showed that whenever f has a $(d_1, 1/2, 0)$-dual object[2], the function $h_m := \mathrm{OR}_m \circ f$ has a d-orthogonalizing distribution for $d = \min\{m, d_1\}$. The goal of this section, and the main technical contribution of the paper, is to prove that whenever f has a $(d_1, 1/2, \eta)$-dual object for $\eta > 0$, the function h_m has a d-orthogonalizing distribution that places significant mass on each input $x \in h_m^{-1}(1)$. More precisely, we show:

▶ **Theorem 11.** *Suppose that $f : \{-1,1\}^k \to \{-1,1\}$ has a $(d_1, 1/2, \eta)$-dual object, and let $h_m = \mathrm{OR}_m \circ f$. Then there exists a d-orthogonalizing distribution $\mu : \{-1,1\}^{mk} \to [0,1]$ for h_m such that $\mu(x) \geq 4^{-(m+d+1)}\eta^{-m/2}2^{-mk}$ for every $x \in h_m^{-1}(1)$, where $d = \min\{m/2, d_1\}$.*

Combining this theorem with Lemmas 7 and 10 yields smooth orthogonalizing distributions for the functions $\mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$ and $\mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}$.

▶ **Corollary 12.** *There exists a d-orthogonalizing distribution μ for $h = \mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$ such that $\mu(x) \geq 2^{-O(d)}2^{-n}$ on each $x \in h^{-1}(1)$, for some $d = \Omega(n^{1/3})$.*

▶ **Corollary 13.** *There exists a d-orthogonalizing distribution μ for $h = \mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}}$ such that $\mu(x) \geq 2^{-O(d)}2^{-n}$ on each $x \in h^{-1}(1)$, for some $d = \tilde{\Omega}(n^{2/5})$.*

---

1 The approximate degree of a Boolean function f is the minimum degree of a real polynomial for which $|p(x) - f(x)| \leq 1/3$ for all Boolean inputs x.

2 The existence of a $(d_1, 1/2, 0)$-dual object for f is in fact a dual formulation of the property that f has *one-sided approximate degree* at least d_1. See [5, 23] for the definition of one-sided approximate degree.

3.1 Proof of Theorem 11

3.1.1 Notation

Let $f : \{-1,1\}^k \to \{-1,1\}$ be as in the statement of Theorem 11, and let $T : \{-1,1\}^k \to D$ be the symmetrization for f associated with the assumed $(d_1, 1/2, \eta)$-dual object for f. Define $T^m : \{-1,1\}^{mk} \to D^m$ by $T^m(x_1, \ldots, x_m) := (T(x_1), \ldots, T(x_m))$. Since T is degree non-increasing, it is easy to see that T^m is also degree non-increasing. Moreover, T^m is a symmetrization for h_m. The map T^m induces a symmetrized version $\widetilde{h}_m : D^M \to \mathbb{R}$ of h_m given by $\widetilde{h}_m = \mathrm{OR}_m \circ \widetilde{f}$.

3.1.2 Proof Outline

Let $Z^+ := \widetilde{h}_m^{-1}(1) \subseteq D^m$. At a high level, our proof will produce, for every $z \in Z^+$, a d-orthogonalizing distribution μ_z that is targeted to z, in the sense that

$$\hat{\mu}_z(z) \geq 2^{-O(m+d)} \cdot \eta^{-O(m)}.$$

Since the property of d-orthogonalization is preserved under averaging, the distribution $\mu = \frac{1}{|Z^+|} \sum_{z \in Z^+} \mu_z$ remains d-orthogonalizing, and places the required amount of probability mass on each input $x \in T^{-1}(Z^+) = h_m^{-1}(1)$. The goal therefore becomes to construct these targeted distributions μ_z. We do this in two stages.

Stage 1. In the first stage (see Claim 15 below), we construct distributions μ_z for every z belonging to a highly structured subset $G \subset Z^+$ that we now describe. Let $c \in \widetilde{f}^{-1}(1)$ denote the point on which the dual object $\hat{\psi}$ for f has $\hat{\psi}(c) \geq \eta$ (cf. Condition (5) within Definition 6). The set G consists of inputs in Z^+ for which $c \in D$ is repeated many times (specifically, at least $m/2$ times).

Stage 2. In the second stage (see Claim 16 below), we show that given the family of distributions $\{\mu_z : z \in G\}$ constructed in Stage 1, we can construct appropriate distributions μ_z for z belonging to the entire set Z^+.

Both stages can be viewed as generalized dual counterparts to analogous statements in the work of Razborov and Sherstov (cf. [17, Lemma 3.4] and [17, Theorem 3.6] respectively). Taking a dual perspective allows us to identify general properties (Definition 6) of a dual object for f that enable the construction of a smooth orthogonalizing distribution. This results in a much more general and modular framework for proving the existence of these distributions. Our framework also has the advantage of constructing smooth orthogonalizing distributions explicitly.

3.1.3 Proof Details

We begin with a relatively simple lemma that shows that the function $\mathrm{OR}_{m/2} \circ f$ has a d-orthogonalizing distribution μ such that $\hat{\mu}$ places a lot of probability mass on a particular highly structured input, where $d = \min\{d_1, m/2\}$. This distribution is an important building block in the proof of Claim 15 below.

▶ **Lemma 14.** *Let $\ell = m/2$, and let f, T, and T^ℓ be as above. Consider the function $h_\ell : \{-1,1\}^{k\ell} \to \{-1,1\}$ defined by $h_\ell(x_1, \ldots, x_\ell) = \mathrm{OR}_\ell(f(x_1), \ldots, f(x_\ell))$. There exists a* *function* $\psi : \{-1,1\}^{k\ell} \to [0,1]$ *symmetrized by* T^ℓ *with the following properties.*

$$\psi \text{ agrees in sign with } h_\ell. \text{ That is, } \psi(x) \cdot h_\ell(x) \geq 0 \text{ for all } x \in \{-1,1\}^{k\ell} \tag{6}$$

$$\|\psi\|_1 = 1 \tag{7}$$

$$\psi \text{ has pure high degree at least } d = \min\{\ell, d_1\} \tag{8}$$

$$\text{There exists a } c \in D \text{ such that } \widetilde{f}(c) = 1 \text{ and } \hat{\psi}(\underbrace{c, \dots, c}_{\ell \text{ times}}) \geq \eta^{-\ell}/2 \tag{9}$$

We remark that Conditions (6)–(8) are equivalent to requiring that $\mu := \psi \cdot h_\ell$ is a d-orthogonalizing distribution for h_ℓ, where $d = \min\{\ell, d_1\}$.

Proof Sketch. Sherstov [23] showed that when the function f has a $(d_1, 1/2, 0)$-dual witness, then there is a function ψ satisfying Conditions (6)–(8). In the full version of this work, we show that if f additionally has a $(d_1, 1/2, \eta)$-dual witness with $\eta > 0$, then Sherstov's construction yields a function $\hat{\psi}$ that also satisfies Condition (9). ◀

To complete Stage 1 of our proof, we show that for every input $w \in D^m$ that is close in Hamming distance to the special point $(\underbrace{c, \dots, c}_{m \text{ times}})$, there is an orthogonalizing distribution for h_m that places substantial weight on w. Let $G \subset Z^+ = \widetilde{h}_m^{-1}(1)$ denote the set of inputs in $\widetilde{h}_m^{-1}(1)$ that take the value c on at least $m/2$ coordinates. That is,

$$G = \{z \in Z^+ : \exists i_1, \dots, i_{m/2} \text{ s.t. } z_{i_1} = \dots = z_{i_{m/2}} = c\}.$$

▶ **Claim 15.** *Let G be as above. For every $w = (w_1, \dots, w_m) \in G$, there exists a d-orthogonalizing distribution $\nu_w : \{-1,1\}^{km} \to [0,1]$ for h_m such that ν_w is symmetrized by T^m and $\hat{\nu}_w(w) \geq \eta^{m/2}/2$.*

Proof Sketch. Let $I = \{i_1, \dots, i_{m/2}\}$ denote the first $m/2$ coordinates on which w takes the value c. Define the distribution $\hat{\nu}_w$ by

$$\hat{\nu}_w(z) = \begin{cases} |\hat{\psi}(z_{i_1}, \dots, z_{i_{m/2}})| & \text{if } z_i = w_i \text{ for all } i \notin I \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{\psi}$ is the function from Lemma 14 for $\ell = m/2$. It is immediate from the definition that $\hat{\nu}_w$ is a distribution on D^m, and hence ν_w is a distribution on $\{-1,1\}^{km}$. Moreover, $\hat{\nu}_w(w) \geq \eta^{m/2}/2$. The fact that ν_w is d-orthogonalizing follows from the fact that ψ has pure high degree at least d. This calculation appears in the full version of this work. ◀

We now proceed to Stage 2 of our proof, in which we use the distributions constructed in Claim 15 to give orthogonalizing distributions that place significant weight on *any* input $x \in h_m^{-1}(1)$.

▶ **Claim 16.** *Let G be as before, and suppose that for every $w \in G$ there exists a d-orthogonalizing distribution $\nu_w : \{-1,1\}^{km} \to [0,1]$ for h_m that is symmetrized by T^m, and satisfies $\hat{\nu}_w(w) \geq \delta$. Then for every $v \in (Z^+ \setminus G)$, there exists a d-orthogonalizing distribution ρ_v that is symmetrized by T^m, and $\hat{\rho}_v(v) \geq \delta/4^{m+d}$.*

The main technical ingredient in the proof of Claim 16 is the construction of a function $\varphi : \{0,1\}^m \to \mathbb{R}$ of pure high degree d for which $\varphi(1^m)$ is "large". This can be viewed as a dual formulation of a bound on the growth of low-degree polynomials. The construction of φ appears as part of the proof of such a bound in [17].

► Remark. We choose to state Lemma 17 below for a function $\varphi : \{0,1\}^m \to \mathbb{R}$, rather than applying our usual convention of working with functions over $\{-1,1\}^m$, because it makes various statements in the proof of Claim 16 cleaner. To clarify the terminology below, we say a function $\varphi : \{0,1\}^m \to \mathbb{R}$ has *pure high degree* d if $\sum_{x\in\{0,1\}^m} \varphi(x)\cdot p(x) = 0$ for every polynomial $p : \{0,1\}^m \to \mathbb{R}$ of degree less than d. The Hamming weight function $|\cdot| : \{0,1\}^m \to [m]$ counts the number of 1's in its input, i.e. $|s| = s_1 + s_2 + \cdots + s_m$.

► **Lemma 17** (cf. [17, Proof of Lemma 3.2]). *Let d be an integer with $0 \le d \le m-1$. Then there exists a function $\varphi : \{0,1\}^m \to \mathbb{R}$ such that*

$$\varphi(1^m) = 1 \tag{10}$$

$$\varphi(x) = 0 \textit{ for all } d \le |x| < m \tag{11}$$

$$\varphi \textit{ has pure high degree at least } d \tag{12}$$

$$\sum_{|x|\le d} |\varphi(x)| \le 2^d \binom{m}{d} \tag{13}$$

Proof of Claim 16. Fix $v \in (Z^+ \setminus G)$. Define an auxiliary function $\hat{\varphi}_v : D^m \to [0,1]$ as follows. For any $z = (z_1, \ldots, z_m)$, let

$$\hat{\varphi}_v(z) := \sum_{\substack{s\in\{0,1\}^m \text{ s.t.}\\ \forall i\ z_i = s_i c + (1-s_i)v_i}} \varphi(s),$$

where φ is as in Lemma 17, with d set as in the conclusion of Claim 15 (observe that if there is some z_i such that $z_i \ne c$ and $z_i \ne v_i$, then $\hat{\varphi}_v(z) = 0$).

Letting φ_v denote the function on $\{-1,1\}^{km}$ induced from $\hat{\varphi}_v$ by T^m, we record some properties of φ_v and $\hat{\varphi}_v$.

$$\hat{\varphi}_v(v) = \varphi(1^m) = 1 \tag{14}$$

$$\operatorname{supp} \hat{\varphi}_v \subset G \cup \{v\} \tag{15}$$

$$\varphi_v \text{ has pure high degree at least } d \tag{16}$$

$$\|\varphi_v\|_1 \le 2^d \binom{m}{d} + 1 \tag{17}$$

$$\hat{\varphi}_v \text{ is supported on at most } \frac{1}{2}2^m + 1 \text{ points in } D^m \tag{18}$$

Verifying Conditions (14)–(18). Conditions (14), (15), and (18) are immediate from the definition of $\hat{\varphi}_v$, combined with Conditions (10) and (11) of Lemma 17. For Condition (16), it is enough to show that if $p_1, \ldots, p_m$ are polynomials over $\{-1,1\}^k$ whose degrees sum to less than d, then $\sum_{x=(x_1,\ldots,x_m)\in\{-1,1\}^{km}} \varphi_v(x) \prod_{i=1}^m p_i(x_i) = 0$. To establish this, let $q_1, \ldots, q_m : D \to \mathbb{R}$ denote polynomials satisfying $\deg(q_i) \le \deg(p_i)$, and such that for all i and all z_i in the image of T, $q_i(z_i) := \mathbb{E}_{x\in T^{-1}(z_i)}[p_i(z_i)]$. Such polynomials are guaranteed to exist, since T is degree non-increasing. Then:

$$\begin{aligned}\sum_{x=(x_1,\ldots,x_m)\in\{-1,1\}^{km}} \varphi_v(x)\prod_{i=1}^m p_i(x_i) &= \sum_{z=(z_1,\ldots,z_m)\in D^m} \hat{\varphi}_v(z)\prod_{i=1}^m q_i(z_i)\\ &= \sum_{z=(z_1,\ldots,z_m)\in D^m} \left(\sum_{\substack{s\in\{0,1\}^m \text{ s.t.}\\ \forall i\ z_i=s_ic+(1-s_i)v_i}} \varphi(s)\right)\prod_{i=1}^m q_i(z_i)\\ &= \sum_{s\in\{0,1\}^m} \varphi(s)\prod_{i=1}^m q_i(s_ic+(1-s_i)v_i)\\ &= 0,\end{aligned}$$

To see that the final equality holds, recall that that degrees of the polynomials q_i sum to strictly less than d. Hence, $p(s_1,\ldots,s_m) := \prod_{i=1}^m q_i(s_ic+(1-s_i)v_i)$ is a polynomial of degree strictly less than d over $\{-1,1\}^m$. The final equality then follows from the fact that φ has pure high degree at least d.

To establish Condition (17), we check that

$$\sum_{z\in D^m, z\neq v} |\hat{\varphi}_v(z)| \leq \sum_{s\in\{0,1\}^m, s\neq 1^m} |\varphi(s)| \leq 2^d\binom{m}{d},$$

where the final inequality holds by Condition (13).

Construction and analysis of ρ_v. Up to normalization, the function $\varphi_v\cdot h_m$ has all of the properties that we need to establish Claim 16, except that there are locations where it may be negative. We obtain our desired orthogonalizing distribution ρ_v by adding correction terms to $\hat{\varphi}_v$ in the locations where $\hat{\varphi}_v$ may disagree with $\widetilde{h}_m$ in sign. These correction terms are derived from the distributions $\hat{\nu}_w$ whose existence are hypothesized in the statement of Claim 16. We start by defining

$$\hat{P}_v(z) = \frac{\delta}{2^d\binom{m}{d}+1}\widetilde{h}_m(z)\hat{\varphi}_v(z) + \sum_{w\in(\operatorname{supp}\hat{\varphi}_v\setminus\{v\})} \hat{\nu}_w(z). \tag{19}$$

Observe that each w appearing in the sum on the right hand side of Eq. (19) is in the set G, owing to Condition (15). This guarantees that each term $\hat{\nu}_w$ in the sum is well-defined.

Now we check that $\hat{P}_v$ is nonnegative. Since each term $\hat{\nu}_w$ appearing in the sum on the right hand side of Eq. (19) is a distribution (and hence non-negative), it suffices to check that $\hat{P}_v(z)\geq 0$ for each point $z\in\operatorname{supp}\hat{\varphi}_v$. On each such point with $z\neq v$, Condition (17) guarantees that $\frac{\delta}{2^d\binom{m}{d}+1}\widetilde{h}_m(z)\hat{\varphi}_v(z)\geq -\delta$. Moreover, the contribution of the sum is at least $\hat{\nu}_z(z)\geq\delta$ by hypothesis. Hence, $\hat{P}_v$ is a non-negative function.

Next, we check that normalizing $\hat{P}_v$ yields a distribution $\hat{\rho}_v := \hat{P}_v/\|P_v\|_1$ for which $\hat{\rho}_v(v)\geq \delta/4^{m+d}$ as required. By construction, $\hat{P}_v(v) = \delta/\left(2^d\binom{m}{d}+1\right)$. Moreover, Conditions (14), (17), and (18) together show that $\|\hat{P}_v\|_1 \leq \delta + \frac{1}{2}2^m \leq 2^m$. Hence, $\hat{P}_v(v)\geq \delta/\left(2^m\cdot\left(2^d\binom{m}{d}+1\right)\right) \geq \delta/\left(2^{m+d+1}\binom{m}{d}\right) \geq \delta/2^{2m+d+1} \geq \delta/4^{m+d}$.

Finally, we must check that $\rho_v = P_v/\|P_v\|_1$ is d-orthogonalizing for h_m. To see this, observe that $P_v\cdot h_m$ is a linear combination of the functions φ_v and $\nu_w\cdot h_m$ for $w\in(\operatorname{supp}\hat{\varphi}_v\setminus\{v\})$. Moreover, each of these functions has pure high degree at least d (φ_v does so by Condition (16), while $\nu_w\cdot h_m$ does by the fact that ν_w is d-orthogonalizing for h_m). By linearity, it follows that $P_v \cdot h_m$ has pure high degree at least d, so ρ_v is d-orthogonalizing for h_m as desired.

This completes the proof of Claim 16. ◀

We are now ready to combine the claims above to prove Theorem 11.

Proof of Theorem 11. By Claim 15, for every $w \in G$ there exists a d-orthogonalizing distribution $\nu_w : \{-1,1\}^{km} \to [0,1]$ for h_m that is symmetrized by T^m, with $\hat{\nu}_w(w) \geq \eta^{m/2}/2$. Thus, by Claim 16, it is also true that for every $v \in (Z^+ \setminus G)$, there is a d-orthogonalizing distribution $\rho_v : \{-1,1\}^{km} \to [0,1]$ that is symmetrized by T^m, with $\hat{\rho}_v(v) \geq \eta^{m/2}4^{-(m+d+1)}$. Now consider the distribution

$$\hat{\mu}(z) = \frac{1}{|Z^+|}\left(\sum_{w\in G}\hat{\nu}_w(z) + \sum_{v\in(Z^+\setminus G)}\hat{\rho}_v(z)\right).$$

The (un-symmetrized) distribution $\mu : (\{-1,1\}^k)^m \to [0,1]$ satisfies $\mu(x) \geq \eta^{m/2}4^{-(m+d+1)}2^{-km}$ for every point $x \in T^{-1}(Z^+) = h_m^{-1}(1)$. Moreover, μ remains d-orthogonalizing for h_m, as it is a sum of d-orthogonalizing distributions for h_m. ◀

4 Sign Rank Lower Bounds for AC^0

We now use the machinery developed by Razborov and Sherstov to translate our construction of a smooth orthogonalizing distribution into a sign-rank lower bound.

▶ **Theorem 18** (Implicit in [17, Theorem 1.1])**.** *Let $h\colon \{-1,1\}^n \to \{-1,1\}$ be a Boolean function, and suppose there exists a d-orthogonalizing distribution μ for h such that $\mu(x) \geq 2^{-cd}2^{-n}$ for all but a 2^{-cd} fraction of inputs $x \in \{-1,1\}^n$. Then there exists a constant C (depending only on c) such that if $F(x,y) := h(\ldots, \wedge_{j=1}^C(x_{ij} \vee y_{ij}), \ldots)$, then the matrix $[F(x,y)]_{x,y}$ has sign-rank $\exp(\Omega(d))$.*

Combining Theorem 18 with Theorem 11 yields the main result of this work.

▶ **Theorem 19.** *Let $f : \{-1,1\}^k \to \{-1,1\}$ be a Boolean function in the class $\mathcal{C}_d$. Let $F : \{-1,1\}^n \to \{-1,1\}^n$ be defined by*

$$F = \mathrm{OR}_{2d} \circ f \circ \mathrm{AND}_C \circ \mathrm{OR}_2,$$

where C is the universal constant of Theorem 18 (and hence $n = O(dk)$). The sign-rank of the matrix $[F(x,y)]_{x,y}$ is $\exp(\Omega(d))$.

Proof. Let $h_{2d} : \{-1,1\}^{2dk} \to \{-1,1\}$ denote the function $h_{2d} = \mathrm{OR}_{2d} \circ f$. By Theorem 11, there exists a d-orthogonalizing distribution μ for h_{2d} such that $\mu(x) \geq 2^{-9d}2^{-2dk}$ for every $x \in h_{2d}^{-1}(1)$. Since $f \in \mathcal{C}_d$, we have by a union bound that $h_{2d}^{-1}(1)$ contains all but a $(2d) \cdot 2^{-d} \leq 2^{-d/2}$ fraction of the points in $\{-1,1\}^{2dk}$. Thus, by Theorem 18, there is a universal constant C for which $[F(x,y)]_{x,y}$ has sign-rank $\exp(\Omega(d))$. ◀

▶ **Corollary 20.** *Let $\mathrm{MP}_n = \mathrm{OR}_{n^{1/3}} \circ \mathrm{AND}_{n^{2/3}}$ be the Minsky-Papert DNF. Then $[\mathrm{MP}_n(x \vee y)]_{x,y}$ has sign-rank $\exp(\Omega(n^{1/3}))$*

Proof. The function AND_k evaluates to TRUE on exactly 1 out of 2^k inputs. Hence, by Lemma 7, we have $\mathrm{AND}_k \in \mathcal{C}_d$ for $d = \Omega(k^{1/2})$. Let $F = \mathrm{MP}_n \circ \mathrm{AND}_C \circ \mathrm{OR}_2$. Applying Theorem 19 implies that the sign-rank of $[F(x,y)]_{x,y} = \exp(\Omega(n^{1/3}))$. Merging the two adjacent layers of AND gates in the natural circuit computing F yields the desired result. ◀

▶ **Corollary 21.** *Let* $F_n^{\mathrm{ED}} = \mathrm{OR}_{n^{2/5}} \circ \mathrm{ED}_{n^{3/5}} \circ \mathrm{AND}_C$. *Then* $[F_n^{\mathrm{ED}}(x \vee y)]_{x,y}$ *has sign-rank* $\exp(\tilde{\Omega}(n^{2/5}))$

Proof. Assume for simplicity that $k = K \log K$. The function ED_k evaluates to TRUE on exactly $K!$ inputs, which is an $\exp(-O(K))$ fraction of the $2^k = K^K$ total inputs. Hence, by Lemma 10, we have $\mathrm{ED}_k \in \mathcal{C}_d$ for $d = \Omega(K^{2/3}/\log K)$. The result follows by applying Theorem 19. ◀

References

1 Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004.

2 Noga Alon, Peter Frankl, and Vojtech Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *26th Annual Symposium on Foundations of Computer Science, Portland, Oregon, USA, 21-23 October 1985*, pages 277–280, 1985.

3 Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank, VC dimension and spectral gaps. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:135, 2014.

4 László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986*, pages 337–347, 1986.

5 Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and Markov-Bernstein inequalities. In *ICALP (1)*, pages 303–314, 2013.

6 Mark Bun and Justin Thaler. Dual polynomials for collision and element distinctness. *CoRR*, abs/1503.07261, 2015. URL: `http://arxiv.org/abs/1503.07261`.

7 Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *Automata, Languages, and Programming – 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 268–280, 2015.

8 Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002.

9 Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt, and Hans-Ulrich Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *FST TCS 2001: Foundations of Software Technology and Theoretical Computer Science, 21st Conference, Bangalore, India, December 13-15, 2001, Proceedings*, pages 171–182, 2001.

10 Jürgen Forster and Hans-Ulrich Simon. On the smallest possible dimension and the largest possible margin of linear arrangements representing given concept classes uniform distribution. In *Algorithmic Learning Theory, 13th International Conference, ALT 2002, Lübeck, Germany, November 24-26, 2002, Proceedings*, pages 128–138, 2002.

11 Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:49, 2015. URL: `http://eccc.hpi-web.de/report/2015/049`.

12 Lisa Hellerstein and Rocco A. Servedio. On PAC learning algorithms for rich boolean function classes. *Theor. Comput. Sci.*, 384(1):66–76, 2007.

13 Adam R. Klivans and Rocco A. Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. Syst. Sci.*, 68(2):303–318, 2004.

14 Nati Linial, Shahar Mendelson, Gideon Schechtman, and Adi Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.

15 Marvin Minsky and Seymour Papert. *Perceptrons – an introduction to computational geometry.* MIT Press, 1969.

16 Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.
17 Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010.
18 A. A. Sherstov. The power of asymmetry in constant-depth circuits. In *Foundations of Computer Science*, 2015.
19 Alexander A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95:59–93, 2008.
20 Alexander A. Sherstov. The unbounded-error communication complexity of symmetric functions. *Combinatorica*, 31(5):583–614, 2011.
21 Alexander A. Sherstov. Approximating the AND-OR tree. *Theory of Computing*, 9(20):653–663, 2013.
22 Alexander A. Sherstov. The intersection of two halfspaces has high threshold degree. *SIAM J. Comput.*, 42(6):2329–2374, 2013.
23 Alexander A. Sherstov. Breaking the Minsky-Papert barrier for constant-depth circuits. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 – June 03, 2014*, pages 223–232, 2014.
24 Robert Spalek. A dual polynomial for OR. *CoRR*, abs/0803.4516, 2008. URL: `http://arxiv.org/abs/0803.4516`.
25 Justin Thaler. Lower bounds for the approximate degree of block-composed functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:150, 2014. URL: `http://eccc.hpi-web.de/report/2014/150`.