SIAM J. COMPUT.
Vol. 51, No. 3, pp. 820–848

A SHORT LIST OF EQUALITIES INDUCES LARGE SIGN-RANK*

ARKADEV CHATTOPADHYAY† AND NIKHIL S. MANDE‡

Abstract. We exhibit a natural function F_n on n variables that can be computed by just a linear-size decision list of "Equalities," but whose sign-rank is $2^{\Omega(n^{1/4})}$. This yields the following two new unconditional complexity class separations. 1. *Boolean circuit complexity.* The function F_n can be computed by linear-size depth-two threshold *formulas* when the weights of the threshold gates are unrestricted ($\mathsf{THR} \circ \mathsf{THR}$), but any $\mathsf{THR} \circ \mathsf{MAJ}$ *circuit* (the weights of the bottom threshold gates are polynomially bounded in n) computing F_n requires size $2^{\Omega(n^{1/4})}$. This provides the first separation between the Boolean circuit complexity classes $\mathsf{THR} \circ \mathsf{MAJ}$ and $\mathsf{THR} \circ \mathsf{THR}$. While Amano and Maruoka [Proceedings of the 30th International Symposium on Mathematical Foundations of Computer Science, 2005, pp. 107–118] and Hansen and Podolskii [Proceedings of the 25th Annual IEEE Conference on Computational Complexity, 2010, pp. 270–279] emphasized that superpolynomial separations between the two classes remained a basic open problem, our separation is in fact exponential. In contrast, Goldmann, Håstad, and Razborov [Comput. Complexity, 2 (1992), pp. 277–300] showed more than twenty-five years ago that functions efficiently computable by $\mathsf{MAJ} \circ \mathsf{THR}$ circuits can also be efficiently computed by $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits. In view of this, it was not even clear if $\mathsf{THR} \circ \mathsf{THR}$ was significantly more powerful than $\mathsf{THR} \circ \mathsf{MAJ}$ until our work, and there was no candidate function identified for the potential separation. 2. *Communication complexity.* The function F_n (under the natural partition of the inputs) lies in the communication complexity class P^{MA}. Since F_n has large sign-rank, this implies $\mathsf{P}^{\mathsf{MA}} \not\subseteq \mathsf{UPP}$, strongly resolving a recent open problem posed by Göös, Pitassi, and Watson [*Comput. Complexity*, 27 (2018), pp. 245–304]. In order to prove our main result, we view F_n as an XOR function and develop a technique to lower bound the sign-rank of such functions. This requires novel approximation-theoretic arguments against polynomials of *unrestricted* degree. Further, our work highlights for the first time the class "decision lists of exact thresholds" as a common frontier for making progress on longstanding open problems in threshold circuits and communication complexity.

Key words. Boolean threshold circuits, communication complexity, sign-rank, approximation theory

AMS subject classifications. 68Q15, 68Q17

DOI. 10.1137/19M1271348

1. Introduction.

Sign-rank is a delicate but powerful notion, which has a matrix rigidity-like flavor. It was introduced in the seminal work of Paturi and Simon [51]. The sign-rank of a $\{-1, 1\}$ valued matrix M is defined to be the minimum rank of a real valued matrix each of whose entries agrees in sign with the corresponding entry of M. Sign-rank has found numerous applications in computer science in areas like communication complexity, Boolean circuit complexity, and computational learning theory. Paturi and Simon showed that the logarithm of the sign-rank of a (communication) matrix is essentially equivalent to the unbounded-error two-party communication complexity of the underlying function. Forster et al. [24] showed that proving lower bounds on the sign-rank of a function gives lower bounds on the minimum size of any $\mathsf{THR} \circ \mathsf{MAJ}$ circuit computing it. Sign-rank is known to be equiv-

*Received by the editors June 28, 2019; accepted for publication (in revised form) January 26, 2022; published electronically June 14, 2022. A preliminary version of this article appeared in *Proceedings of the 59th IEEE Annual Symposium on Foundations of Computer Science*, 2018.
https://doi.org/10.1137/19M1271348

Funding: The first author was partially supported by a Ramanujan fellowship of the DST. The second author was partially supported by a TCS fellowship.

†School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, 400005, India (Arkadev.c@tifr.res.in).
‡QuSoft and CWI, Amsterdam, 1098XG, The Netherlands (nikhil.s.mande@gmail.com).

alent to dimension complexity, a geometric notion that is of fundamental importance in computational learning theory. Even proving lower bounds on the sign-rank of a random function is nontrivial and was first done by Alon, Frankl, and Rödl [3]. On the other hand, proving strong lower bounds on the sign-rank of an explicit function, IP, was a breakthrough achieved by Forster [23] fifteen years later. Since that work, there have relatively been just a few results proving strong sign-rank lower bounds on explicit functions [59, 54, 12, 9, 14, 61]. While many basic questions about sign-rank remain unanswered, new connections between it and other areas of mathematics keep showing up (see, for example, [4]).

We consider the following easily describable function F_n: The input, of length $n = 2m\ell$, is split into two disjoint parts, $X \in \{-1,1\}^{m\ell}$ and $Y \in \{-1,1\}^{m\ell}$.[1] X and Y are each further divided into ℓ disjoint blocks $X_1, \dots, X_\ell, Y_1, \dots, Y_\ell$, of length m each. The function F_n outputs -1 iff the largest index $i \in [\ell]$ for which $X_i = Y_i$ holds is an odd index (in particular, F_n outputs 1 if there is no such index). For the purposes of this paper, we set $m = \ell^{1/3} + \log \ell$. It is not hard to see that F_n can be easily described as a decision list (formally defined in section 2) of Equalities. Decision lists are a natural class of functions to study and have widespread applications in learning theory, for example, [55, 42, 43].

Our main theorem shows a strong lower bound on the sign-rank of M_{F_n}, where the rows of M_{F_n} are indexed by the inputs X, the columns by Y, and the (x,y)th entry is $F_n(x,y)$. We overload notation and refer to the sign-rank of M_{F_n} as the sign-rank of F_n.

THEOREM 1.1 (*main*). *The function F_n has sign-rank $2^{\Omega(n^{1/4})}$.*

An active research program is to search for functions in AC^0 that are increasingly hard to *approximate* under various natural measures. For example, recently Bun and Thaler [13] gave near-optimal lower bounds for the approximate degree for functions in AC^0. Sign-rank is arguably one of the hardest notions of approximability to analyze. There have been a series of works [54, 12, 14] on improving sign-rank lower bounds for functions computable in AC^0. Most recently, Sherstov and Wu [61] showed near-optimal sign-rank lower bounds for functions computable in AC^0. All these results exploit the considerable computing power of AC^0 to come up with more intricate functions that are harder to approximate. In our work, Theorem 1.1 contrasts with these efforts by finding in some sense a *simpler* function F_n in AC^0, that still remarkably has large sign-rank. The building block of F_n is Equality, which is a very simple function under various models of computation. It turns out F_n can be computed by a *depth*-2 linear-size threshold formula. This simplicity of F_n, mainly in its depth complexity, enables us to settle two open problems. The first is a twenty-five year old (open since the work of Goldmann, Håstad, and Razborov [25]) and very basic problem of understanding the relative power of weights in depth-2 threshold circuits. This application of our result is outlined in subsection 1.1. The second problem, posed much more recently by Göös, Pitassi, and Watson [29], is a communication complexity class separation, outlined in subsection 1.2. Interestingly, our resolution of these two problems also serves to highlight an emerging barrier, which we call the "sign-rank barrier," against proving new lower bounds against depth-2 threshold circuits and communication protocols just above the first level of the polynomial hierarchy.

[1]Throughout this paper, we consider the input and output domains to be $\{-1,1\}^n$ and $\{-1,1\}$, rather than $\{0,1\}^n$ and $\{0,1\}$, respectively. -1 is identified with "True" and 1 with "False."

1.1. Application: Bottom weights can matter. Linear threshold functions (LTFs) form one of the most central classes of Boolean functions that are studied. Every such function corresponds to the halfspace induced by a real weight vector $\mathbf{w} \in \mathbb{R}^{n+1}$, denoted by $\mathsf{THR}_{\mathbf{w}}$, in the following way: For each $x \in \{-1,1\}^n$,

$$\mathsf{THR}_{\mathbf{w}}(x) = \operatorname{sgn}\left(w_0 + \sum_{i=1}^{n} w_i x_i\right).$$

It is well known [49] that for every threshold function with n inputs, there exists a threshold representation for it that uses only integer weights of magnitude at most $2^{O(n \log n)}$. From now on, we only consider threshold representations of threshold functions that use integer weights. The power of an LTF depends on the magnitude of the weights allowed. For instance, the Boolean function $\mathrm{GT}(x, y)$ that determines if the n-bit integer x is at least as large as the n-bit integer y is an example of an LTF that has no representation as an LTF with subexponentially small weights. Indeed in various areas, several questions and problems have been solved when the LTFs arising in the study are restricted to having small weights, but extending them to unrestricted weights are either open or have been solved after spending much research effort. Examples of such areas are learning theory [41, 60], pseudorandom generators [57], analysis of Boolean functions [33], and Boolean circuit complexity [21]. Understanding the relative power of large weights vs. small weights in the context of small-depth circuits having LTFs as gates has attracted attention in several works [5, 25, 62, 31, 32, 53, 30, 36, 26].

In this section, we describe the applicability of our main theorem in answering a longstanding open question that completes the picture of the role weights play in depth-2 threshold circuits. The class of all Boolean functions that can be computed by circuits of depth d and polynomial size, comprising gates computing LTFs (of polynomially bounded weights), is denoted by LT_d ($\widehat{LT}_d$). Indeed, the class of functions computable by constant-depth threshold circuits, commonly denoted TC^0, has received wide attention lately in various areas [18, 20, 63, 19], including those mentioned earlier in this section. In the context of proving explicit lower bounds against small-depth threshold circuits, the seminal work of Minsky and Papert [46] showed that a simple function, Parity, is not in LT_1. While it is not very hard to verify that Parity is in $\widehat{LT}_2$, an outstanding problem is to exhibit an explicit function that is not in LT_2. This problem is now a well-identified frontier for research in circuit complexity.

By contrast, the relatively early work of Hajnal et al. [30] established the fact that the Inner-Product modulo 2 function (denoted by IP), which is easily seen to be in $\widehat{LT}_3$, is not in $\widehat{LT}_2$. It turns out that there is a natural class sitting between $\widehat{LT}_2$ and LT_2, denoted by $\mathsf{THR} \circ \mathsf{MAJ}$, where the top THR gate has unrestricted weights, but the weights of the bottom MAJ gates are restricted to being only polynomially large. Goldmann, Håstad, and Razborov [25] proved several interesting results, which implied the following structure:

$$\widehat{LT}_2 \overset{[25]}{=} \mathsf{MAJ} \circ \mathsf{THR} \overset{[25]}{\subsetneq} \mathsf{THR} \circ \mathsf{MAJ} \subseteq LT_2 \overset{[25]}{\subseteq} \widehat{LT}_3.$$

In a breakthrough work, Forster [23] showed that IP has sign-rank $2^{\Omega(n)}$ for the natural partition of input variables. Forster et al. [24] observed that sign-rank lower bounds for a function f imply lower bounds on the size of $\mathsf{THR} \circ \mathsf{MAJ}$ circuits computing f.

This yielded an exponential separation between $\mathsf{THR} \circ \mathsf{MAJ}$ and $\widehat{LT}_3$. This meant that at least one of the two containments $\mathsf{THR} \circ \mathsf{MAJ} \subseteq LT_2$ and $LT_2 \subseteq \widehat{LT}_3$ is strict. Which of these containments is strict has intrigued researchers. For instance, Pudlàk in private communication [52] conjectured that $\mathsf{THR} \circ \mathsf{MAJ} = LT_2$. This was motivated by the fact that Goldmann and co-authors showed that in the related setting of the top gate being weight-restricted, weights at the bottom gates do not give more power, i.e., $\mathsf{MAJ} \circ \mathsf{MAJ} = \mathsf{MAJ} \circ \mathsf{THR}$. Alman and Williams [2] recently showed interesting upper bounds on the "probabilistic sign-rank" for functions in LT_2. In contrast, Amano and Maruoka [5] and Hansen and Podolskii [31] state that *separating* $\mathsf{THR} \circ \mathsf{MAJ}$ from $\mathsf{THR} \circ \mathsf{THR} = LT_2$ would be an important step for shedding more light on the structure of depth-2 Boolean circuits. However, as far as we know, there was no clear target function identified for the purpose of separating the two classes. No progress on this question was made until our work.

We show that indeed $\mathsf{THR} \circ \mathsf{MAJ} \subsetneq \mathsf{THR} \circ \mathsf{THR}$ and the function F_n achieves the desired separation. To see why it does, we first note that F_n can be conveniently expressed as a composed function in the following way: consider a simple adaptation of the well-known $\mathsf{ODD\text{-}MAX\text{-}BIT}$ function, which we denote by OMB_ℓ^0. The function OMB_ℓ^0 outputs -1 precisely if the rightmost bit that is set to 1 occurs at an odd index (in particular, OMB_ℓ^0 outputs 1 if there is no such index). It is simple to observe that it is a linear threshold function:

$$\mathsf{OMB}_\ell^0(x) = -1 \iff \sum_{i=1}^{\ell} (-1)^{i+1} 2^i (1 + x_i) \geq 0.5.$$

For functions $f_m : \{-1,1\}^m \to \{-1,1\}$ and $g_n : \{-1,1\}^n \to \{-1,1\}$,[2] let $f_m \circ g_n : \{-1,1\}^{m \times n} \to \{-1,1\}$ denote the composed function on mn input bits, where each of the m input bits to the outer function f_m is obtained by applying the inner function g_n to a block of n bits. It is not hard to verify that $F_n = \mathsf{OMB}_\ell^0 \circ \mathsf{OR}_{\ell^{1/3}+\log \ell} \circ \mathsf{XOR}_2$.

We first observe that F_n can be computed by linear-size $\mathsf{THR} \circ \mathsf{THR}$ formulas. For each $x \in \{-1,1\}^n$, let $\mathsf{ETHR}_{\mathbf{w}}(x) = -1 \iff w_0 + w_1 x_1 + \cdots + w_n x_n = 0$. Thus, ETHR gates are also called exact threshold gates. By first observing that every function computed by a formula of the form $\mathsf{THR} \circ \mathsf{OR}$ can also be computed by a formula of the form $\mathsf{THR} \circ \mathsf{AND}$ with a linear blow-up in size, it follows that F_n can be computed by linear-size formulas of the form $\mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}_2$. Note that each $\mathsf{AND} \circ \mathsf{XOR}_2$ is computable by an ETHR gate. Hence, F_n is computable by a linear-size $\mathsf{THR} \circ \mathsf{ETHR}$ formula. A result of Hansen and Podolskii [31] shows that every linear-size $\mathsf{THR} \circ \mathsf{ETHR}$ formula can be simulated by a linear-size $\mathsf{THR} \circ \mathsf{THR}$ formula. Thus, along with the fact that sign-rank lower bounds yield lower bounds on $\mathsf{THR} \circ \mathsf{MAJ}$ circuit size [24], our main theorem (Theorem 1.1) and the above observation yield the following circuit class separation.

THEOREM 1.2. *The function F_n can be computed by linear-size $\mathsf{THR} \circ \mathsf{THR}$ formulas, but any $\mathsf{THR} \circ \mathsf{MAJ}$ circuit computing it requires size $2^{\Omega(n^{1/4})}$.*

The message of Theorem 1.2 may be contrasted with previous knowledge as follows: *While weights at the bottom do not matter if the top is light, they do matter if the top is heavy.* Further, Theorem 1.2 provides the first explanation for why current lower bound methods fail to get traction with $\mathsf{THR} \circ \mathsf{THR}$. Interestingly, it also suggests some new paths along which progress seems feasible. This is discussed in section 8.

[2] We sometimes drop the subscript when the arity of the underlying function is clear from context.

1.2. Application: Communication complexity. Göös [27] pointed out that F_n can be used to demonstrate another complexity class separation, this time in communication complexity. Complexity classes in communication complexity were first introduced in the seminal work of Babai, Frankl, and Simon [6] as an analogue to the standard Turing complexity classes. While *unconditionally* understanding the relative power of (non)determinism and randomness in the context of Turing machines seems well beyond current techniques, Babai and co-authors had hoped that making progress in the miniworld of communication protocols would be less difficult. Indeed, we understand a lot more in the latter world. For instance, the class P^{cc} is *strictly* contained in both BPP^{cc} and NP^{cc}, while BPP^{cc} and NP^{cc} are provably different. A major goal, set by Babai and co-authors, is to prove lower bounds against the polynomial hierarchy for which the simple function IP has long been identified as a target. Unfortunately, it even remains open to exhibit a function that is not in the second level of the hierarchy. Our result explains this lack of progress by showing that a total function, conceivably well below the second level, has large sign-rank.

Henceforth, we often drop cc from the superscript for convenience since we deal exclusively with communication complexity classes. The strongest lower bound technique currently known in communication complexity is the sign-rank method, as discussed earlier in section 1. Functions whose communication matrix of dimension $2^n \times 2^n$ have sign-rank upper bounded by a quasi-polynomial in n were shown in [51] to correspond exactly to the complexity class UPP. The lower bound on sign-rank by Razborov and Sherstov [54] implied that PH (in fact, $\Pi_2\mathsf{P}$) contains functions with large sign-rank, rendering the sign-rank technique essentially useless to prove lower bounds against the second level. A natural question is to understand until where, between the first and second level, does the sign-rank method suffice to prove lower bounds.

Indeed, there is a rich landscape of communication complexity classes below the second level as discussed in a recent, almost exhaustive survey by Göös, Pitassi, and Watson [29]. To motivate our contributions, we informally define MA protocols.[3] Merlin, an all powerful prover, has access to Alice and Bob's inputs. He sends a (purported) proof string to Alice and Bob, who then run a randomized protocol to verify the proof. The protocol accepts an input iff the verification goes through. We say the protocol computes a function F if for all inputs to Alice and Bob, the probability of outputting the correct answer is at least $2/3$. The cost of the protocol on an input is the sum of the length of Merlin's proof string and the number of bits communicated between Alice and Bob. A function is said to be in the complexity class MA if there is such a protocol computing it with polylogarithmic worst-case cost (in the size of the input). For example, the function $\mathsf{OR} \circ \mathsf{EQ}$ can be seen to be in MA as follows: Merlin sends Alice and Bob the index of an input to the OR gate (if it exists) where EQ outputs -1, and Alice and Bob run an efficient randomized protocol for EQ to verify this. The class MA is a natural generalization of NP and has received a lot of attention, starting with the work of [40]. It is known that MA is strictly contained in UPP.

One could similarly define AM, but its power remains much less understood. Göös, Pitassi, and Watson [29] conjectured that the (potentially incomparable) classes $\mathsf{AM} \cap \mathsf{coAM}$ and $\mathsf{S_2P}$ contain functions of large sign-rank. In a very recent work, Bouland et al. [9] showed that there is a *partial function* in $\mathsf{AM} \cap \mathsf{coAM}$ which has large

[3] The definition of UPP can be found in section 2. A formal description of all other communication complexity classes defined in this section can be found in the appendix.

sign-rank, (partially) resolving the first conjecture.[4] We provide a strong confirmation of the second conjecture.

In order to state our result, let us consider the complexity class P^{MA} that is contained in $\mathsf{S}_2\mathsf{P}$. A function is in P^{MA} if it can be computed by deterministic protocols of polylogarithmic cost, where Alice and Bob have oracle access to any function in MA. The function F_n under the natural input partition (recall that it can be expressed as a decision list of equalities) can be efficiently solved by P^{MA} protocols by an appropriate binary search, and querying an $\mathsf{OR} \circ \mathsf{EQ}$ oracle at each step. A formal description of this protocol is given in Algorithm 6.1.

We thus prove the following as a consequence of our main theorem.

THEOREM 1.3. *The function F_n witnesses the following communication complexity class separation:*

$$\mathsf{P}^{\mathsf{MA}} \not\subseteq \mathsf{UPP}.$$

Our result thus strongly confirms the second conjecture of Göös and co-authors by exhibiting the first *total function* in a complexity class contained, plausibly strictly, in Π_2P, that has large sign-rank.

On the other hand, it is known that $\mathsf{P}^{\mathsf{NP}} \subsetneq \mathsf{UPP}$ and $\mathsf{MA} \subsetneq \mathsf{PP} \subsetneq \mathsf{UPP}$. These facts combined with Theorem 1.3 show that P^{MA} is right on the frontier between what we understand and what we do not. Thus, proving lower bounds against P^{MA} protocols emerges as a natural program for advancing the set of currently known techniques, given our work. Future directions are further discussed in section 8.

1.3. Related work. Long after Forster [23] showed that upper bounding the spectral norm of a $\{-1, 1\}$ valued matrix suffices to show sign-rank lower bounds, Sherstov [59] introduced an innovative method that designed a passage to a suitable approximation problem via LP duality. This basic framework was again used by Razborov and Sherstov [54], developing more approximation-theoretic tools, to prove the first exponential lower bounds of $2^{\Omega(n^{1/3})}$ on the sign-rank of a function in AC^0. This function can be computed by a *depth*-3 *linear-size circuit*. After a series of works improving this bound [12, 14], Sherstov and Wu [61] recently proved a near-optimal $2^{\Omega(n^{1-\epsilon})}$ lower bound on the sign-rank of a function computable in AC^0. Bouland et al. [9] recently proved strong sign-rank bounds for a *partial function* with interesting applications. All of these works [59, 54, 12, 9, 14, 61] rely on the passage, invented by Sherstov [59], to an approximation-theoretic problem involving *low degree* polynomials. This passage is made possible by exploiting the elegant spectral properties of communication matrices of the target functions, following the basic pattern matrix method of Sherstov [58].

Unfortunately, it seems difficult to embed a pattern matrix in a function in $\mathsf{THR} \circ \mathsf{THR}$. Consequently, we come up with a different type of function, F_n, that is an *XOR function*. Proving lower bounds on the communication complexity of XOR functions, in general, has received a lot of attention recently [48, 65, 64, 34, 39, 17]. However, there seem to be just two previous works that prove a lower bound on the sign-rank of an XOR function, due to Hatami and Qian [35] and independently by Ada, Fawzi, and Kulkarni [1]. Their result characterizes the sign-rank of functions of the form $f \circ \mathsf{XOR}$ when f is symmetric. In contrast, our target function F_n is not a symmetric XOR function. Moreover, both the works [35] and [1] obtain their result using neat reductions from pattern matrices of symmetric functions, which had been

[4] It still remains unknown if there are *total* functions in $\mathsf{AM} \cap \mathsf{coAM}$ that have large sign-rank.

analyzed by Sherstov [59]. Such a reduction for a function in $\mathsf{THR} \circ \mathsf{THR}$ is unknown, and plausibly impossible. This forces us to use a first-principle based argument for bounding the sign-rank of an XOR function. Such functions also have nice spectral properties that are, however, different from those of pattern matrices. More specifically, the approximation-theoretic problem that one is led to in this case involves polynomials with *unrestricted* degree but low Fourier weight. A similar flavored but simpler problem had been tackled in a recent work of the authors [15], which characterized the *discrepancy* of XOR functions. Roughly speaking, in that work, the primal program asked for a distribution μ such that f correlates poorly with all parities w.r.t. μ. However, there was no smoothness constraint imposed on μ in [15], which is what we are constrained to have in this work. Analyzing this combination of high degree parity constraints and the smoothness constraints is the main new technical challenge that our work addresses.

It is simple to verify that F_n is computed by a linear-size AC^0 circuit. Theorem 1.1 therefore yields a new argument to show that AC^0 has large sign-rank. While our bounds on the sign-rank of F_n are weaker than that of [54, 12, 14, 61], F_n is simpler than the earlier functions in other ways. It is just a decision list of "Equalities" that is, therefore, both in the Boolean circuit class $\mathsf{THR} \circ \mathsf{THR}$ and the communication complexity class P^{MA}. It is precisely this property of F_n that allows us to simultaneously answer two open questions.

1.4. Our techniques. We strive to prove a lower bound on the sign-rank of a function $F \in \mathsf{THR} \circ \mathsf{THR}$. We are guided by a communication complexity theoretic interpretation of sign-rank, due to Paturi and Simon [51]. Paturi and Simon introduced a model of two-party randomized communication, called the unbounded-error model. In this model, Alice and Bob are only required to give the right answer with probability strictly greater than $1/2$ on every input. This is the strongest known two-party model against which we know how to prove lower bounds. Paturi and Simon [51] showed that the sign-rank of the communication matrix of F essentially characterizes its unbounded-error communication complexity.

Why should some function $F \in \mathsf{THR} \circ \mathsf{THR}$ have large unbounded-error communication complexity? A natural protocol one is tempted to use is the following. Sample a subcircuit of the top gate with a probability proportional to its weight. Then use the best protocol for the sampled bottom THR gate. Note that for any given input x, with probability $1/2 + \epsilon$, one samples a bottom gate that agrees with the value of $F(x)$. Here, ϵ can be inverse exponentially small in the input size. Thus, if we had a small cost randomized protocol for the bottom THR gate that errs with probability significantly less than ϵ we would have a small cost unbounded-error protocol for F. Fortunately for us (the lower bound prover), there does not seem to exist any such efficient randomized protocol for THR, when $\epsilon = 1/2^{n^{\Omega(1)}}$.

Taking this a step further, one could hope that the bottom gates could be any function that is hard to compute with such tiny error ϵ. The simplest such canonical function is Equality (denoted by EQ). Therefore, a plausible target is $\mathsf{THR} \circ \mathsf{EQ}$. This still turns out to be in $\mathsf{THR} \circ \mathsf{THR}$ as $\mathsf{EQ} \in \mathsf{ETHR}$. Moreover, EQ has a nice composed structure. It is just $\mathsf{AND} \circ \mathsf{XOR}$, which lets us reexpress our target as $F = \mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}$, for some top THR that is "suitably" hard. At this point, we view F as an XOR function whose outer function, f, needs to have sufficiently good analytic properties for us to prove that $f \circ \mathsf{XOR}$ has high sign-rank.

We are naturally drawn to the work of Razborov and Sherstov [54] for inspiration as they bound the sign-rank of a three-level composed function as well. They showed

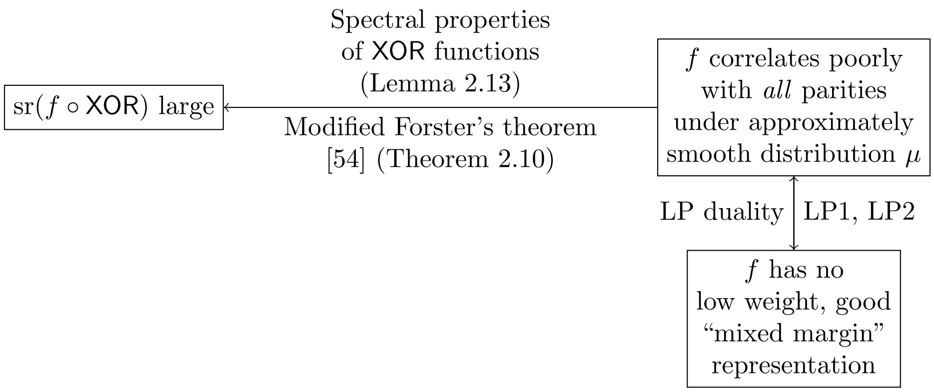

FIG. 1. *Approximation-theoretic hardness of f implies large sign-rank of $f \circ \mathsf{XOR}$ (Theorem* 3.1*).*

that $\mathsf{AND} \circ \mathsf{OR} \circ \mathsf{AND}_2$ has high sign-rank. They exploited the fact that this function embeds a *pattern matrix* inside it, which has nice convenient spectral properties as observed in [58]. These spectral properties dictate them to look for an *approximately smooth orthogonalizing* distribution w.r.t. which the outer function $f = \mathsf{AND} \circ \mathsf{OR}$ has zero correlation with small degree parities. This naturally gives rise to a linear program that seeks to maximize the *smoothness* of the distribution under the constraints of low-degree orthogonality. The main technical challenge that Razborov and Sherstov overcome is the analysis of the dual of this LP using and building appropriate approximation-theoretic tools. We follow this general framework of analyzing the dual of a suitable LP. However, as we are forced to work with an XOR function, there are new challenges that crop up. This is understandable, for if we take the same outer function of $\mathsf{AND} \circ \mathsf{OR}$, then the resulting XOR function has small sign-rank. Indeed, this remains true even if one were to harden the outer function to $\mathsf{MAJ} \circ \mathsf{OR}$. This is simply because $\mathsf{OR} \circ \mathsf{XOR}$ is nonequality (NEQ). A simple efficient UPP protocol for $\mathsf{MAJ} \circ \mathsf{NEQ}$ exists: pick a random NEQ and then execute a protocol of cost $O(\log n)$ that solves this NEQ with error less than $1/n^2$.

Figure 1 describes a general passage from the problem of lower bounding the sign-rank of a function $f \circ \mathsf{XOR}$ to a sufficient problem of proving an approximation-theoretic hardness property of f, namely, f has no good "mixed margin" representation by low weight polynomials. Theorem 3.1 states the precise connection between the approximation-theoretic property of f and the sign-rank of $f \circ \mathsf{XOR}$. This passage is made possible by using well known spectral properties of XOR functions and LP duality. This is similar to earlier works [54, 59, 12, 9, 14, 61], where the spectral properties of pattern matrices were analyzed. The key difference between our work and theirs is in the nature of the approximation-theoretic problem that we end up with. While all these previous works had to rule out good *low degree* representations, our Theorem 3.1 stipulates us to rule out good *low weight* representations of otherwise *unrestricted degree*.

Our main technical contribution is Theorem 4.1, which shows that the function $\mathsf{OMB}^0 \circ \mathsf{OR}$ is inapproximable by low weight polynomials of *unrestricted degree*, in a sense which we elaborate on below. We prove this by a novel combination of ideas, sketched in Figure 2, that differs entirely from analysis in earlier works. One may view this result as a hardness amplification result, albeit specific to the function OMB^0. We start with the function OMB^0, which has no low weight "worst case margin"

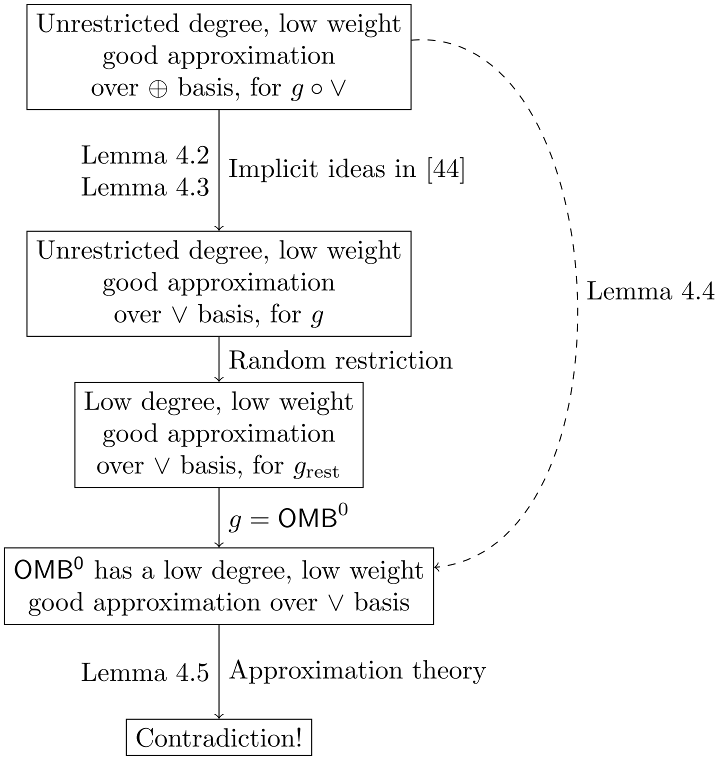

FIG. 2. *Approximation-theoretic analysis (Theorem* 4.1*).*

representation when the degree of the approximating polynomial is bounded [7]. We show that on composition with large fan-in OR gates, the function $\mathsf{OMB}^0 \circ \mathsf{OR}$ becomes "*mixed margin*"-inapproximable by low weight polynomials, even with *unrestricted degree.* We believe this result to be of independent interest in the area of analysis of Boolean functions and approximation theory.

The first step in our method is to borrow an averaging idea from Krause and Pudlák [44] to show the following: a low weight good approximation of $g \circ \mathsf{OR}_m$ by a polynomial p over the parity (Fourier) basis implies that there exists a low weight polynomial q over the OR basis which approximates g as well as p approximates $g \circ \mathsf{OR}_m$, save an additive loss of at most 2^{-m}. This transformation to q is very useful because although it is still unrestricted in degree, it is over the OR basis, which is vulnerable to random restrictions. Indeed, in the next step, we hit q with random restrictions to reduce its degree. At this point, we extract a low weight *and* low degree polynomial r that still approximates g_{rest}, the restriction of g. We now appeal to interesting properties of the $\mathsf{ODD\text{-}MAX\text{-}BIT}$ function by setting $g = \mathsf{OMB}^0$. First, we observe that OMB^0 on ℓ bits, under random restrictions, retains its hardness as it contains OMB^0 on $\ell/8$ bits with high probability. Next, we show that OMB^0 does not have low degree good approximations by appealing to classical approximation-theoretic tools, suitably modifying the arguments of Buhrman, Vereshchagin, and de Wolf [11] and Beigel [7]. This provides us with the required contradiction.

2. Preliminaries.

In this section, we provide some necessary preliminaries.

2.1. Notation.

All logarithms in this paper are taken base 2. For a positive integer n and a set $X \subseteq \{-1, 1\}^n$, let X^c denote the complement of X in $\{-1, 1\}^n$.

For positive integers $m, \ell > 0$ and a string $x = (x_1, \ldots, x_\ell) \in \{-1,1\}^{m\ell}$, let $\bigvee(x) = (\bigvee_m(x_1), \ldots, \bigvee_m(x_\ell)) \in \{-1,1\}^\ell$. Here $\bigvee_m$ denotes the OR function on m input bits, which outputs -1 if there exists an index where the input is -1, and outputs 1 otherwise. We drop the subscripts when m, ℓ are clear from context.

We require the following form of the Chernoff bound (see, for example, [47, Theorem 4.5]).

LEMMA 2.1 (Chernoff bound). *Let $X_1, X_2, \ldots, X_n$ be independent $\{0,1\}$-valued random variables. Let $X = \sum_{i=1}^n X_i$ denote their sum, and let $\mu = \mathbb{E}(X)$. Then for all $0 < \delta < 1$,*

$$\Pr[X \le (1-\delta)\mu] \le e^{-\mu\delta^2/2}.$$

DEFINITION 2.2 (OR polynomials). *Define a function $p : \{-1,1\}^n \to \mathbb{R}$ of the form $p(x) = \sum_{S\subseteq[n]} a_S \bigvee_{i\in S} x_i$ to be an* OR polynomial. *Define the weight of p (in the OR basis) to be $\sum_{S\subseteq[n]} |a_S|$ and its degree to be $\max_{S\subseteq[n]}\{|S| : a_S \neq 0\}$.*

Remark 2.3. In the above definition, "OR monomials" are defined as follows:

$$\bigvee_{i\in S} x_i = \begin{cases} 1 & x_i = 1 \ \forall i \in S, \\ -1 & \text{otherwise.} \end{cases}$$

Unless mentioned otherwise, all polynomials we consider will be over the *parity* basis.

DEFINITION 2.4 (decision lists). *A decision list of length k is a sequence $D = (L_1, a_1), (L_2, a_2), \ldots, (L_k, a_k)$, where each $a_i \in \{-1,1\}$, and L_k is the constant -1 function. The decision list computes a function $f : \{-1,1\}^n \to \{-1,1\}$ as follows. If $L_1(x) = -1$, then $f(x) = a_1$; else if $L_2(x) = -1$, then $f(x) = a_2$; else if $\ldots$, else if $L_k(x) = -1$, then $f(x) = a_k$. That is,*

$$f(x) = \bigvee_{i=1}^{k} \left(a_i \bigwedge L_i(x) \bigwedge_{j<i} \neg L_j(x) \right).$$

2.2. Threshold circuits. We permit negated variables as inputs in all circuit classes under consideration in this paper.

DEFINITION 2.5 (threshold functions). *A function $f : \{-1,1\}^n \to \{-1,1\}$ is called a* linear threshold function *if there exist integer weights $a_0, a_1, \ldots, a_n$ such that for all inputs $x \in \{-1,1\}^n$, $f(x) = \mathrm{sgn}(a_0 + \sum_{i=1}^n a_i x_i)$. Let THR denote the class of all such functions.*

DEFINITION 2.6 (exact threshold functions). *A function $f : \{-1,1\}^n \to \{-1,1\}$ is called an* exact threshold function *if there exist reals $w_1, \ldots, w_n, t$ such that*

$$f(x) = -1 \iff \sum_{i=1}^n w_i x_i = t.$$

Let ETHR denote the class of exact threshold functions.

Hansen and Podolskii [31] showed the following.

THEOREM 2.7 (Hansen and Podolskii [31]). *If a function $f : \{-1,1\}^n \to \{-1,1\}$ can be represented by a $\mathsf{THR} \circ \mathsf{ETHR}$ formula of size s, then it can be represented by a $\mathsf{THR} \circ \mathsf{THR}$ formula of size $2s$.*

For the sake of completeness and clarity, we provide the proof below.

Proof. Let h be an exact threshold function with the representation $\sum_{i=1}^{n} w_i x_i = t$. There exists an $\epsilon_h > 0$ such that $\sum_{i=1}^{n} w_i x_i > t \implies \sum_{i=1}^{n} w_i x_i > t + \epsilon_h$. Consider a $\mathsf{THR}\circ\mathsf{ETHR}$ formula of size s which computes f. Say it computes $\mathrm{sgn}(c_0 + \sum_{i=1}^{s} c_i f_i)$, where the f_i's have exact threshold representations $\sum_{j=1}^{n} w_{i,j} x_j = t_i$, respectively. Consider the $\mathsf{THR} \circ \mathsf{THR}$ formula of size $2s$, given by $\mathrm{sgn}\left(\sum_{i=1}^{s} c_i (g_{i,1} - g_{i,2} + 1)\right)$, where the g_i's are threshold functions with representations as follows:

$$g_{i,1} = 1 \iff \sum_{j=1}^{n} w_{i,j} x_j \geq t_i,$$

$$g_{i,2} = 1 \iff \sum_{j=1}^{n} w_{i,j} x_j \geq t_i + \epsilon_{f_i}.$$

It is easy to verify that this formula computes f. ◻

Remark 2.8. In fact, Hansen and Podolskii [31] showed that the circuit class $\mathsf{THR} \circ \mathsf{THR}$ is identical to the circuit class $\mathsf{THR} \circ \mathsf{ETHR}$. However, we do not require the full generality of their result.

We now note that any function computable by a $\mathsf{THR} \circ \mathsf{OR}$ formula can be computed by a $\mathsf{THR} \circ \mathsf{AND}$ formula without a blowup in the size.

LEMMA 2.9. *Suppose $f : \{-1,1\}^n \to \{-1,1\}$ can be computed by a $\mathsf{THR} \circ \mathsf{OR}$ formula of size s. Then, f can be computed by a $\mathsf{THR} \circ \mathsf{AND}$ formula of size s.*

Proof. Consider a $\mathsf{THR} \circ \mathsf{OR}$ formula of size s, computing f, say,

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{s} w_i \bigvee_{j \in S_i} x_j\right).$$

Note that

$$\sum_{i=1}^{s} w_i \bigvee_{j \in S_i} x_j = \sum_{i=1}^{s} -w_i \bigwedge_{j \in S_i} x_j^c.$$

Thus, $\mathrm{sgn}\left(\sum_{i=1}^{s} -w_i \bigwedge_{j \in S_i} x_j^c\right)$ is a $\mathsf{THR} \circ \mathsf{AND}$ representation of f, of size s. ◻

2.3. Sign-rank. Define the *sign-rank* of a real valued matrix $A = [A_{ij}]$, denoted by $\mathrm{sr}(A)$, to be the least rank of a real matrix $B = [B_{ij}]$ such that $A_{ij} B_{ij} > 0$ for all (i,j) such that $A_{ij} \neq 0$.

We require the following generalization of Forster's theorem [23] by Razborov and Sherstov [54].

THEOREM 2.10 (Razborov and Sherstov [54]). *Let $A = [A_{xy}]_{x \in X, y \in Y}$ be a real valued matrix with $s = |X||Y|$ entries, such that $A \neq 0$. For arbitrary parameters $h, \gamma > 0$, if all but h of the entries of A satisfy $|A_{xy}| \geq \gamma$, then*

$$\mathrm{sr}(A) \geq \frac{\gamma s}{||A||\sqrt{s} + \gamma h}.$$

Forster et al. [24] showed that functions with efficient $\mathsf{THR} \circ \mathsf{MAJ}$ representations have small sign-rank.

LEMMA 2.11 (Forster et al. [24]). *Let $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ be a Boolean function computed by a $\mathsf{THR} \circ \mathsf{MAJ}$ circuit of size s. Then,*

$$\mathrm{sr}(M_F) \leq sn,$$

where M_F denotes the communication matrix of F.

For the purposes of this paper, we abuse notation and use $\mathrm{sr}(F)$ and $\mathrm{sr}(M_F)$ interchangeably to denote the sign-rank of M_F.

2.4. Fourier analysis. Consider the vector space of functions from $\{-1,1\}^n$ to $\mathbb{R}$, equipped with the following inner product:

$$\langle f, g \rangle = \mathbb{E}_{x \in \{-1,1\}^n} [f(x)g(x)] = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)g(x).$$

Define "characters" χ_S for every $S \subseteq [n]$ by $\chi_S(x) = \prod_{i \in S} x_i$. The set $\{\chi_S : S \subseteq [n]\}$ forms an orthonormal basis for this vector space. Thus, every $f : \{-1,1\}^n \to \mathbb{R}$ can be uniquely written as $f = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S$, where

$$\widehat{f}(S) = \langle f, \chi_S \rangle = \mathbb{E}_{x \in \{-1,1\}^n} [f(x)\chi_S(x)] . \tag{2.1}$$

For a polynomial $p = \sum_{S \subseteq [n]} c_S \chi_S$, define

$$\mathrm{mon}(p) = |S \subseteq [n] : c_S \neq 0|.$$

Also define the *weight* of p by

$$\mathrm{wt}(p) = \sum_{S \subseteq [n]} |c_S|.$$

Since every function $f : \{-1,1\}^n \to \mathbb{R}$ has a unique polynomial exactly representing it, we abuse notation and use $\mathrm{mon}(f)$ and $\mathrm{wt}(f)$ to denote the corresponding quantities for this polynomial.

We require the following well-known identity (see, for example, [50]).

FACT 2.12 (Plancherel's identity). *For any functions $f, g : \{-1,1\}^n \to \mathbb{R}$,*

$$\mathbb{E}_{x \in \{-1,1\}^n} [f(x)g(x)] = \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{g}(S).$$

The following lemma characterizes the spectral norm of the communication matrix of XOR functions (see, for example, [8]).

LEMMA 2.13 (folklore). *Let $f : \{-1,1\}^n \to \mathbb{R}$ be any real valued function and let M denote the communication matrix of $f \circ \mathsf{XOR}$. Then,*

$$||M|| = 2^n \cdot \max_{S \subseteq [n]} \left|\widehat{f}(S)\right|.$$

2.5. Polynomial approximations. The following is a well-known lemma by Minsky and Papert [46].

LEMMA 2.14 (Minsky and Papert [46]). *Let $p : \{-1,1\}^n \to \mathbb{R}$ be any symmetric real polynomial of degree d. Then, there exists a univariate polynomial q of degree at most d, such that for all $x \in \{-1,1\}^n$,*

$$p(x) = q(\#1(x)),$$

where $\#1(x) = |\{i \in [n] : x_i = 1\}|$.

We require the following approximation-theoretic lemma by Ehlich and Zeller [22] and Rivlin and Cheney [56].

LEMMA 2.15 ([22, 56]). *The following holds true for any real valued $\alpha > 0$ and integer $k > 0$. Let $p : \mathbb{R} \to \mathbb{R}$ be a univariate polynomial of degree $d < \sqrt{k/4}$, such that $p(0) \geq \alpha$, and $p(i) \leq 0$ for all $i \in [k]$. Then, there exists $i \in [k]$ such that $p(i) < -2\alpha$.*

2.6. Communication complexity. In the model of communication we consider, two players, say, Alice and Bob, are given inputs $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ for some finite input sets $\mathcal{X}, \mathcal{Y}$. They are given access to *private* randomness and wish to compute a given function $F : \mathcal{X} \times \mathcal{Y} \to \{-1,1\}$. We will use $\mathcal{X} = \mathcal{Y} = \{-1,1\}^n$ for the purposes of this paper. Alice and Bob communicate using a randomized protocol which has been agreed upon in advance. The cost of the protocol is the maximum number of bits communicated in the worst case input and coin toss outcomes. A protocol Π computes F with advantage ϵ if the probability of F agreeing with Π is at least $1/2+\epsilon$ for all inputs. We denote the cost of the best such protocol to be $R_\epsilon(F)$. Note here that we deviate from standard notation (used in [45], for example, where the subscript denotes the error probability rather than advantage). Unbounded-error communication complexity was introduced by Paturi and Simon [51] and is defined as follows:

$$\mathsf{UPP}(F) = \inf_{\epsilon>0}(R_\epsilon(F)).$$

This measure gives rise to the natural communication complexity class UPP^{cc}, defined as $\mathsf{UPP}^{cc}(F) \equiv \{F : \mathsf{UPP}(F) = \mathrm{polylog}(n)\}$.

Paturi and Simon [51] showed an equivalence between $\mathsf{UPP}(F)$ and the sign-rank of M_F.

THEOREM 2.16 (Paturi and Simon [51]). *For any function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$,*

$$\mathsf{UPP}(F) = \log \mathrm{sr}(M_F) \pm O(1).$$

3. Sign-rank to polynomial approximation. In this section, we prove how a certain approximation-theoretic hardness property of f implies that the sign-rank of $f \circ \mathsf{XOR}$ is large, as outlined in Figure 1.

Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function, $\delta > 0$ be a parameter, and X be any subset of $\{-1,1\}^n$. We consider the following linear program, which has exactly the same structure as in (LP1) in [59] except for one crucial difference, which is described below:

$$\text{(LP1)}\quad \begin{array}{|llll|} \hline \text{Variables} & \epsilon, \{\mu(x) : x \in \{-1,1\}^n\} & & \\ \text{Minimize} & \epsilon & & \\ \text{s.t.} & \left|\sum_x \mu(x) f(x) \chi_S(x)\right| & \leq \epsilon & \forall S \subseteq [n] \\ & \sum_x \mu(x) & = 1 & \\ & \epsilon \geq 0 & & \\ & \mu(x) \geq \frac{\delta}{2^n} & & \forall x \in X \\ & \mu(x) \geq 0 & & \forall x \in \{-1,1\}^n \\ \hline \end{array}$$

The first constraint in (LP1) specifies that the correlation of f against *all parities* needs to be small w.r.t. a distribution μ. Note that in [59], this constraint was only imposed for *low degree* parities. This difference between the two linear programs forces us to entirely change the analysis of the dual from the one in [59]. As discussed earlier in subsection 1.4, this analysis is one of our main technical innovations. The second-to-last constraint enforces the fact that μ is "δ-smooth" over the set X. As we had indicated earlier in subsection 1.4, these constraints make analyzing the LP challenging.

Standard manipulations (as in [15], for example) and strong linear programming duality reveal that the optimum of (LP1) equals the optimum of (LP2). Let OPT denote the optima of these programs.

$$\text{(LP2)}\quad \begin{array}{|llll|} \hline \text{Variables} & \Delta, \{\alpha_S : S \subseteq [n]\}, \{\xi_x : x \in X\} & & \\ \text{Maximize} & \Delta + \frac{\delta}{2^n} \sum_{x \in X} \xi_x & & \\ \text{s.t.} & f(x) \sum_{S \subseteq [n]} \alpha_S \chi_S(x) & \geq \Delta & \forall x \in \{-1,1\}^n \\ & f(x) \sum_{S \subseteq [n]} \alpha_S \chi_S(x) & \geq \Delta + \xi_x & \forall x \in X \\ & \sum_{S \subseteq [n]} |\alpha_S| & \leq 1 & \\ & \Delta \in \mathbb{R} & & \\ & \alpha_S \in \mathbb{R} & & \forall S \subseteq [n] \\ & \xi_x \geq 0 & & \forall x \in X \\ \hline \end{array}$$

The objective function of (LP2) consists of a "worst-case component" Δ and an "average-case component" $\frac{\delta}{2^n} \sum_{x \in X} \xi_x$. The first set of constraints indicates that a dual polynomial p must satisfy $f(x)p(x) \geq \Delta$ for all $x \in \{-1,1\}^n$ (justifying the terminology "worst-case component").[5] The second set of constraints says that at each point x over the smooth set X, the dual polynomial has to better the worst-case component by at least ξ_x (justifying the terminology "average-case component"). If OPT is large, then it means that on average, the dual polynomial did significantly better than its worst-case component. It is for this reason we call the optimum the "mixed margin"[6] as mentioned in subsection 1.4.

We now show that upper bounding OPT for any function f yields sign-rank lower bounds against $f \circ \mathsf{XOR}$. The proof idea is depicted in Figure 1.

THEOREM 3.1. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be any function. For any $\delta > 0$ and $X \subseteq \{-1,1\}^n$, suppose the value of the optimum of (LP2) (and hence (LP1)) is at most*

[5] Note that Δ might be negative.

[6] The mixed margin is always nonnegative, since a suitably scaled version of the polynomial exactly representing f acts as a feasible dual polynomial.

OPT. *Then,*

$$\mathrm{sr}(f \circ \mathsf{XOR}) \geq \frac{\delta}{\mathsf{OPT} + \delta \cdot \frac{|X^c|}{2^n}}.$$

Proof. By (LP1), there exists a distribution μ on $\{-1,1\}^n$ such that $\mu(x) \geq \frac{\delta}{2^n}$ for all $x \in X$, and $\max_{S \subseteq [n]} |\widehat{f\mu}(S)| \leq \frac{\mathsf{OPT}}{2^n}$. By Lemma 2.13,

$$||M_{f\mu \circ \mathsf{XOR}}|| = 2^n \cdot \max_{S \subseteq [n]} \left|\widehat{f\mu}(S)\right| \leq \mathsf{OPT}.$$

Each $x \in X$ contributes to 2^n entries of $M_{f\mu \circ \mathsf{XOR}}$ whose absolute value is at least δ. Plugging values into Theorem 2.10, we obtain

$$\mathrm{sr}(f \circ \mathsf{XOR}) \geq \mathrm{sr}(f\mu \circ \mathsf{XOR}) \geq \frac{\frac{\delta}{2^n} \cdot 2^{2n}}{\mathsf{OPT} \cdot 2^n + \frac{\delta}{2^n} \cdot 2^n \cdot |X^c|} = \frac{\delta}{\mathsf{OPT} + \delta \cdot \frac{|X^c|}{2^n}},$$

which proves the desired sign-rank lower bound. ∎

The above theorem can also be seen to imply sign-monomial complexity lower bounds, a measure that is more basic than sign-rank. We provide a definition of this measure and a self-contained proof of the resultant lower bounds in Appendix B, which does not involve the use of Forster's theorem.

4. Hardness of approximating $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_m$.

Below is our main technical result, capturing the essence of Figure 2, which says that no dual polynomial exists with a large optimum value for (LP2) when $f = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log \ell} : \{-1,1\}^{\ell^{4/3}+\ell \log \ell} \to \{-1,1\}$, even when the smoothness parameter δ is as high as $1/4$.

THEOREM 4.1 (main technical result). *For a positive integer ℓ, let $m = \ell^{1/3} + \log \ell$. Define $f = \mathsf{OMB}^0_\ell \circ \bigvee_m : \{-1,1\}^{\ell m} \to \{-1,1\}, \delta = 1/4$, and $X = \{x \in \{-1,1\}^{\ell m} : \bigvee(x) = -1^\ell\}$. Then for sufficiently large values of ℓ, the optimal value* OPT *of (LP2) is less than $2^{-\frac{\ell^{1/3}}{81}}$.*

Theorem 4.1 can be viewed as a hardness amplification theorem as follows. Our base function is OMB^0, which is known to be hard to approximate in the worst case by low degree sign representing polynomials [7, 11]. We show that a lifted version of this function, $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_m$, cannot be approximated well under a significantly weaker notion of approximation where we permit any approximating polynomial to have the following additional power:

- unrestricted degree but low weight,
- it need not sign represent $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_m$, but its "mixed margin" is small (see (LP2)).

We prove Theorem 4.1 toward the end of this section. We first outline the various tools that go into proving Theorem 4.1, following the schematic in Figure 2.

We first use an idea from Krause and Pudlák [44] that enables us to work with polynomial approximations for g, given a polynomial approximation for $g \circ \bigvee_m$. We use the following notation for the following two lemmas. For any set $H \subseteq [\ell] \times [m]$, define $I \subseteq [\ell]$ to be the projection of H on $[\ell]$; $i \in I \iff \exists j,\ x_{i,j} \in H$. Note that I depends on H. However, we do not make this dependence explicit to avoid clutter. The set H from which I is obtained will be clear from context. For any $y \in \{-1,1\}^\ell$, let μ_y denote the uniform distribution over all inputs $x \in \{-1,1\}^{m\ell}$ such that $\bigvee_m(x) = y$. Lemmas 4.2 and 4.3 represent the first implication in Figure 2. The first tool we use is an approximation of monomials (in the parity basis) by OR functions, with a small error.

LEMMA 4.2. *Let $\ell \geq 4$ and m be positive integers such that $m > 2\log \ell$. For any set $H \subseteq [\ell] \times [m], y \in \{-1,1\}^\ell$,*

$$\left| \mathbb{E}_{\mu_y}\left[\bigoplus_{(i,j)\in H} x_{i,j} \right] - \frac{1}{2} - \frac{1}{2}\bigvee_{i\in I} y_i \right| \leq 2\ell 2^{-m}.$$

The proof of Lemma 4.2 appears in the proof of Lemma 2.3 in [44]. However, we reproduce the proof below for clarity and completeness.

Proof. First observe that for all $y \in \{-1,1\}^\ell$, and for all x satisfying $\bigvee_m(x) = y$, the monomial corresponding to H equals

$$\bigoplus_{(i,j)\in H} x_{i,j} = \bigoplus_{(i,j)\in H, y_i=-1} x_{i,j}.$$

Let $A = \{i \in [\ell] : y_i = -1\}$. If $A \cap I = \emptyset$, then

$$\mathbb{E}_{\mu_y}\left[\bigoplus_{(i,j)\in H} x_{i,j} \right] = \bigvee_{i\in I} y_i = 1.$$

Else, $\bigvee_{i\in I} y_i = -1$. Also,

$$\mathbb{E}_{\mu_y}\left[\bigoplus_{(i,j)\in H} x_{i,j} \right] = \mathbb{E}_{x\in\{-1,1\}^{(A\cap I)\times[m]}:\bigvee(x)=-1^{|A\cap I|}}\left[\bigoplus_{(i,j)\in H, y_i=-1} x_{i,j} \right]. \tag{4.1}$$

Note that

$$\mathbb{E}_{x\in\{-1,1\}^{(A\cap I)\times[m]}}\left[\bigoplus_{(i,j)\in H, y_i=-1} x_{i,j} \right] = 0. \tag{4.2}$$

Denote $|A \cap I| = t$. Using (4.2) and a simple counting argument, the absolute value of the right-hand side (and thus the left-hand side) of (4.1) can be upper bounded as follows (note that we require $1 \leq t \leq \ell$ in the following computations):

$$\begin{aligned} \left| \mathbb{E}_{\mu_y}\left[\bigoplus_{(i,j)\in H} x_{i,j} \right] \right| &\leq \frac{2^{mt} - (2^m-1)^t}{(2^m-1)^t} \\ &\leq \frac{2^{mt} - (2^{mt} - t2^{m(t-1)})}{(2^m-1)^t} \\ &\leq \frac{t \cdot 2^{mt-m}}{2^{mt}/2} \\ &\leq 2\ell 2^{-m}. \end{aligned}$$

The second inequality above holds because the sum of the remaining terms in binomial expansion of $(2^m-1)^t$ is positive as $m > 2\log \ell$. The third inequality above holds for the following reason: $(2^m-1)^t = 2^{mt}(1-1/2^m)^t \geq 2^{mt}(1-1/\ell^2)^t$ since $m > 2\log \ell$. As $\ell \geq 4$ and $t \leq \ell$, we can use the standard fact that $(1-x) \geq e^{-2x}$ for all $x < 1/2$ to conclude that $2^{mt}(1-1/\ell^2)^t \geq 2^{mt}e^{-2t/\ell^2} \geq 2^{mt}e^{-1/2} > 2^{mt}/2$. Hence, for all $y \in \{-1,1\}^\ell$, we have

$$\left| \mathbb{E}_{\mu_y}\left[\bigoplus_{(i,j)\in H} x_{i,j} \right] - \frac{1}{2} - \frac{1}{2}\bigvee_{i\in I} y_i \right| \leq 2\ell 2^{-m}. \tag{4.3}$$ ◻

The next lemma states that g can be approximated well over the OR basis, given a good approximation for $g \circ \bigvee$ over the parity basis.

LEMMA 4.3. *Let $\ell \geq 4$ and m be positive integers such that $m > 2\log \ell$, and $g : \{-1,1\}^\ell \to \{-1,1\}$ be any function. Define $f = g \circ \bigvee_m : \{-1,1\}^{m\ell} \to \{-1,1\}, \Delta \in \mathbb{R}, e_x \geq 0$ for all $x \in X$, where X denotes the set of all inputs x in $\{-1,1\}^{m\ell}$ such that $\bigvee_m(x) = -1^\ell$, and let p be a real polynomial such that*

$$\forall x \in \{-1,1\}^{m\ell}, \qquad f(x)p(x) \geq \Delta,$$
$$\forall x \in X, \qquad f(x)p(x) \geq \Delta + e_x.$$

Then there exists an OR polynomial q, of weight at most $\mathrm{wt}(p)$, such that

$$\forall y \in \{-1,1\}^\ell, \qquad q(y)g(y) \geq \Delta - \mathrm{wt}(p)\left(2\ell \cdot 2^{-m}\right),$$
$$q(-1^\ell)g(-1^\ell) \geq \Delta + \frac{\sum_{x\in X} e_x}{|X|} - \mathrm{wt}(p)\left(2\ell \cdot 2^{-m}\right).$$

Proof. Note that for any $y \in \{-1,1\}^\ell$,

$$\mathbb{E}_{\mu_y}[f(x)p(x)] = g(y) \cdot \mathbb{E}_{\mu_y}[p(x)] \geq \Delta \tag{4.4}$$

and

$$\mathbb{E}_{\mu_{-1^\ell}}[f(x)p(x)] = g(-1^\ell) \cdot \mathbb{E}_{\mu_{-1^\ell}}[p(x)] \geq \Delta + \frac{\sum_{x\in X} e_x}{|X|}. \tag{4.5}$$

Denote the unique multilinear expansion of p by $p = v_0 + \sum_k v_k p_k$, where $p_k(x) = \oplus_{(i,j)\in H_k} x_{i,j}$. Let I_k denote the projection of H_k on $[\ell]$. Define

$$q = v_0 - \frac{\sum_k v_k}{2} - \sum_k \frac{v_k}{2} \bigvee_{i \in I_k} y_i.$$

Note that

$$\mathrm{wt}(q) = \mathrm{wt}\left(v_0 - \frac{\sum_k v_k}{2} - \sum_k \frac{v_k}{2} \bigvee_{i\in I_k} y_i\right) = \left|v_0 - \frac{\sum_k v_k}{2}\right| + \sum_k \left|\frac{v_k}{2}\right| \leq \mathrm{wt}(p).$$

Thus, using linearity of expectation and Lemma 4.2, (4.4) and (4.5) yield that for all $y \in \{-1,1\}^\ell$,

$$q(y) \cdot g(y) \geq \Delta - \mathrm{wt}(p)\left(2\ell \cdot 2^{-m}\right)$$

and

$$q(-1^\ell) \cdot g(-1^\ell) \geq \Delta + \frac{\sum_{x\in X} e_x}{|X|} - \mathrm{wt}(p)\left(2\ell \cdot 2^{-m}\right). \qquad \square$$

Next, we use random restrictions that reduce the degree of the approximating OR polynomial, at the cost of a small error. In particular, we consider the case when $g = \mathsf{OMB}^0_\ell$. This represents the dashed implication in Figure 2.

LEMMA 4.4. *Let $\ell \geq 4$ and m be any positive integers such that $m > 2\log \ell$. Let $g_\ell = \mathsf{OMB}^0_\ell : \{-1,1\}^\ell \to \{-1,1\}, f = g_\ell \circ \bigvee_m$, and $\Delta, \{e_x \geq 0 : x \in X\}$ (where X is defined as in Lemma 4.3), and p be a real polynomial such that*

$$\forall x \in \{-1,1\}^{m\ell},\ f(x)p(x) \geq \Delta,$$
$$\forall x \in X,\ p(x) \geq \Delta + e_x.$$

Then for any integer $d > 0$, there exists an OR *polynomial $r : \{-1,1\}^{\ell/8} \to \mathbb{R}$, of degree d and weight at most* $\mathrm{wt}(p)$*, such that*

$$\forall y \in \{-1,1\}^{\ell/8},\ r(y)g_{\ell/8}(y) \geq \Delta - \mathrm{wt}(p)\left(2\ell \cdot 2^{-m} + 2^{-(d-1)}\right)$$

$$\text{and} \qquad r(-1^{\ell/8}) \geq \Delta + \frac{\sum_{x\in X} e_x}{|X|} - \mathrm{wt}(p)\left(2\ell \cdot 2^{-m} + 2^{-(d-1)}\right).$$

Proof. Lemma 4.3 guarantees the existence of an OR polynomial q, of weight at most $\mathrm{wt}(p)$, such that

$$\forall y \in \{-1,1\}^{\ell}, \qquad q(y)g_\ell(y) \geq \Delta - \mathrm{wt}(p)\left(2\ell \cdot 2^{-m}\right), \tag{4.6}$$
$$q(-1^\ell)g(-1^\ell) \geq \Delta + \frac{\sum_{x\in X} e_x}{|X|} - \mathrm{wt}(p)\left(2\ell \cdot 2^{-m}\right).$$

Now, set each of the ℓ variables to -1 with probability $1/2$, and leave it unset with probability $1/2$. Call this random restriction R. Any OR monomial of degree at least d gets fixed to -1 with probability $1-2^{-d}$. Thus, by linearity of expectation, the expected weight of surviving monomials of degree at least d in q is at most $\mathrm{wt}(p)\cdot 2^{-d}$. Let $M|_R$ denote the value of a monomial M after the restriction R. By Markov's inequality,

$$\Pr_R\left[\sum_{M:\deg(M|_R)\geq d} \mathrm{wt}(M|_R) > \mathrm{wt}(p)\cdot 2^{-d+1}\right] < 1/2.$$

Consider $\ell/2$ pairs of variables, $\{(x_{2i}, x_{2i+1}) : i \in [\ell/2]\}$ (assume without loss of generality that ℓ is even). For any pair, the probability that both of its variables remain unset is $1/4$. This probability is independent over pairs. Thus, by a Chernoff bound (Lemma 2.1), the probability that at most $\ell/16$ pairs remain unset is at most $2^{-\frac{\ell}{64}}$, since the expected number of pairs with both variables unset equals $\ell/8$.

By a union bound, there exists a setting of variables such that at least $\ell/16$ pairs of variables are left free, and the weight of degree $\geq d$ monomials in q is at most $\mathrm{wt}(p)\cdot 2^{-d+1}$. Set the remaining $7\ell/8$ variables to the value -1. After the restriction, drop the monomials of degree $\geq d$ from q to obtain r, which is now an OR polynomial of degree less than d and weight at most $\mathrm{wt}(p)$. Note that the function g_ℓ hit with this restriction just becomes $g_{\ell/8}$.

Thus, (4.6) yields the following:

$$\forall y \in \{-1,1\}^{\ell/8},\ r(y)g_{\ell/8}(y) \geq \Delta - \mathrm{wt}(p)\left(2\ell \cdot 2^{-m} + 2^{-(d-1)}\right)$$

$$\text{and} \qquad r(-1^{\ell/8}) \geq \Delta + \frac{\sum_{x\in X} e_x}{|X|} - \mathrm{wt}(p)\left(2\ell \cdot 2^{-m} + 2^{-(d-1)}\right). \qquad \square$$

4.1. Hardness of OMB^0. The following lemma states that approximating OMB^0 well by a low weight polynomial p is not possible unless the degree of p is large. This captures the last implication in Figure 2.

LEMMA 4.5. *Suppose $p : \{-1,1\}^n \to \mathbb{R}$ is a polynomial of degree $d < \sqrt{n/4}$ and let $a > 0, b \in \mathbb{R}$ be reals such that $p(-1^n) \geq a$ and $\mathsf{OMB}^0_n(x)p(x) \geq b$ for all $x \in \{-1,1\}^n$. Define*

$$p_{\max} = \max_{i\in\{0,\ldots,\lfloor n/10d^2\rfloor\}} \{2^i a + \left(3\cdot 2^i - 3\right) b\}.$$

Then there exists $x \in \{-1,1\}^n$ such that $|p(x)| \geq p_{\max}$.

A simple consequence of the above lemma is that the weight of a polynomial p (in either the OR basis or the parity basis) satisfying the assumptions of Lemma 4.5 is at least $p_{\max}$. This property of p suffices for our need.

The proof of Lemma 4.5 follows an iterative argument, making repeated use of Lemma 2.15, inspired by the arguments of Beigel [7] and Buhrman, Vereshchagin, and de Wolf [11].

Remark 4.6. We remark here that this strengthens the result of Beigel [7], who proved that any good approximation by a low degree *sign representing* polynomial for OMB^0 must have large weight. Our approximating polynomial is not constrained to be sign representing (b might be negative in Lemma 4.5). In fact, it might disagree in sign on all inputs but -1^n.

We now proceed to prove Lemma 4.5. We first require the following intermediate claim.

CLAIM 4.7. *If a and b are reals such that $a > 0, b \in \mathbb{R}$, and $2^i a + \left(3 \cdot 2^i - 2\right) b \leq 0$ for some integer $i \geq 0$, then $2^j a + \left(3 \cdot 2^j - 3\right) b < 0$ for all integers $j > i$.*

Proof. Note that since $a > 0$ and $2^i a + \left(3 \cdot 2^i - 2\right) b \leq 0$, b must be negative. For any $j > i$, write $2^j a + \left(3 \cdot 2^j - 3\right) b = 2^{j-i}\left(2^i a + \left(3 \cdot 2^i - 2\right) b\right) + (2^{j-i+1} - 3)b < 0.$ ☐

Proof of Lemma 4.5. Divide the n variables into $\lfloor n/10d^2 \rfloor$ contiguous blocks of size $10d^2$ each. Define $i_{\max} = \min\left\{\lfloor n/10d^2 \rfloor, j\right\}$, where j is the smallest nonnegative integer such that $2^j a + (3 \cdot 2^j - 2)b \leq 0$. We prove the following hypothesis by induction.

Induction hypothesis: For each $i \in \{0, \ldots, i_{\max}\}$, there exists an input $x^i \in \{-1, 1\}^n$ such that the following hold:

- $x^i_j = -1$ for all indices j to the right of the ith block (thus, $x^0 = (-1)^n$).
- The values of x^i_j for indices j to the left of the ith block are set as dictated by the previous step. That is, $x^i_j = x^{i-1}_j$ for all indices j to the left of the ith block.
- $p(x^i) \geq 2^i a + \left(3 \cdot 2^i - 3\right) b$ for i even and $p(x^i) \leq -2^i a - \left(3 \cdot 2^i - 3\right) b$ for i odd.

To see why proving this hypothesis would prove Lemma 4.5, we first note that the hypothesis implies the existence of an $x \in \{-1, 1\}^n$ satisfying $|p(x)| \geq \max_{i \in \{0, \ldots, i_{\max}\}}\{2^i a + \left(3 \cdot 2^i - 3\right) b\}$. This is true for the following reason: consider any $i \leq i_{\max}$, for which $\mathrm{val}_i = 2^i a + \left(3 \cdot 2^i - 3\right) b \leq 0$. For such an i, $p(x^0) = a > \mathrm{val}_i$. For all other $i \leq i_{\max}$, the hypothesis directly shows $|p(x^i)| \geq \mathrm{val}_i$. By the definition of $i_{\max}$ and Claim 4.7, we have $2^i a + \left(3 \cdot 2^i - 3\right) b < 0$ for all $i > i_{\max}$. Thus, again $p(x^0) = a > \mathrm{val}_i$ for all $i > i_{\max}$. This yields Lemma 4.5. All that remains is to prove the induction hypothesis, which we do now.

- **Base case:** Say $i = 0$. By assumption, $p(-1^n) \geq a$. If $i_{\max} = 0$, the proof is complete. Else we proceed to the inductive step.
- **Inductive step:** Say the hypothesis is true for all $0 \leq j \leq i - 1$ for some $i \geq 1$. In the ith block, set the variables corresponding to the even indices to -1 if i is odd, and set the odd indexed variables to -1 if i is even. Set the variables outside the ith block as set in the $(i-1)$'th step. Assume that i is odd (the argument for even i follows in a similar fashion, with suitable sign changes). Denote the free variables by $y_1, \ldots, y_{5d^2}$. Define a polynomial $p_i : \{-1, 1\}^{5d^2} \to \mathbb{R}$ by $p_i(y) = \mathbb{E}_{\sigma \in S_{5d^2}} \tilde{p}(\sigma(y))$, where $\tilde{p}(y)$ denotes the value of p on input $y_1, \ldots, y_{5d^2}$, and the remaining variables are set as described earlier. The expectation is over the uniform distribution. Note that p_i is a

symmetric polynomial of degree at most d and satisfies

$$p_i(-1^{5d^2}) \geq 2^{i-1}a + \left(3 \cdot 2^{i-1} - 3\right) b, \qquad p_i(y) \leq -b \ \forall y \neq -1^{5d^2}.$$

By Lemma 2.14, there exists a univariate polynomial p_i' such that for all $j \in \{0\} \cup [5d^2]$,

$$p_i'(j) = p_i(y) \ \forall y \text{ such that } \#1(y) = j.$$

Thus,

$$p_i'(0) \geq 2^{i-1}a + \left(3 \cdot 2^{i-1} - 3\right) b, \qquad p_1'(j) \leq -b \ \forall j \in [5d^2].$$

Define $p_i'' = p_i' + b$. Thus,

$$p_i''(0) \geq 2^{i-1}a + \left(3 \cdot 2^{i-1} - 2\right) b, \qquad p_i''(j) \leq 0 \ \forall j \in [5d^2].$$

Note that $p_i''(0) > 0$ since $i - 1 < i_{\max}$. Hence the conditions of Lemma 2.15 are satisfied, and it implies the existence of a $j \in [5d^2]$ such that $p_i''(j) \leq -2^i a - \left(3 \cdot 2^i - 4\right) b$, and hence $p_i'(j) \leq -2^i a - \left(3 \cdot 2^i - 3\right) b$. This implies the existence of an x^i in $\{-1, 1\}^n$ (with all variables to the right of the ith block still set to -1, and variables to the left of the ith block as dictated by the previous step) such that $p(x^i) < -2^i a - \left(3 \cdot 2^i - 3\right) b$. ☐

We are now ready to prove our main technical result, following the schematic depicted in Figure 2.

Proof of Theorem 4.1. Let p be a polynomial of weight 1, for which (LP2) attains its optimum. Denote the values taken by the variables at the optimum by $\Delta_{\mathsf{OPT}}, \{\xi_{x,\mathsf{OPT}} : x \in X\}$. Toward a contradiction, assume $\mathsf{OPT} \geq 2^{-\frac{\ell^{1/3}}{81}}$.

Lemma 4.4 (set $m = \ell^{1/3} + \log \ell$) shows the existence of an OR polynomial r, on $\ell/8$ variables, of degree $\ell^{1/3}$ and weight 1, such that

$$\text{for all } y \in \{-1,1\}^{\ell/8}, \ r(y)\mathsf{OMB}^0_{\ell/8}(y) \geq \Delta_{\mathsf{OPT}} - 4 \cdot 2^{-\ell^{1/3}}$$
$$\text{and} \qquad r(-1^{\ell/8}) \geq \Delta_{\mathsf{OPT}} + \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-\ell^{1/3}}.$$

Note that

$$\mathsf{OPT} \geq 2^{-\frac{\ell^{1/3}}{81}} \implies \Delta_{\mathsf{OPT}} \geq 2^{-\frac{\ell^{1/3}}{81}} - \delta \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{2^n}. \tag{4.7}$$

The polynomial r satisfies the assumptions of Lemma 4.5 with $d = \deg(r) = \ell^{1/3} < \sqrt{\ell/32}$ (since any OR polynomial of degree d can be represented by a polynomial of degree at most d), $a = \Delta_{\mathsf{OPT}} + \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-\ell^{1/3}}$, and $b = \Delta_{\mathsf{OPT}} - 4 \cdot 2^{-\ell^{1/3}}$. Here, a is positive because of the following:

$$a = \Delta_{\mathsf{OPT}} + \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-\ell^{1/3}} \geq 2^{-\frac{\ell^{1/3}}{81}} - 4 \cdot 2^{-\ell^{1/3}} > 0.$$

Set $k = \ell^{1/3}/80$ for the remainder of this proof. By Lemma 4.5, there exists an $x \in \{-1,1\}^{\ell/8}$ such that

$$\begin{aligned}
|r(x)| &\geq 2^k a + \left(3 \cdot 2^k - 3\right) b \\
&= \Delta_{\mathsf{OPT}}(4 \cdot 2^k - 3) + 2^k \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-80k}(4 \cdot 2^k - 3) \\
&\geq \left(4 \cdot 2^k - 3\right) \left(2^{-\frac{\ell^{1/3}}{81}} - \delta \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{2^n}\right) \\
&+ 2^k \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-80k}(4 \cdot 2^k - 3) \\
&\geq \left(4 \cdot 2^k - 3\right) \left(2^{-80k/81} - 4 \cdot 2^{-80k}\right) > 1.
\end{aligned}$$

The second inequality above holds because of (4.7), and the last inequality follows because $\delta = 1/4$, and $k \geq 1$. This yields a contradiction, since r was a polynomial of weight (in the OR basis) at most 1. ∎

5. Proof of main theorem. We are now ready to prove our sign-rank lower bound.

THEOREM 5.1 (restatement of Theorem 1.1). *Let* $f = \mathsf{OMB}_\ell^0 \circ \bigvee_{\ell^{1/3}+\log \ell} : \{-1,1\}^{\ell^{4/3}+\ell\log\ell} \to \{-1,1\}$. *Then, for sufficiently large values of* ℓ,

$$\mathrm{sr}(f \circ \mathsf{XOR}) \geq 2^{\frac{\ell^{1/3}}{81}-3}.$$

Proof. Let $n = \ell^{4/3} + \ell \log \ell$. Theorem 4.1 says that the optimum of (LP2) (and hence (LP1), by duality) is at most $2^{-\frac{\ell^{1/3}}{81}}$, when $f = \mathsf{OMB}_\ell^0 \circ \bigvee_{\ell^{1/3}+\log\ell}, \delta = 1/4$, and $X = \{x \in \{-1,1\}^{\ell^{4/3}+\ell\log\ell} : \bigvee(x) = -1^\ell\}$. We now estimate the size of X^c. The probability (over the uniform distribution on the inputs) of a particular OR gate firing a 1 is $2^{-\ell^{1/3}+\log\ell}$. By a union bound, the probability of any OR gate firing a 1 is at most $2^{-\ell^{1/3}}$, and hence $|X^c| \leq 2^n \cdot 2^{-\ell^{1/3}}$. Plugging these values into Theorem 3.1, we obtain

$$\mathrm{sr}(f \circ \mathsf{XOR}) \geq \frac{1/4}{2^{-\frac{\ell^{1/3}}{81}} + 2^{-\ell^{1/3}-2}} \geq 2^{\frac{\ell^{1/3}}{81}-3}.$$ ∎

6. Applications. In this section, we outline some applications of Theorem 1.1.

6.1. A separation of depth-2 threshold circuit classes. We are now ready to prove Theorem 1.2, which gives us a lower bound on the size of $\mathsf{THR} \circ \mathsf{MAJ}$ circuits computing $F_n = \mathsf{OMB}_\ell^0 \circ \bigvee_{\ell^{1/3}+\log\ell} \circ \mathsf{XOR}_2$ and resolves an open question posed in [5, 31] by yielding an exponential separation between the circuit classes $\mathsf{THR} \circ \mathsf{MAJ}$ and $\mathsf{THR} \circ \mathsf{THR}$.

Proof of Theorem 1.2. First, we show that F_n is computable by linear-size $\mathsf{THR} \circ \mathsf{THR}$ formulas. Let $n = 2\ell^{4/3} + 2\ell\log\ell$ denote the number of input bits to $F_n = \mathsf{OMB}_\ell^0 \circ \bigvee_{\ell^{1/3}+\log\ell} \circ \mathsf{XOR}_2$. By Lemma 2.9, F_n can be computed by a $\mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}_2$ formula of size $2\ell^{4/3} + 2\ell\log\ell$. Hence $F_n \in \mathsf{THR} \circ \mathsf{ETHR} = \mathsf{THR} \circ \mathsf{THR}$, by Theorem 2.7.

Next, we show a lower bound on the size of any $\mathsf{THR} \circ \mathsf{MAJ}$ circuit computing F_n. Suppose $\mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} \circ \mathsf{XOR}_2$ could be represented by a $\mathsf{THR} \circ \mathsf{MAJ}$ circuit of size s. By Lemma 2.11 and Theorem 5.1,

$$s\left(2\ell^{4/3} + 2\ell\log\ell\right) \geq \mathrm{sr}(f) \geq 2^{\frac{\ell^{1/3}}{81}-3}.$$

Thus, $s = 2^{\Omega\left(n^{1/4}\right)}$. ∎

6.2. Communication complexity class separations. We now show explicit separations between certain communication complexity classes, resolving an open question posed in [29]. This application of our main result was brought to our attention by Göös [27]. Precise definitions of communication complexity classes of interest may be found in the appendix.

The function we use for the class separations is $F = \mathsf{OMB}_\ell^0 \circ \bigvee_{\ell^{1/3}+\log \ell} \circ \mathsf{XOR}_2$.

Theorem 6.1. *For a positive integer ℓ, let $m = \ell^{1/3} + \log \ell$. Let $f = \mathsf{OMB}_\ell^0 \circ \bigvee_m : \{-1,1\}^{\ell m} \to \{-1,1\}$, and let $n = \ell m$ denote the number of input variables. Then, for sufficiently large values of n,*

$$\mathsf{UPP}(f \circ \mathsf{XOR}) = \Omega\left(n^{1/4}\right).$$

Proof. It follows from Theorems 5.1 and 2.16. ∎

Note that $F_n = \mathsf{OMB}_\ell \circ \mathsf{EQ}_{\ell^{1/3}+\log \ell}$, where OMB_ℓ outputs -1 iff the rightmost bit of the input set to -1 occurs at an odd index.

It is not hard to see that there is an MA protocol for $\bigvee_\ell \circ \mathsf{EQ}_{\ell^{1/3}+\log \ell}$ of cost polylogarithmic in ℓ. Using this, and a binary search, we exhibit a P^{MA} upper bound for F_n under the natural partition of the inputs in Algorithm 6.1.

Protocol 6.1 P^{MA} protocol for $\mathsf{OMB}(\mathsf{EQ}_1, \ldots, \mathsf{EQ}_\ell)$.

if $\bigvee_{i=1}^{\ell} (\mathsf{EQ}_i) = 1$ **then** Output 1.
end if
$start = 1$
$end = \ell$
$mid = \lceil \frac{start+end}{2} \rceil$
while $start \neq end$ **do**
  if $\bigvee_{i=mid}^{end} (\mathsf{EQ}_i) = -1$ **then** $start \leftarrow mid$
  else if $\bigvee_{i=mid}^{end} (\mathsf{EQ}_i) = 1$ **then** $end \leftarrow mid - 1$
  end if
end while
Output -1 iff $start$ is odd.

Hence, we obtain $F_n \in \mathsf{P}^{\mathsf{MA}cc}$. Along with Theorem 6.1, this yields the following result.

Theorem 6.2.

$$\mathsf{P}^{\mathsf{MA}cc} \not\subseteq \mathsf{UPP}^{cc}.$$

It is known that $\mathsf{P}^{\mathsf{MA}cc} \subseteq \mathsf{S2P}^{cc}$ and $\mathsf{P}^{\mathsf{MA}cc} \subseteq \mathsf{BPP}^{\mathsf{NP}cc}$ (see, e.g., [29] for references for such containments and an excellent overview on the landscape of communication complexity classes).

Thus, Theorem 6.2 yields

$$\mathsf{S2P}^{cc} \not\subseteq \mathsf{UPP}^{cc} \qquad \text{and} \qquad \mathsf{BPP}^{\mathsf{NP}cc} \not\subseteq \mathsf{UPP}^{cc}.$$

The first noninclusion resolves an open question posed in [29]. To the best of our knowledge, ours is the first explicit total function to witness the second noninclusion. We remark here that Bouland et al. [9] used a partial function to witness the same separation.

7. An upper bound. In this section, we observe that the function F_n has sign-rank $2^{O(n^{1/4})}$, showing that our lower bound in Theorem 1.1 is essentially tight for F_n.

THEOREM 7.1. *The function F_n has sign-rank $2^{O(n^{1/4})}$.*

Proof. As noted in the previous section, F_n is expressible as a circuit of the form $\mathsf{THR}_\ell \circ \mathsf{EQ}_{\ell^{1/3}+\log \ell}$. A natural unbounded-error protocol for F_n is to sample an input to the top threshold with probability proportional to its weight and solve the corresponding Equality deterministically. The cost associated with sampling an input to the threshold is $\log \ell$, and the cost of solving an Equality deterministically is $\ell^{1/3}+\log \ell$, which is at most $2\ell^{1/3}$ for large enough values of ℓ. Since $n = \ell^{4/3}+\ell \log \ell > \ell^{4/3}$, the cost of the unbounded-error protocol is $O(n^{1/4})$. By Theorem 2.16, F_n has sign-rank $2^{O(n^{1/4})}$. ∎

8. Conclusions and future directions. We exhibit the first function known to be computable efficiently (in fact in linear size) by depth-2 threshold circuits, but which has exponentially large sign-rank. This result solves two open problems in one go: the first is a basic and old open problem, arising from the classical work of Goldmann, Håstad, and Razborov [25] from the early nineties, of determining the power of weights in depth-2 threshold circuits. Can such circuits be efficiently simulated by depth-2 circuits in which the bottom gates are restricted to have small weights? Goldmann and co-authors showed that they can be if we allow only small weights to appear at the top gate in the circuit we want to simulate. We prove that in general, such efficient simulations are impossible. This, along with previous work, yields the following fine structure of depth-2 threshold circuit classes:

$$\widehat{LT}_1 \subsetneq LT_1 \subsetneq \widehat{LT}_2 = \mathsf{MAJ} \circ \mathsf{THR} \subsetneq \underbrace{\mathsf{THR} \circ \mathsf{MAJ} \subsetneq LT_2}_{\text{This work}} \subseteq \widehat{LT}_3 \subseteq \mathrm{NP/poly}.$$

The currently best known lower bounds against $\mathsf{THR} \circ \mathsf{THR}$ circuits are subquadratic, due to Kane and Williams [38]. Our work provides the first formal explanation of why current techniques have failed so far to prove strong lower bounds against $\mathsf{THR} \circ \mathsf{THR}$ circuits. It also suggests following directions along which progress can be made on this longstanding problem:

- How large can the sign-rank of a function in $\mathsf{THR} \circ \mathsf{THR}$ be? We showed that it can be as large as $2^{\Omega(n^{1/4})}$. Is it possible that the sign-rank of all functions in $\mathsf{THR} \circ \mathsf{THR}$ is $2^{O(n^\epsilon)}$ for some constant $\epsilon < 1$? Even an upper bound of $2^{o(n)}$ is enough to show IP is not in $\mathsf{THR} \circ \mathsf{THR}$. On the other hand, finding a function of sign-rank $2^{\Omega(n)}$ would also be quite interesting!
- Our function is just a short decision list of Equalities. While it is not hard to show that decision lists of Equalities cannot compute[7] IP, can we prove strong lower bounds on the size of *decision lists of exact thresholds* for computing an explicit (say, in NP) function? This is a subclass of $\mathsf{THR} \circ \mathsf{THR}$. Our main result shows that this subclass already inherits the curse of large sign-rank.

[7]Since they are in AC^0, for instance.

This raises the challenge of proving lower bounds on their size as a natural next step.

On a second front, our main result shows that the communication complexity class P^{MA} has functions with large sign-rank, strongly resolving an open problem posed recently by Göös, Pitassi, and Watson [29]. This is in contrast to the known facts that all functions in P^{NP} and MA have small sign-rank. As the sign-rank lower bound technique remains the strongest known technique for proving lower bounds against communication protocols (including quantum protocols), it suggests that new techniques need to be developed for proving bounds against P^{MA}. Indeed, there are specialized techniques for proving lower bounds against the class P^{NP} (see [37, 28]). Can they be generalized to P^{MA}? In particular, note that every function expressible as a short decision list of exact thresholds is in P^{MA}. Proving lower bounds on the length of such decision lists for computing an explicit function is also a natural first step for proving lower bounds against P^{MA} communication protocols.

In conclusion, our work puts the spotlight on the basic and simple computational model of "decision lists of exact thresholds" that is capable of very efficiently computing a function of large sign-rank. Proving lower bounds on the size of such decision lists is a necessary step for proving lower bounds against both $\mathsf{THR} \circ \mathsf{THR}$ circuit size and P^{MA} communication cost.

Appendix A. Communication complexity classes. For any communication model $\mathcal{C}$ and function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$, denote $\mathcal{C}(F)$ to be the minimum cost of a correct protocol for F in the model $\mathcal{C}$. We denote by $\mathcal{C}^{cc}$ the class of all functions F with $\mathcal{C}(F)$ at most polylogarithmic in n.

Definition A.1 (NP). *An* NP *protocol* Π *outputs* -1 *or* 1 *indicating whether or not the input is in* $\bigcup_{w\in\{-1,1\}^k} R_w$*, where* $\{R_w : w \in \{-1,1\}^k\}$ *are rectangles. The protocol correctly computes* $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ *if for all* $(x,y) \in \{-1,1\}^n \times \{-1,1\}^n$, $\Pi(x,y) = F(x,y)$*. The cost of the protocol is* k*.*

Definition A.2 (MA). *An* MA *protocol is a distribution over deterministic protocols* Π *that take an additional input* $w \in \{-1,1\}^k$ *(Merlin's proof string), visible to both Alice and Bob. The protocol correctly computes* $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ *if it satisfies the following properties:*

$$\begin{aligned} &\textit{Completeness:} && \textit{If } F(x,y) = -1, \ \textit{then } \exists w : \Pr[\Pi(x,y,w) = -1] \geq 2/3. \\ &\textit{Soundness:} && \textit{If } F(x,y) = 1, \ \textit{then } \forall w : \Pr[\Pi(x,y,w) = -1] \leq 1/3. \end{aligned}$$

The cost of the protocol is the sum of the maximum cost of any constituent deterministic protocol and k*.*

Definition A.3 ($\mathsf{S_2P}$). *An* $\mathsf{S_2P}$ *protocol can be viewed as a matrix* Π*, with rows indexed by* $r \in \{-1,1\}^k$*, columns indexed by* $c \in \{-1,1\}^k$*, where each entry contains a deterministic protocol. The protocol correctly computes* $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ *if the matrix satisfies the following properties:*

$$\begin{aligned} &\textit{If } F(x,y) = -1, \ \textit{then } \exists c \ \forall r : \ \Pi_{r,c}(x,y) = -1, \\ &\textit{if } F(x,y) = 1, \ \textit{then } \exists r \ \forall c : \ \Pi_{r,c}(x,y) = 1. \end{aligned}$$

The cost of the protocol is the sum of the maximum cost of any constituent deterministic protocol and k*.*

We now define protocols where Alice and Bob have access to certain oracles.

Definition A.4 (P^{NP}). *A* P^{NP} *protocol* Π *is a protocol in which at each step, one of the following actions occur:*

- *For cost* 1*, Alice sends a bit to Bob.*
- *For cost* 1*, Bob sends a bit to Alice.*
- *For cost* k*, Alice and Bob compute the value of* $g(x,y)$*, where* g *has an* NP *protocol of cost* k*.*

The protocol correctly computes $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ *if* $\Pi(x,y) = F(x,y)$ *for all* $x, y \in \{-1,1\}^n$*.*

Definition A.5 (P^{MA}). *A* P^{MA} *protocol* Π*, is a protocol in which at each step, one of the following actions occur:*

- *For cost* 1*, Alice sends a bit to Bob.*
- *For cost* 1*, Bob sends a bit to Alice.*
- *For cost* k*, Alice and Bob compute the value of* $g(x,y)$*, where* g *has an* MA *protocol of cost* k*.*

The protocol correctly computes $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ *if* $\Pi(x,y) = F(x,y)$ *for all* $x, y \in \{-1,1\}^n$*.*

Definition A.6 ($\mathsf{BPP}^{\mathsf{NP}}$). *A* $\mathsf{BPP}^{\mathsf{NP}}$ *protocol is a distribution over* P^{NP} *protocols* Π*. The protocol correctly computes* $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ *if* $\Pr[\Pi(x,y) = F(x,y)] \geq 2/3$ *for all* $x, y \in \{-1,1\}^n$*.*

Appendix B. Sign-monomial complexity lower bounds.

Definition B.1 (sign-monomial complexity). *The* sign-monomial complexity *of a Boolean function* $f : \{-1,1\}^n \to \{-1,1\}$*, denoted by* $\mathrm{mon}_{\pm}(f)$*, is the minimum number of monomials required by a polynomial* p *to sign represent* f *on all inputs.*

Remark B.2. Note that the sign-monomial complexity of a function f exactly corresponds to the minimum size Threshold of Parity circuit computing f.

In this section, we show how upper bounding the optimum of LP1 (and LP2) w.r.t. a function f yields sign-monomial complexity lower bounds for representing it. This is already implied by Theorem 3.1, as a sign-rank lower bound on $f \circ \mathsf{XOR}$ directly implies a sign-monomial complexity lower bound on f. The use of Theorem 3.1, whose proof makes use of the deep result of Forster [23], seems an overkill to just lower bound sign-monomial complexity. In this section, we give a much more direct proof of this fact, entirely avoiding the use of Forster's theorem. This also allows us to generalize a classical result of Bruck [10] that gave a sufficient condition for lower bounding complexity. One may note that our generalization is analogous to Razborov and Sherstov's [54] generalization of Forster's theorem. Further, our generalized result, Theorem B.4, along with Theorem 4.1, will directly imply that there are functions in poly-size $\mathsf{THR} \circ \mathsf{OR}$ circuits that cannot be computed in subexponential size by $\mathsf{THR} \circ \mathsf{XOR}$ circuits. Such a result was first proved by Krause and Pudlák [44], using a different technique. Interestingly, Krause and Pudlák expressed the belief that such a separation cannot be done based on a spectral technique like that of Bruck's theorem [10]. Our argument here shows that this belief was false.

We recall Bruck's theorem below.

Theorem B.3 ([10]). *Let* $f : \{-1,1\}^n \to \{-1,1\}$ *be any Boolean function. If* $\max_{S \subseteq [n]} \left|\hat{f}(S)\right| \leq \epsilon$*, then*

$$\mathrm{mon}_{\pm}(f) \geq \frac{1}{\epsilon}.$$

The following is our generalization of Theorem B.3.

THEOREM B.4. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function and X any subset of $\{-1,1\}^n$. If there exists a distribution μ on $\{-1,1\}^n$ such that $\max_{S\subseteq[n]} \left|\widehat{f\mu}(S)\right| \le \epsilon$ and $\min_{x\in X} \mu(x) \ge \delta$, then*

$$\mathrm{mon}_{\pm}(f) \ge \frac{\delta}{\epsilon + \delta \cdot \frac{|X^c|}{2^n}}.$$

We require the following lemma (see, for example, [50, Exercise 3.9]).

LEMMA B.5 (folklore). *For any function $f : \{-1,1\}^n \to \mathbb{R}$,*

$$\mathbb{E}_{x\in\{-1,1\}^n}\left[|f(x)|\right] \ge \max_{S\subseteq[n]} \left|\widehat{f}(S)\right|.$$

Proof of Theorem B.4. Let $p : \{-1,1\}^n \to \mathbb{R}$ be any polynomial which sign represents f. By Fact 2.12,

$$\begin{aligned}\text{(B.1)}\qquad \mathbb{E}_x[f(x)\mu(x)p(x)] &= \sum_{S\subseteq[n]} \widehat{f\mu}(S)\widehat{p}(S) \le \max_{S\subseteq[n]} \left|\widehat{f\mu}(S)\right| \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| \cdot \mathrm{mon}(p)\\ &\le \epsilon \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| \cdot \mathrm{mon}(p).\end{aligned}$$

Note that

$$\begin{aligned}\mathbb{E}_x[f(x)\mu(x)p(x)] &= \frac{1}{2^n}\sum_{x\in X} f(x)\mu(x)p(x) + \frac{1}{2^n}\sum_{x\in X^c} f(x)\mu(x)p(x)\\ &\ge \frac{\min_{x\in X}\mu(x)}{2^n}\left[\sum_{x\in\{-1,1\}^n} |p(x)| - |X^c| \cdot \max_{x\in X^c} |p(x)|\right]\\ &\ge \delta \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| - \frac{\delta}{2^n} \cdot |X^c| \cdot \max_{x\in\{-1,1\}^n} |p(x)|,\end{aligned}$$

where the first inequality holds since p sign represents f, and the last inequality uses Lemma B.5. Combining the above and (B.1), we obtain

$$\begin{aligned}&\epsilon \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| \cdot \mathrm{mon}(p) \ge \delta \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| - \frac{\delta}{2^n} \cdot |X^c| \cdot \max_{x\in\{-1,1\}^n} |p(x)|\\ &\Longrightarrow \epsilon \cdot \mathrm{mon}(p) \ge \delta - \frac{\delta}{2^n} \cdot |X^c| \cdot \frac{\max_{x\in\{-1,1\}^n} |p(x)|}{\max_{S\subseteq[n]} |\widehat{p}(S)|} \ge \delta - \frac{\delta}{2^n} \cdot |X^c| \cdot \mathrm{mon}(p)\\ &\Longrightarrow \mathrm{mon}(p) \ge \frac{\delta}{\epsilon + \delta \cdot \frac{|X^c|}{2^n}}.\end{aligned}$$

∎

The following theorem yields a sign-monomial complexity lower bound against a function in $\mathsf{THR} \circ \mathsf{OR}$.

THEOREM B.6. *Let $f = \mathsf{OMB}^0_\ell \circ \bigvee_{\ell^{1/3}+\log \ell} : \{-1,1\}^{\ell^{4/3}+\ell\log\ell} \to \{-1,1\}$. Then,*

$$\mathrm{mon}_{\pm}(f) \ge 2^{\frac{\ell^{1/3}}{81}-3}.$$

Proof. The proof follows from Theorems B.4 and 4.1 in the same way as the proof of Theorem 5.1 follows from Theorems 3.1 and 4.1. ∎

This gives us a function f on n input variables, computable by linear-size $\mathsf{THR} \circ \mathsf{AND}$ circuits, such that for large enough n,

$$\mathrm{mon}_{\pm}(f) \geq 2^{\Omega\left(n^{1/4}\right)}.$$

Acknowledgments. We are grateful to Kristoffer Hansen for bringing to our attention the question of separating the classes $\mathsf{THR}\circ\mathsf{MAJ}$ and $\mathsf{THR}\circ\mathsf{THR}$ at the Summer School on Lower Bounds, held in Prague in 2015. We also thank Michal Koucký for organizing and inviting us to the workshop. We thank Mika Göös for pointing out to us the consequences of our results in communication complexity class separations (subsection 6.2), and Justin Thaler, Ryan Williams, and Jaikumar Radhakrishnan for invaluable feedback on an earlier version of this manuscript.

REFERENCES

[1] A. Ada, O. Fawzi, and R. Kulkarni, *On the Spectral Properties of Symmetric Functions*, CoRR, abs/1704.03176, 2017.

[2] J. Alman and R. R. Williams, *Probabilistic rank and matrix rigidity*, in Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC, 2017, pp. 641–652.

[3] N. Alon, P. Frankl, and V. Rödl, *Geometrical realization of set systems and probabilistic communication complexity*, in Proceedings of the 26th Annual Symposium on Foundations of Computer Science, 1985, pp. 277–280.

[4] N. Alon, S. Moran, and A. Yehudayoff, *Sign rank versus VC dimension*, in Proceedings of the 29th Conference on Learning Theory, COLT, 2016, pp. 47–80.

[5] K. Amano and A. Maruoka, *On the complexity of depth-2 circuits with threshold gates*, in Proceedings of the 30th International Symposium on Mathematical Foundations of Computer Science, 2005, pp. 107–118; also available online from https://www.cs.gunma-u.ac.jp/~amano/paper/mfcs_2k5_ex.ps.

[6] L. Babai, P. Frankl, and J. Simon, *Complexity classes in communication complexity theory (preliminary version)*, in Proceedings of the 27th Annual Symposium on Foundations of Computer Science, FOCS, 1986, pp. 337–347.

[7] R. Beigel, *Perceptrons, PP, and the polynomial hierarchy*, Comput. Complexity, 4 (1994), pp. 339–349.

[8] A. Bernasconi and B. Codenotti, *Spectral analysis of Boolean functions as a graph eigenvalue problem*, IEEE Trans. Comput., 48 (1999), pp. 345–351, https://doi.org/10.1109/12.755000.

[9] A. Bouland, L. Chen, D. Holden, J. Thaler, and P. N. Vasudevan, *On the power of statistical zero knowledge*, SIAM J. Comput., 49 (2019), FOCS17-1.

[10] J. Bruck, *Harmonic analysis of polynomial threshold functions*, SIAM J. Discrete Math., 3 (1990), pp. 168–177.

[11] H. Buhrman, N. Vereshchagin, and R. de Wolf, *On computation and communication with small bias*, in Proceedings of the 22nd Annual IEEE Conference on Computational Complexity, CCC '07, IEEE Computer Society, 2007, pp. 24–32.

[12] M. Bun and J. Thaler, *Improved bounds on the sign-rank of AC^0*, in Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming, ICALP, 2016, pp. 37:1–37:14.

[13] M. Bun and J. Thaler, *A nearly optimal lower bound on the approximate degree of AC^0*, SIAM J. Comput., 49 (2019), FOCS17-59.

[14] M. Bun and J. Thaler, *The large-error approximate degree of AC^0*, Theory Comput., 17 (2021), pp. 1–46.

[15] A. Chattopadhyay and N. S. Mande, *Dual polynomials and communication complexity of XOR functions*, Electron. Colloq. Comput. Complexity (ECCC), 24 (2017), 62.

[16] A. Chattopadhyay and N. S. Mande, *A short list of equalities induces large sign rank*, in Proceedings of the 59th IEEE Annual Symposium on Foundations of Computer Science, FOCS, Paris, 2018, pp. 47–58.

[17] A. Chattopadhyay, N. S. Mande, and S. Sherif, *The log-approximate-rank conjecture is false*, J. ACM, 67 (2020), pp. 1–28.

[18] L. Chen, *Toward Super-Polynomial Size Lower Bounds for Depth-Two Threshold Circuits*, CoRR, abs/1805.10698, 2018.

[19] L. Chen and R. Tell, *Bootstrapping results for threshold circuits "just beyond" known lower bounds*, in Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, 2019, pp. 34–41.

[20] R. Chen, R. Santhanam, and S. Srinivasan, *Average-case lower bounds and satisfiability algorithms for small threshold circuits*, Theory Comput., 14 (2018), pp. 1–55.

[21] X. Chen, I. C. Oliveira, and R. A. Servedio, *Addition is exponentially harder than counting for shallow monotone circuits*, in Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC, 2017, pp. 1232–1245.

[22] H. Ehlich and K. Zeller, *Schwankung von polynomen zwischen gitterpunkten*, Math. Z., 86 (1964), pp. 41–44.

[23] J. Forster, *A linear lower bound on the unbounded error probabilistic communication complexity*, in Proceedings of the 16th Annual IEEE Conference on Computational Complexity, CCC, 2001, pp. 100–106.

[24] J. Forster, M. Krause, S. V. Lokam, R. Mubarakzjanov, N. Schmitt, and H. U. Simon, *Relations between communication complexity, linear arrangements, and computational complexity*, in Proceedings of the 21st International Conference on Foundations of Software Technology and Theoretical Computer Science, 2001, pp. 171–182.

[25] M. Goldmann, J. Håstad, and A. A. Razborov, *Majority gates vs. general weighted threshold gates*, Comput. Complexity, 2 (1992), pp. 277–300.

[26] M. Goldmann and M. Karpinski, *Simulating threshold circuits by majority circuits*, SIAM J. Comput., 27 (1998), pp. 230–246.

[27] M. Göös, *private communication*, 2017.

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

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

[30] A. Hajnal, W. Maas, P. Pudlák, M. Szegedy, and G. Turán, *Threshold circuits of bounded depth*, J. Comput. System Sci., 46 (1993), pp. 129–154.

[31] K. A. Hansen and V. V. Podolskii, *Exact threshold circuits*, in Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC, 2010, pp. 270–279.

[32] K. A. Hansen and V. V. Podolskii, *Polynomial threshold functions and Boolean threshold circuits*, Inform. and Comput., 240 (2015), pp. 56–73.

[33] P. Harsha, A. R. Klivans, and R. Meka, *An invariance principle for polytopes*, J. ACM, 59 (2012), pp. 29:1–29:25.

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

[35] H. Hatami and Y. Qian, *The Unbounded-Error Communication Complexity of Symmetric XOR Functions*, CoRR, abs/1704.00777, 2017.

[36] T. Hofmeister, *A note on the simulation of exponential threshold weights*, in Proceedings of Computing and Combinatorics, Second Annual International Conference, COCOON, 1996, pp. 136–141.

[37] R. Impagliazzo and R. Williams, *Communication complexity with synchronized clocks*, in Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC, 2010, pp. 259–269.

[38] D. M. Kane and R. Williams, *Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits*, in Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC, 2016, pp. 633–643.

[39] S. Kannan, E. Mossel, S. Sanyal, and G. Yaroslavtsev, *Linear sketching over $\mathbb{F}_2$*, in Proceedings of the 33rd Computational Complexity Conference, CCC, 2018, pp. 8:1–8:37.

[40] H. Klauck, *Rectangle size bounds and threshold covers in communication complexity*, in Proceedings of the 18th Annual IEEE Conference on Computational Complexity, CCC, 2003, pp. 118–134.

[41] A. R. Klivans, R. O'Donnell, and R. A. Servedio, *Learning intersections and thresholds of halfspaces*, J. Comput. System Sci., 68 (2004), pp. 808–840.

[42] A. R. Klivans and R. A. Servedio, *Toward attribute efficient learning of decision lists and parities*, J. Mach. Learning Res., 7 (2006), pp. 587–602.

[43] M. Krause, *On the computational power of boolean decision lists*, Comput. Complexity, 14 (2006), pp. 362–375.

[44] M. Krause and P. Pudlák, *Computing Boolean functions by polynomials and threshold circuits*, Comput. Complexity, 7 (1998), pp. 346–370.

[45] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, Cambridge, 1997.
[46] M. Minsky and S. Papert, *Perceptrons—An Introduction to Computational Geometry*, MIT Press, Cambridge, MA, 1987.
[47] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, Cambridge, 2005, https://doi.org/10.1017/CBO9780511813603.
[48] A. Montanaro and T. Osborne, *On the Communication Complexity of XOR Functions*, CoRR, abs/0909.3392, 2009.
[49] S. Muroga, *Threshold Logic and Its Applications*, Wiley-Interscience, New York, 1971.
[50] R. O'Donnell, *Analysis of Boolean Functions*, Cambridge University Press, Cambridge, 2014.
[51] R. Paturi and J. Simon, *Probabilistic communication complexity*, J. Comput. System Sci., 33 (1986), pp. 106–123.
[52] P. Pudlák, *private communication*, 2013.
[53] A. A. Razborov, *On small depth threshold circuits*, in Proceedings of the 3rd Scandinavian Workshop on Algorithm Theory (SWAT), 1992, pp. 42–52.
[54] A. A. Razborov and A. A. Sherstov, *The sign-rank of AC^0*, SIAM J. Comput., 39 (2010), pp. 1833–1855.
[55] R. L. Rivest, *Learning decision lists*, Mach. Learn., 2 (1987), pp. 229–246.
[56] T. J. Rivlin and E. W. Cheney, *A comparison of uniform approximations on an interval and a finite subset thereof*, SIAM J. Numer. Anal., 3 (1966), pp. 311–320.
[57] R. A. Servedio and L. Tan, *Fooling intersections of low-weight halfspaces*, in Proceedings of the 58th IEEE Annual Symposium on Foundations of Computer Science, FOCS, 2017, pp. 824–835.
[58] A. A. Sherstov, *The pattern matrix method*, SIAM J. Comput., 40 (2011), pp. 1969–2000.
[59] A. A. Sherstov, *The unbounded-error communication complexity of symmetric functions*, Combinatorica, 31 (2011), pp. 583–614.
[60] A. A. Sherstov, *Optimal bounds for sign-representing the intersection of two halfspaces by polynomials*, Combinatorica, 33 (2013), pp. 73–96.
[61] A. A. Sherstov and P. Wu, *Near-Optimal Lower Bounds on the Threshold Degree and Sign-rank of AC^0*, SIAM J. Comput., (2021), STOC19-1.
[62] K. I. Siu and J. Bruck, *On the power of threshold circuits with small weights*, SIAM J. Discrete Math., 4 (1991), pp. 423–435.
[63] R. Tell, *Quantified derandomization of linear threshold circuits*, in Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC, Los Angeles, 2018, pp. 855–865.
[64] S. Zhang, *Efficient quantum protocols for XOR functions*, in Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA, 2014, pp. 1878–1885.
[65] Z. Zhang and Y. Shi, *Communication complexities of symmetric XOR functions*, Quantum Inf. Comput., 9 (2009), pp. 255–263.