

# $AC^0 \circ MOD_2$ lower bounds for the Boolean Inner Product

Mahdi Cheraghchi<sup>1</sup>, Elena Grigorescu<sup>2</sup>, Brendan Juba<sup>3</sup>, Karl Wimmer<sup>4</sup>, and Ning Xie<sup>5</sup>

- 1 Department of Computing, Imperial College London, London, UK  
m.cheraghchi@imperial.ac.uk
- 2 Department of Computer Science, Purdue University, IN, USA  
elena-g@purdue.edu
- 3 Department of Computer Science and Engineering, Washington University, St. Louis, MO, USA  
bjuba@wustl.edu
- 4 Department of Mathematics and Computer Science, Duquesne University, Pittsburgh, PA, USA  
wimmerk@duq.edu
- 5 School of Computing and Information Sciences, Florida International University, Miami, FL, USA  
nxie@cs.fiu.edu

---

## Abstract

$AC^0 \circ MOD_2$  circuits are  $AC^0$  circuits augmented with a layer of parity gates just above the input layer. We study  $AC^0 \circ MOD_2$  circuit lower bounds for computing the Boolean Inner Product functions. Recent works by Servedio and Viola (ECCC TR12-144) and Akavia et al. (ITCS 2014) have highlighted this problem as a frontier problem in circuit complexity that arose both as a first step towards solving natural special cases of the matrix rigidity problem and as a candidate for constructing pseudorandom generators of minimal complexity. We give the first superlinear lower bound for the Boolean Inner Product function against  $AC^0 \circ MOD_2$  of depth four or greater. Specifically, we prove a superlinear lower bound for circuits of arbitrary constant depth, and an  $\tilde{\Omega}(n^2)$  lower bound for the special case of depth-4  $AC^0 \circ MOD_2$ . Our proof of the depth-4 lower bound employs a new “moment-matching” inequality for bounded, nonnegative integer-valued random variables that may be of independent interest: we prove an optimal bound on the maximum difference between two discrete distributions’ values at 0, given that their first  $d$  moments match.

**Keywords and phrases** Boolean analysis, circuit complexity, lower bounds

**Digital Object Identifier** 10.4230/LIPIcs.CVIT.2016.23

## 1 Introduction

We study lower bounds for computing the inner product function by  $AC^0$  circuits with parity gates on the level just above the input gates ( $AC^0 \circ MOD_2$ ). As we will review, this problem has emerged as a common, particularly simple special case of several major open problems in Computational Complexity, about which we know surprisingly little. We therefore view progress on this special case as a benchmark for new techniques in circuit complexity for these larger questions.

A core program in Computational Complexity is to understand the power of restricted circuit families. One facet of such understanding is to identify functions that these circuits cannot compute. In practice, it turns out that once we can prove such lower bounds, then we



licensed under Creative Commons License CC-BY

42nd Conference on Very Important Topics (CVIT 2016).

Editors: John Q. Open and Joan R. Access; Article No. 23; pp. 23:1–23:20

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

become surprisingly facile with the class, gaining the ability to learn the functions computed by such circuits [14] (and this is necessary in some form [10, 11, 29]), the ability to generate inputs that are pseudorandom for the class [16, 18] (again necessary in some form [7, 24, 27]), and more. As a consequence, “understanding” the class is often identified with proving such lower bounds. It is therefore interesting when this intuition *fails* to hold.

Shaltiel and Viola [25] noticed such a gap: although we can prove that, e.g., the  $MOD_3$  function has constant hardness for  $AC^0[2]$  circuits [21, 26] (where  $AC^0[2]$  is  $AC^0$  equipped with parity gates), we still *do not* have pseudorandom generators for  $AC^0[2]$ . The trouble is that known constructions of pseudorandom generators require strongly hard on average functions [18], and proofs of hardness amplification require the class in question to compute the majority function, which  $AC^0[2]$  cannot even approximate [21]. Shaltiel and Viola therefore highlight the problem of establishing such strong average case hardness against  $AC^0[2]$  circuits as a challenge in circuit complexity. Servedio and Viola [23] pointed out that such strong hardness is not even known for  $AC^0 \circ MOD_2$ , and suggest the problem as a natural special case. In particular, they conjecture that, for this special case, the Inner Product function (IP), defined below, is an example of such a function (although it is trivially computable by  $AC^0[2]$ ).

► **Definition 1.**  $IP(x, y): \{0, 1\}^{2n} \rightarrow \{0, 1\}$  is the function  $\sum_{i=1}^n x_i y_i \pmod{2}$ .

Thus, showing that IP cannot be computed by small  $AC^0 \circ MOD_2$  circuits is a natural step towards a better understanding of  $AC^0[2]$ .

On the other hand, a better understanding of the class  $AC^0 \circ MOD_2$  turns out to be of interest to practical cryptography as well. Along similar lines, Akavia et al. [1], in the course of proposing a candidate weak pseudorandom function of minimal complexity (computable in  $AC^0 \circ MOD_2$  in this case), made a strong conjecture; namely that every  $AC^0 \circ MOD_2$  circuit has a quasipolynomially heavy Fourier coefficient. Since IP only has small Fourier coefficients, this conjecture also entails the same consequence considered by Servedio and Viola, and simply showing that IP cannot be computed by small  $AC^0 \circ MOD_2$  circuits is again a special case of this problem.

Finally, Servedio and Viola [23] note that a special case of Valiant’s matrix rigidity problem [28] is to exhibit a function that has low correlation with all sparse polynomials.  $AC^0 \circ MOD_2$  circuits are in turn well-approximated by such sparse polynomials, so giving explicit functions that are not correlated with any  $AC^0 \circ MOD_2$  functions is again a natural special case; and IP is again the natural candidate for such a function.

Proving lower bounds for  $AC^0 \circ MOD_2$  circuits computing IP is challenging since the usual techniques from the literature do not immediately apply. Specifically, although Razborov’s technique [21] establishes strong lower bounds against  $AC^0[2]$ , we note that IP *does* have small  $AC^0[2]$  circuits. There is thus no hope in using Razborov’s technique directly to prove lower bounds for IP. And of course, techniques based on random restrictions are helpless against the input layer parity gates.

Servedio and Viola noted that it follows from Jackson’s work [8, Fact 8] that depth-3  $AC^0 \circ MOD_2$  circuits (i.e., a DNF of parities) cannot approximate IP. Also, Jukna [9] has shown that such circuits computing IP must have exponential size (a bound recently optimized by Cohen and Shinkar [4]). And yet, as Servedio and Viola noted, nothing is known about depth-4 circuits, let alone  $AC^0 \circ MOD_2$  circuits of arbitrary depth.

## Our results.

In this work, we give the first nontrivial (superlinear) lower bound for IP against (arbitrary depth)  $AC^0 \circ MOD_2$ . In fact, our result is slightly stronger and applies to the broader class of *bent* functions (i.e., functions whose Fourier coefficients are all equal in magnitude, IP being a special case).

► **Theorem 2.** *If  $C$  is an  $AC^0 \circ MOD_2$  circuit of depth  $k$  and size  $S$  that computes the IP function on  $n$  variables, then  $S = \Omega(n^{1+4^{-k}})$ .*

The proof of this theorem follows by an adaptation of the results of Chaudhuri and Radhakrishnan [3] who showed a similar bound for  $AC^0$  circuits; a similar adaptation for  $AC^0[2]$  circuits was previously given by Kopparty and Srinivasan [13].

Our main theorem is an  $\tilde{\Omega}(n^2)$   $AC^0 \circ MOD_2$  lower bound for IP:

► **Theorem 3.** *Any depth-4  $AC^0 \circ MOD_2$  circuit computing the IP function on  $n$  variables must have size  $s = \Omega(n^2 / \log^6 n)$ .*

An intuitive interpretation of the above results is the following. IP is a means to “generate” all possible parities on  $n$  bits.  $AC^0 \circ MOD_2$  circuits are merely  $AC^0$  circuits that are given access to an arbitrary but fixed set of parity functions, bounded in number by the size of the circuit. Our results address the question of how much these few parities can aid the computation of most remaining parities.

## Our technique: a moment-matching bound.

At the heart of this second lower bound is a lemma that may be of independent interest:

► **Lemma 4 (Moment-matching bound).** *Let  $X$  and  $Y$  be random variables taking values in  $\{0, 1, 2, \dots, s\}$ . Suppose that the first  $d$  moments of  $X$  and  $Y$  are equal. Then,  $\Pr(Y = 0) \leq \Pr(X = 0) + e^{-\Omega(d/\sqrt{s})}$ .*

Several other “moment-matching” bounds appear in the literature, and here we briefly discuss the relationship of our work to these bounds. First, the classical “truncated moments” problem concerns the conditions for the existence of a probability distribution on a given set with a given sequence of moments [2, 5]. But, as noted by Rashkodnikova et al. [20], the solutions generated by these techniques do not necessarily lie on integers, and so the conditions refer to a different class of random variables. Klivans and Meka [12] likewise considered bounds on the difference in probability of general events that may be induced by distributions with  $d$  matching moments. Their bounds apply to much more general properties (than simply the event  $X = 0$ ) and much more general distributions; as such, in spite of some similarities in the techniques employed in their work<sup>1</sup>, they do not obtain bounds in a useful form for our purposes. Rashkodnikova et al. [20] in turn consider nonnegative, bounded, and integer-valued random variables as we do, but they consider a different property; namely, given that the first  $d$  moments are *proportional* (not necessarily identical), they maximize their ratio.

Interestingly, it turns out that the moment-matching bound we obtain has a close technical relationship to the *approximate inclusion-exclusion* bounds obtained by Linial and Nisan [15].<sup>2</sup>

<sup>1</sup> Indeed, although like us, Klivans and Meka related this problem to the existence of some polynomials via LP duality, for Klivans and Meka, constructing these (sandwiching) polynomials was the *problem*, not the *solution*.

<sup>2</sup> We are indebted to Johan Håstad for pointing out to us the similarity in the underlying technique.

Indeed, the technique we use to prove Lemma 4 is essentially the same as the core technique underlying Linial and Nisan's work, and in fact, we can show that our moment-matching lemma is essentially equivalent to Linial and Nisan's approximate inclusion-exclusion bounds (see Appendix C.1 for details). In view of the naturalness of the statement of our moment-matching bound, we believe that this lemma may be of interest, even if one is familiar with the approximate inclusion-exclusion bounds.

## 1.1 Overview of the depth-4 lower bound

Our argument consists of two main steps: (1) We show that any depth-4  $AC^0 \circ MOD_2$  circuit (without loss of generality, with an AND top gate) of size  $s \leq n^2$  computing the Inner Product function must have a one-sided approximation by a DNF of parities in which the terms are all small: It is correct when it outputs 0, and the circuit outputs zero on at least a  $1/n^2$  fraction of inputs. (2) We then show that such one-sided approximators for the Inner Product function can only output 0 with small probability, which can be made smaller than  $1/n^2$  for some  $s = O(n^2/\text{poly log } n)$ .

The first part is relatively straightforward. We let a candidate circuit for the inner product function of size  $s \leq n^2$  be given. We first obtain a one-sided approximation to our circuit by invoking the Discriminator Lemma of Hajnal et al. [6] to obtain a depth-3 circuit (eliminating the top AND layer) that is correct whenever it reports 0, and reports 0 on a large ( $\geq 1/n^2$ ) fraction of the inputs. We then reduce the fan-in of the second (from bottom) layer of AND gates by trimming the AND gates with large fan-in at a slight cost in the approximation error (asymptotically smaller than  $1/n^2$ ).

Towards the second part of our argument, we consider the *degree* of an arbitrary parity in the  $\{\pm 1\}$ -representation in terms of the original variables as well as the bottom layer parities. That is, the degree of a parity  $\chi$  is now defined as the minimum number of variables and/or bottom layer parities that need to be added together (over  $\mathbb{F}_2$ ) to obtain  $\chi$ : e.g., a single parity gate (new variable) has degree 1, and a parity of  $k$  new variables (parity gates or old variables) has degree  $\leq k$ . Given the size of the circuit  $s$ , we obtain that w.h.p. over the setting of the input  $y$  variables, the inner product function  $IP(x, y)$  is a parity in the  $x$  variables that remains of high degree (at least  $\Omega(n/\log s)$ ) over these new variables.

We show that, for a  $1 - o(1)$  fraction of fixings of the  $y$  variables, the probability that our circuit outputs 0 when  $IP(x, y) = 0$  is small as follows. We apply the above-mentioned moment-matching bound (Lemma 12) to the random variable  $N(x)$  (over a random  $x$ ) that counts the number of the AND gates in the depth-3 approximator obtained by the Discriminator Lemma that output 1. We can then show that the first  $m = \tilde{\Omega}(n)$  moments of  $(N(x) \mid IP(x, y) = 0)$  and  $(N(x) \mid IP(x, y) = 1)$  are identical and  $\Pr_x(N(x) = 0 \mid IP(x, y) = 1) = 0$  since  $N(x) = 0$  precisely when the OR gate at the output of the depth-3 one-sided approximator outputs 0, in which case the circuit is correct. Using this information in a linear-programming based proof, we show that  $\Pr_x(N(x) = 0 \mid IP(x, y) = 0) \lesssim e^{-\tilde{\Omega}(m/\sqrt{s})}$ . For our  $m$ , if  $s \leq n^2/\text{poly log } n$ , the upper bound becomes smaller than  $1/n^2$ , completing the second part and finishing the proof.

To see that the low-degree moments match, we note that  $N(x)$  is represented by a low-degree polynomial: In the  $\{0, 1\}$ -representation, it is simply the summation of monomials of degree  $O(\log n)$  corresponding to the second-level AND gates (recall that the degree remains the same in the  $\{\pm 1\}$ -representation). In the  $\{\pm 1\}$ -representation, however, it is then clear that the parity in  $x$  that we obtain from our setting of the  $y$  variables in  $IP(x, y)$  is (w.h.p. over  $y$ ) uncorrelated with  $N(x)$ . In other words,  $\mathbf{E}_x(N(x) \mid IP(x, y) = 0) = \mathbf{E}_x(N(x) \mid IP(x, y) = 1)$ . This argument can be seen to hold for larger moments as well.

We prove the moment-matching bound by writing a linear program for the probability distribution satisfying the given moment constraints over  $\{0, \dots, s\}$  that maximizes the probability of obtaining 0. We bound the value of this LP by giving an explicit dual-feasible solution; It turns out that the dual can be rewritten as maximizing the lower bound on the values a bounded degree polynomial attains at the integer points in  $[0, s]$ , given that it takes value 0 at the origin and is also upper bounded by 1 at these integer points. This is quite similar to the conditions for approximators for the OR function sought by Nisan and Szegedy [17], and our solution follows theirs, using Chebyshev polynomials to construct the desired (essentially optimal, cf. Paturi [19]) polynomial.

### An alternative proof.

There is also an alternative way of completing the proof of our lower bound that uses approximate inclusion-exclusion directly. We believe that the main proof we describe here, using moment-matching, is more natural. For completeness, a sketch of this alternative proof appears in Appendix C.2.

## 1.2 Preliminaries

Most of our notation is standard, but for reference we include an overview in Appendix A. We also include a brief review of discrete Fourier analysis in Appendix A.2.

A *linear threshold gate*  $T_k^{\mathbf{a}}(x_1, \dots, x_t)$  of fan-in  $t$  outputs 1 if and only if  $\sum_{i=1}^t a_i x_i \geq k$ , where  $\mathbf{a} = (a_1, \dots, a_t)$  is vector of weights. The Discriminator Lemma of Hajnal et al. is a powerful tool for proving lower bounds of threshold circuits.

► **Lemma 5** (Discriminator lemma, Lemma 3.3 in [6]). *Let  $C = T_k^{\mathbf{a}}(C_1, \dots, C_m)$  be a circuit on  $n$  inputs with a threshold gate at the top level, and write  $a = \sum_{i=1}^m |a_i|$ . Let  $A, B \subseteq \{0, 1\}^n$  be any two disjoint sets of inputs such that the circuit  $C$  accepts  $A$  and rejects  $B$ . Then there exists a subcircuit  $C_i$ ,  $i \in [m]$ , such that  $|\Pr_A(C_i(x) = 1) - \Pr_B(C_i(x) = 1)| \geq 1/a$ , where  $\Pr_A(C_i(x))$  (resp.,  $\Pr_B(C_i(x))$ ) denotes the uniform probability over the set  $A$  (resp.,  $B$ ).*

## 2 Lower bound for depth-4 circuit

In this section we will show an  $\tilde{\Omega}(n^2)$  lower bound for any depth-4  $\text{AC}^0 \circ \text{MOD}_2$  circuit that computes  $\text{IP}(x, y)$ . Note that all circuits here are allowed to have negations below the XOR gates; these negations are not counted in the depth of the circuit.

### 2.1 Depth-3 discriminator

Let  $C$  be any depth-4  $\text{AC}^0 \circ \text{MOD}_2$  circuit that computes  $\text{IP}(x, y)$ . First, without loss of generality, we may assume the top layer gate of  $C$  is an AND gate; the case that top layer gate is an OR gate follows a similar argument<sup>3</sup>. Second, suppose  $C = \text{AND}(C_1, \dots, C_m)$ , where each subcircuit  $C_i$  is a parity-DNF circuit; then because  $C(x, y) = \text{IP}(x, y)$  for every input, each subcircuit  $C_i$  must compute IP with *one-sided* error only. Specifically, for every input  $(x, y)$  with  $\text{IP}(x, y) = 1$  and every  $i$ ,  $C_i(x, y) = 1$ .

<sup>3</sup> One way to see this is to notice that our proof also shows the same lower bound for the negation of the Inner Product function (since negating only incurs an affine shift that our methods are not sensitive to). Thus it suffices to note that when the top gate is an OR one can just negate the layers and get a circuit in which the top gate is AND that computes the negation of the Inner Product.

We invoke a consequence of the Discriminator Lemma of Hajnal et al. [6].

► **Claim 6 (Consequence of Lemma 5).** There is a subcircuit  $C_i$ ,  $i \in [m]$ , such that  $\Pr_{(x,y): IP(x,y)=1}(C_i(x,y) = 1) = 1$ , and  $\Pr_{(x,y): IP(x,y)=0}(C_i(x,y) = 1) \leq 1 - 1/m$ .

We call such a depth-3  $AC^0 \circ MOD_2$  circuit  $C_i$  a *one-sided  $1/m$ -discriminator* for IP. Our main lemma is an upper bound on the discriminator parameter  $1/m$  of such discriminators in terms of its size.

► **Lemma 7 (Main).** *Suppose that a depth-3  $AC^0 \circ MOD_2$  circuit of size  $s$  is a one-sided  $\epsilon$ -discriminator for IP. Then  $\epsilon$  satisfies*

$$\epsilon \leq 4 \exp\left(-\sqrt{\frac{n^2}{128s \log^2 n \log^2 s}}\right) + \frac{4s}{n^4} + 2^{-n/2}.$$

The proof of Lemma 7 is discussed in Section 2.3. Assuming Lemma 7, the proof of Theorem 3 is straightforward. If  $m \geq n^2$ , then we are done already. Suppose otherwise, so  $\epsilon \geq 1/n^2$ . Then by Lemma 7, the size of discriminator subcircuit  $C_i$  is of size at least  $s = \Omega\left(\frac{n^2}{\log^6 n}\right) = \tilde{\Omega}(n^2)$ .

## 2.2 Random $y$ -restrictions

Let  $C'$  be a size- $s$  depth-3  $AC^0 \circ MOD_2$  circuit which is a one-sided  $\epsilon$ -discriminator for IP. So  $\Pr_{(x,y): IP(x,y)=0}(C'(x,y) = 0) \geq \epsilon$ , and  $C' = \text{OR}(f_1, f_2, \dots, f_{s'})$ , where each  $f_i$  is an AND of parities and  $s' < s$ . Without loss of generality, we can assume that none of these AND gates are constant (i.e., always 0 or 1).

### Reducing the fan-in of AND gates.

Define the *codimension* of  $f_i$  (each of which is an AND of parities) to be the codimension of the subspace corresponding to the coset of inputs on which  $f_i$  evaluates to 1.

For example, if  $f_1 = \text{AND}(x_1 + x_2, x_1 + x_3, \neg(x_2 + x_3))$ , then  $x_1 + x_2$  and  $x_1 + x_3$  both evaluating to 1 necessarily implies that  $\neg(x_2 + x_3)$  evaluates to 1. Hence, the set of inputs for which  $f_1(x) = 1$  is the affine subspace specified by  $\{x_1 + x_2 = 1 \wedge x_2 + x_3 = 1\}$ ; consequently, the codimension of  $f_1$  is 2.

The codimension of  $f_i$  measures the “effective” fan-in of the AND gate in  $C'$ . It is straightforward that without loss of generality one can assume the co-dimension of each AND gate to be equal to its fan-in (see Appendix B.1). From now on, we assume that all redundant parity inputs have already been removed and each AND gate in  $C'$  has its fan-in equal to its codimension. Our next step is trim those AND gates of  $C'$  whose fan-in is large.

Call an AND gate in  $C'$  “bad” if its fan-in is larger than  $4 \log n$ . We reduce  $C'$  to a circuit  $C''$  by trimming all “bad” AND gates to an arbitrary set of  $4 \log n$  inputs in their fan-in. Note that each trimmed AND gate may cause an error, only from 0 to 1, and only when all its (non-trimmed) inputs evaluate to 1 (an event that happens with probability at most  $2^{-4 \log n}$ , since the inputs of each gate are uniform and independent). Define  $\tau = \Pr_{x,y}(C'(x,y) \neq C''(x,y))$ . By the union bound,  $\tau \leq s2^{-4 \log n} = s/n^4$ . Further, if  $C'(x,y) \neq C''(x,y)$  then we must have  $C'(x,y) = 0$  and  $C''(x,y) = 1$ . In other words, if  $\epsilon' := \Pr_{(x,y): IP(x,y)=0}(C''(x,y) = 0)$ , then  $\epsilon' \geq \epsilon - \tau$ , and moreover, if  $C'$  approximates IP with a one-sided error (i.e.,  $C' = 1$  whenever  $IP = 1$ ), then so does  $C''$ .

► **Definition 8.** For a function  $F(x,y)$  (resp., a circuit  $C(x,y)$ ) that maps  $\{0,1\}^n \times \{0,1\}^n$  to  $\{0,1\}$ , a  $y$ -restriction  $\rho \in \{0,1\}^n$  is an assignment of all the  $y$  variables in the input

according to  $\rho$ . Denote the resulting function  $F$  (resp., circuit  $C$ ) after applying restriction  $\rho$  by  $F|_\rho$  (resp.,  $C|_\rho$ ).

A simple fact exploited in the proof is that, for any  $y$ -restriction  $\rho$ ,  $\text{IP}|_\rho$  is a parity over the  $x$  variables, which we denote by  $\ell_\rho$ . Note that  $\ell_\rho(x) = \sum_{i:\rho_i=1} x_i \bmod 2$ . We next argue that for *any* fixed depth-3  $\text{AC}^0 \circ \text{MOD}_2$  circuit  $C''$ , the parity function  $\ell_\rho$  resulting from a random  $y$ -restriction  $\rho$  is of “high degree” with respect to the parity inputs of  $C''$ , and thus “hard” for the circuit.

Fix an arbitrary depth-3  $\text{AC}^0 \circ \text{MOD}_2$  circuit  $C''$ , of which the fan-in of each AND gates is at most  $4 \log n$ . Let the parity inputs for  $C''$  be  $\ell_{(a_1, S_1^x, S_1^y)}, \dots, \ell_{(a_{s'}, S_{s'}^x, S_{s'}^y)}$ , where  $a_i \in \{0, 1\}$ ,  $S_i^x, S_i^y \subseteq [n]$ ,  $\ell_{S_i^x, S_i^y}(x, y) = a_i + \sum_{j \in S_i^x} x_j + \sum_{j \in S_i^y} y_j$ , and  $s' < s$ .

Observe that after applying a  $y$ -restriction  $\rho$  to  $C''$ , the inputs to  $C''|_\rho$  become the  $x$ -part of the original parities or their negations, namely  $\ell_{(a_i, S_i^x, S_i^y)}|_\rho = a'_i + \sum_{j \in S_i^x} x_j$  and  $a'_i = a_i$  or  $a'_i = 1 - a_i$ . Since there is a natural one-to-one correspondence between subsets of  $[n]$  and vectors in  $\mathbb{F}_2^n$ , we may use a set of vectors  $S \subseteq \mathbb{F}_2^n$  to identify the set of parities (or their negations), namely  $\{S_i^x\}_{i \in [s']}$ , that are fed into  $C''|_\rho$ . A key point is that the subset  $S$  depends only on the circuit  $C''$ , and is essentially independent of the choice of  $y$ -restriction  $\rho$ . Note also that  $|S| \leq s' < s$ . In the following, we will slightly abuse notation and use a parity and the subset of  $[n]$  corresponding to that parity interchangeably.

### IP results in high degree parity under random restriction.

Following standard additive combinatorial notation, for a subset  $S \subseteq \mathbb{F}_2^n$  and a positive integer  $k$ , let  $kS = \{x_1 + \dots + x_k : x_1, \dots, x_k \in S\}$ . Clearly we have  $|S \cup 2S \cup \dots \cup kS| \leq (|S| + 1)^k$ .

► **Definition 9.** For any  $S \subseteq \mathbb{F}_2^n$  and  $z \in \mathbb{F}_2^n$ , the  $S$ -degree of  $z$  is the smallest integer  $d$  such that  $z \in dS$ , or  $\infty$  if no such  $d$  exists. Further, the  $S$ -degree of a parity function  $\langle \alpha, x \rangle$  for  $\alpha \in \mathbb{F}_2^n$  is the  $S$ -degree of  $\alpha$ .

Our next claim shows that for any fixed size- $s$  depth-3  $\text{AC}^0 \circ \text{MOD}_2$  circuit  $C''$ , after applying a random  $y$ -restriction, then almost surely, the resulting parity function  $\ell_\rho$  is of high degree with respect to the parity inputs of  $C''|_\rho$ .

► **Claim 10.** Let  $S \subseteq \mathbb{F}_2^n$  be the set of input parities (or their negations) of  $C''|_\rho$ . Then with probability at least  $1 - 2^{-n/2}$  over the choice of  $\rho$ ,  $\ell_\rho$  has  $S$ -degree larger than  $n/(2 \log s)$ .

**Proof.** Set  $k = n/(2 \log s)$ . We have  $|S \cup 2S \cup \dots \cup kS| \leq (|S| + 1)^k \leq s^k = s^{n/(2 \log s)} = 2^{n/2}$ , so the probability that the  $S$ -degree of  $\ell_\rho$  being at most  $k$  is no more than  $2^{n/2}/2^n = 2^{-n/2}$ . ◀

We will call a  $y$ -restriction  $\rho$  *good* (for circuit  $C''$ ) if the  $S$ -degree of  $\ell_\rho$  is larger than  $n/(2 \log s)$  and *bad* otherwise. Therefore a random  $\rho$  is bad with probability at most  $2^{-n/2}$ . Let  $N_\rho : \{0, 1\}^n \rightarrow \mathbb{N}$  be the function that counts the number of AND gates of  $C''|_\rho$  that are 1.

► **Lemma 11.** Let  $S \subseteq \mathbb{F}_2^n$  be the set of input parities (or their negations) of  $C''|_\rho$ . Suppose  $\ell_\rho$  has  $S$ -degree larger than  $k$  and each AND gate in  $C''|_\rho$  has fan-in at most  $w$ , then  $N_\rho^i$  is uncorrelated with  $\ell_\rho$  for  $i = 1, 2, \dots, k/w$ . In other words,  $\mathbf{E}_x(N_\rho^i(x) \mid \ell_\rho(x) = 0) = \mathbf{E}_x(N_\rho^i(x) \mid \ell_\rho(x) = 1)$  for  $i = 1, 2, \dots, k/w$ .

The above lemma is proved in Appendix B.2. Since each of the AND gates in  $C''$  has fan-in at most  $4 \log n$  and the  $S$ -degree of  $\ell_\rho$  is larger than  $n/(2 \log s)$  for every good  $\rho$ , Lemma 11 implies that  $N_\rho^i$  is uncorrelated with  $\ell_\rho$  for  $i$  up to  $d := n/(8 \log n \log s)$  for every good  $y$ -restriction.



### 2.3 Linear programming and feasible solutions based on Chebyshev polynomials (Proof of Lemma 7)

Let  $X_\rho$  (resp.  $Y_\rho$ ) be the (conditional) random variable of  $N_\rho(x) \mid (\ell_\rho(x) = 1)$  (resp.,  $N_\rho(x) \mid (\ell_\rho(x) = 0)$ ). Our key observation is that, by Lemma 11, these two random variables both take values in  $\{0, 1, \dots, s'\}$  and their moments match up to  $n/8 \log n \log s$ . So intuitively, if  $s'$  is not too large, these two random variables should have close to identical distributions; in particular, we should have  $\Pr(X_\rho = 0) \approx \Pr(Y_\rho = 0)$ . Since  $C'$  (and thus,  $C''$ ) computes IP with only one-sided error, we have that for *every*  $y$ -restriction  $\rho$ ,  $\Pr(C''|_\rho(x) = 1 \mid \ell_\rho(x) = 1) = 1$  and consequently  $\Pr(X_\rho = 0) = 0$ . Combining this with the consequence of moment-matching condition between  $X_\rho$  and  $Y_\rho$  implies that  $\Pr(Y_\rho = 0) \approx 0$  for every good  $\rho$ .

Fix a good  $y$ -restriction  $\rho$ . The following key lemma provides the desired upper bound on  $\Pr(Y_\rho = 0) = \Pr_x(C''|_\rho(x) = 0 \mid \ell_\rho(x) = 0)$ . The lemma allows an additional parameter  $\xi_\rho$  which in our application is set to zero (since we have  $\Pr_x(C''|_\rho(x) = 0 \mid \ell_\rho(x) = 1) = 0$ ). However, since the lemma applies to general random variables with matching moments and may be of independent interest, it is stated in the more general form.

► **Lemma 12.** *Let  $X_\rho$  and  $Y_\rho$  be random variables supported on  $\{0, 1, \dots, s'\}$  such that (i)  $\mathbf{E}(X_\rho^i) = \mathbf{E}(Y_\rho^i)$  for  $i = 1, \dots, d$ ; and (ii)  $\Pr(X_\rho = 0) = \xi_\rho$ . Then  $\Pr(Y_\rho = 0) \leq \xi_\rho + 4(1 - \xi_\rho)e^{-d/\sqrt{2s'}}$ .*

**Proof.** We set up a linear program to maximize  $\Pr(Y_\rho = 0)$  over the choices of random variables  $X_\rho$  and  $Y_\rho$ . The variables in the LP are  $x_i$  and  $y_i$  where  $x_i = \Pr(X_\rho = i)$  and  $y_i = \Pr(Y_\rho = i)$ . Aside from nonnegativity and an upper bound constraint for  $x_0$ , we have  $d + 2$  equality constraints; 2 of them to force  $X_\rho$  and  $Y_\rho$  to have probability distributions, and the other  $d$  for the moment matching condition. The linear program and the corresponding dual are listed in Fig. 1. In order to upper bound the value of the primal program (i.e.,  $\Pr(Y_\rho = 0)$ ) and prove Lemma 12, it suffices to find a feasible solution to the corresponding dual program. We show that by choosing the polynomial  $p$  in the dual to be a Chebyshev polynomial (appropriately shifted and scaled), an essentially optimal bound on the primal value can be found. Full details are included in Appendix B.3. ◀

Considering the overwhelming fraction  $1 - 2^{-n/2}$  of good  $\rho$ 's and averaging on  $\rho$ , using Lemma 12 above we get that  $\Pr_{x,y}(C''(x,y) = 0 \mid \text{IP}(x,y) = 0) \approx 0$ . On the other hand, since  $C'$  is an  $\epsilon$ -discriminator for IP, then  $C''$  is an  $\epsilon'$ -discriminator for IP for some  $\epsilon' \geq \epsilon - \tau$  (where we recall  $\tau = \Pr_{x,y}(C'(x,y) \neq C''(x,y))$ ). Therefore  $\epsilon'$  must be small when the circuit size  $s$  of  $C'$  is small, and thus we obtain the desired upper bound on  $\epsilon$ . We include the remaining details of the proof of Lemma 7 (as sketched above) in Appendix B.4.

### 2.4 Limitations of our approach

We remark that the  $\tilde{\Omega}(n^2)$  lower bound is optimal (up to a polylogarithmic factor) for our current approach. This follows from a theorem of Paturi [19], which states that if  $p(x)$  is a degree  $d$  polynomial such that  $0 \leq p(i) \leq 1$  for  $i = 0, 1, \dots, s$  and  $|p(1) - p(0)| \geq c$  for some constant  $c$ , then  $d = \Omega(\sqrt{s})$ , or equivalently  $s = O(d^2)$ . Since in our setting  $d = \Theta(n/\log n \log s)$ , the best lower bound one can show in the current framework is  $\tilde{O}(n^2)$ .



$\begin{aligned} \text{Max } & y_0 && \text{(primal LP)} \\ \text{s.t. } & \sum_{i=0}^{s'} i^j x_i - \sum_{i=0}^{s'} i^j y_i = 0, \quad j = 1, \dots, d \\ & \sum_{i=0}^{s'} x_i = 1 \\ & \sum_{i=0}^{s'} y_i = 1 \\ & x_0 = \xi_\rho \\ & x_i, y_i \geq 0, \quad i = 0, \dots, s' \end{aligned}$
$\begin{aligned} \text{Min } & 1 - (1 - \xi_\rho)z && \text{(dual LP)} \\ \text{s.t. } & p \text{ is a polynomial of degree at most } d \\ & p(0) = 0 \\ & z \leq p(i) \leq 1 \quad i = 1, \dots, s' \end{aligned}$

■ **Figure 1** Primal LP for finding maximum  $\Pr(Y = 0)$  (top), and the final dual LP for finding maximum  $\Pr(Y = 0)$  (bottom).

### 3 Superlinear Lower Bound for General Circuits

In this section we prove the following superlinear lower bound for  $\text{AC}^0 \circ \text{MOD}_2$  circuits of arbitrary depth. Throughout this section we find it more convenient to use  $(x_1, \dots, x_n)$  as the entire input to IP rather than the two-input notation  $(x_1, \dots, x_n, y_1, \dots, y_n)$  used previously. We remark that the results of this section hold for a more general class of functions than IP, namely *bent functions*<sup>4</sup>. We state the results here for IP, and prove them for bent functions in the appendix.

► **Theorem 13.** *If  $C$  is an  $\text{AC}^0 \circ \text{MOD}_2$  circuit of depth  $k$  and size  $S$  that computes IP:  $\{0, 1\}^n \rightarrow \{0, 1\}$ , then  $S = \Omega(n^{1+4^{-k}})$ .*

#### Deterministic restrictions.

The high level idea of the proof is to adapt the technique of “deterministic restrictions” [3] to  $\text{AC}^0 \circ \text{MOD}_2$  circuits. In contrast to *random restrictions* which simplify circuits probabilistically, deterministic restrictions aim to show that, if the circuit size is small, then one can find a (small) set of input variables *deterministically* based on the structure of the circuit, such that fixing them forces the circuit to output a constant. This implies that small circuits fail to compute functions that cannot be made constant without setting a large number of input variables. The only twist when applying this framework to  $\text{AC}^0 \circ \text{MOD}_2$  circuits is, instead of fixing independent input variables, one now fixes linear functions which in general are no longer independent. We use a folklore result (called the *Folk Lemma* below) that IP can not be made constant by imposing less than  $n/2$  linear constraints on the inputs; i.e., IP is not constant on a linear subspace of dimension more than  $n/2$  (see Appendix D.1).

<sup>4</sup> A Boolean function is bent if all its Fourier coefficients are equal in magnitude. The Inner Product function (IP) is a special case (see Appendix A.2). In fact, our result holds for any function whose Fourier coefficients are all exponentially small in magnitude.

The main ingredient of the proof is the following lemma, which is the exact analogue of a result of Chaudhuri and Radhakrishnan [3] for  $AC^0$  circuits.

► **Lemma 14.** *Let  $C(x)$  be an  $AC^0 \circ MOD_2$  circuit of depth  $k$  and size  $S$ , with variable inputs  $x_1, \dots, x_n$  and bottom parity gates  $p_1(x), \dots, p_r(x)$ . Then there exists a set of  $t$  linearly independent linear restrictions,  $t < 5S^{1-4^{-k}}$ , such that imposing them on  $x_1, \dots, x_n$  makes  $C(x)$  constant (on the restricted space).*

**Proof.** We will adapt the argument of [3]. The algorithm of [3] constructs a partial assignment to the inputs of an  $AC^0$  circuits so that the output is fixed and the number of fixed variables is small. In particular, it fixes the values of gates at each level (by fixing the bottom variables and propagating the values up the circuit), starting at level 0 (the input level), and proceeding successively up to the output gate at level  $k$ . The specific way of fixing these gates ensures that after level  $i$  is fixed, all gates at levels  $j \leq i$  have both small fan-in and small fan-out (fan-ins and fan-outs are defined with respect to the current partial restriction and gates that are not fixed yet). At the end of such fixing, a so-called “regular” circuit is obtained. Then it is straightforward to show that one can fix an additional small number of variables of such regular circuit to make it output a constant. Our argument proceeds in an almost identical way. However, we fix *parities* in addition to input variables, and once a new parity gate is fixed, we need to fix the free parity gates which linearly depend on the fixed parity gates. This can only possibly reduce the number of parity gates needed to be fixed in the process, thus the original proof works in the setting of  $AC^0 \circ MOD_2$  circuits as well. We defer the details to Appendix D.2. ◀

Now we are ready to prove the main theorem of this section.

**Proof of Theorem 13.** Suppose  $C$  has size  $S < \frac{1}{5}n^{1+4^{-k}}$  (hence, it has at most that many parity gates) and computes the IP function. By Lemma 14, there exists a set of linearly independent linear restrictions of size at most  $5S^{1-4^{-k}} < (n^{1+4^{-k}})^{1-4^{-k}} = n^{1-16^{-k}} < n/2$  (for large enough  $n$ ), under which  $C$  becomes a constant function. But by the Folk Lemma, we must impose at least  $n/2$  linear restrictions to make IP a constant; a contradiction. ◀

---

## References

- 1 A. Akavia, A. Bogdanov, S. Guo, A. Kamath, and A. Rosen. Candidate weak pseudorandom functions in  $AC^0 \circ MOD_2$ . In *Proc. 5th ITCS*, pages 251–260, 2014.
- 2 T. Ando. Truncated moment problems for operators. *Acta Sci. Math. (Szeged)*, 31:319–334, 1970.
- 3 S. Chaudhuri and J. Radhakrishnan. Deterministic restrictions in circuit complexity. In *Proc. 28th STOC*, pages 30–36, 1996.
- 4 G. Cohen and I. Shinkar. The complexity of DNF of parities. To appear in ITCS’16, 2016.
- 5 R.E. Curto and L.A. Fialow. Recursiveness, positivity, and truncated moment problems. *Houston J. Math.*, 17:603–635, 1991.
- 6 A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *JCSS*, 46:129–154, 1993. Earlier version in FOCS’87.
- 7 R. Impagliazzo, R. Shaltiel, and A. Wigderson. Reducing the seed length in the Nisan-Wigderson generator. *Combinatorica*, 26(6):647–681, 2006. Earlier version in FOCS’01.
- 8 J.C. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *JCSS*, 55(3):414–440, 1997.
- 9 S. Jukna. On graph complexity. *Combinatorics, Probability, and Computing*, 15(6):855–876, 2006.

- 10 A. Klivans and L. Fortnow. Efficient learning algorithms yield circuit lower bounds. *JCSS*, 75:27–36, 2009.
- 11 A. Klivans, P. Kothari, and I.C. Oliveira. Constructing hard functions using learning algorithms. In *Proc. 28th CCC*, pages 86–97, 2013.
- 12 A. Klivans and R. Meka. Moment-matching polynomials. Technical Report TR13-008, ECCO, 2013.
- 13 S. Kopparty and S. Srinivasan. Certifying polynomials for  $AC^0(\oplus)$  circuits, with applications. In *Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, pages 36–47, 2012.
- 14 N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993.
- 15 N. Linial and N. Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990.
- 16 N. Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.
- 17 N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994. Earlier version in STOC’92.
- 18 N. Nisan and A. Wigderson. Hardness versus randomness. *JCSS*, 49:149–167, 1994.
- 19 R. Paturi. On the degree of polynomials that approximate symmetric Boolean functions. In *Proc. 24th STOC*, pages 468–474, 1992.
- 20 S. Raskhodnikova, D. Ron, A. Shpilka, and A. Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. *SIAM J. Comput.*, 39(3):813–842, 2009. Earlier version in FOCS’07.
- 21 A.A. Razborov. Lower bounds on the size of bounded-depth networks over a complete basis with logical addition. *Math. Notes Acad. of Sci. USSR*, 41(4):333–338, 1987.
- 22 T. Rivlin. *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory*. John Wiley & Sons, Inc., 1990.
- 23 R. Servedio and E. Viola. On a special case of rigidity. Manuscript, 2012.
- 24 R. Shaltiel and C. Umans. Simple extractors for all min-entropies and a new pseudo-random generator. *J. ACM*, 52(2):172–216, 2005. Earlier version appeared in FOCS’01.
- 25 R. Shaltiel and E. Viola. Hardness amplification proofs require majority. *SIAM J. Comput.*, 39(7):3122–3154, 2010.
- 26 R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proc. 19th STOC*, pages 77–82, 1987.
- 27 C. Umans. Pseudo-random generators for all hardnesses. *JCSS*, 67(2):419–440, 2003. Earlier version appeared in CCC’02.
- 28 L.G. Valiant. Graph-theoretic arguments in low-level complexity. In *Proc. 6th MFCS*, volume 53 of *LNCS*, pages 162–176. Springer, 1977.
- 29 I. Volkovich. On learning, lower bounds and (un)keeping promises. In *ICALP 2014, Part I*, volume 8572 of *LNCS*, pages 1027–1038. Springer, 2014.

## **A** Notation and Additional Preliminaries

### **A.1** Notation

All logarithms in this paper are to the base 2. Let  $n \geq 1$  be a natural number. We use  $[n]$  to denote the set  $\{1, \dots, n\}$ . We use  $\mathbb{F}_2$  for the field with 2 elements  $\{0, 1\}$ , where addition and multiplication are performed modulo 2. We view elements in  $\mathbb{F}_2^n$  as  $n$ -bit binary strings – that is elements of  $\{0, 1\}^n$  – alternatively. If  $x$  and  $y$  are two  $n$ -bit strings, then  $x + y$  (or  $x - y$ ) denotes bitwise addition (i.e. XOR) of  $x$  and  $y$ . We view  $\mathbb{F}_2^n$  as a vector space equipped with

an inner product  $\langle x, y \rangle$ , which we take to be the standard dot product:  $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ , where all operations are performed in  $\mathbb{F}_2$ .

Often times, it is convenient to switch the range of Boolean functions between  $\{0, 1\}$  and  $\{-1, 1\}$ . We use  $f^\pm$  to denote the  $\{-1, 1\}$ -valued Boolean function corresponding to  $f$ . They are related by  $f^\pm = (-1)^f = 1 - 2f$  and  $f = (1 - f^\pm)/2$ .

For every  $\alpha \in \mathbb{F}_2^n$ , one can define a *linear function* (or *parity function*) mapping  $\mathbb{F}_2^n$  to  $\{0, 1\}$  as  $\ell_\alpha(x) = \langle \alpha, x \rangle$ . Let  $\chi_\alpha = (-1)^{\ell_\alpha}$ , which are commonly known as *characters*.

Characters play a central role in *Fourier analysis* of Boolean functions. Background on Fourier analysis is reviewed next.

## A.2 Background on Fourier analysis

For functions  $f, g: \mathbb{F}_2^n \rightarrow \mathbb{C}$  the inner product is defined as  $\langle f, g \rangle := \mathbf{E}_{x \in \mathbb{F}_2^n}(f(x)g(x))$ . For  $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{F}_2^n$ , the corresponding character function  $\chi_\alpha$  is defined as  $\chi_\alpha(x_1, \dots, x_n) = \prod_{i: \alpha_i=1} (-1)^{x_i}$ . For  $\alpha, \beta \in \mathbb{F}_2^n$ , the inner product between  $\chi_\alpha$  and  $\chi_\beta$  is 1 if  $\alpha = \beta$ , and 0 otherwise. Therefore the characters form an orthonormal basis for complex-valued functions over  $\mathbb{F}_2^n$ , and we can expand any  $f$  defined on  $\mathbb{F}_2^n$  using  $\{\chi_\alpha\}_{\alpha \in \mathbb{F}_2^n}$  as a basis.

► **Definition 15 (Fourier Transform).** Let  $f: \mathbb{F}_2^n \rightarrow \mathbb{C}$ . The *Fourier transform*  $\hat{f}: \mathbb{F}_2^n \rightarrow \mathbb{C}$  of  $f$  is defined to be  $\hat{f}(\alpha) = \mathbf{E}_x(f(x)\chi_\alpha(x))$ . The quantity  $\hat{f}(\alpha)$  is called the *Fourier coefficient* of  $f$  at  $\alpha$ .

The Fourier inversion formula is given by  $f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \hat{f}(\alpha)\chi_\alpha(x)$ , and the Parseval's identity is  $\sum_{\alpha \in \mathbb{F}_2^n} \hat{f}(\alpha)^2 = \mathbf{E}_x(f(x)^2)$ .

A Boolean function  $f: \mathbb{F}_2^n \rightarrow \{0, 1\}$  is called a *bent function* if all the Fourier coefficients of  $f^\pm := (-1)^f$  have the same absolute value. That is,  $|\hat{f}^\pm(\alpha)| = 2^{-n/2}$  for every  $\alpha \in \mathbb{F}_2^n$ . It is well known that the Inner Product function IP is a bent function.

## B Omitted details of Section 2

### B.1 Fixing codimensions

The following claim shows that we can ensure that each AND gate at the bottom layer has fan-in equal to its codimension.

► **Claim 16.** For any AND gate in  $C'$ , there is an equivalent AND of a subset of its inputs with fan-in equal to its codimension.

**Proof.** Consider the input wires of the AND gate in any order. We say that the  $i^{\text{th}}$  input is *redundant* if, given that the first  $i-1$  input wires are 1, then in the subspace the  $i^{\text{th}}$  input wire must also be 1. Notice that eliminating the redundant wires yields an equivalent function. To see that the fan-in of this new gate is equal to the codimension of the original AND gate, consider the dimension of the coset of inputs that make the first  $i$  inputs 1. Observe that each with non-redundant input, the dimension decreases by one; so, the codimension of the original gate equals the number of non-redundant inputs, which is precisely the fan-in of this new AND gate. ◀

### B.2 Proof of Lemma 11

For convenience, we switch to the  $\{-1, 1\}$  representation of Boolean values for parities, i.e.  $\chi(x) = (-1)^{\ell(x)}$ . Let  $\chi_1, \dots, \chi_{s'}$  be the input parities of  $C''|_\rho$ , and let  $f'_1, f'_2, \dots, f'_{t'}$

(each still taking value in  $\{0, 1\}$ ) be the functions computed by the AND gates in  $C''|_\rho$ . Then  $N_\rho(x) = f'_1(x) + f'_2(x) + \dots + f'_{t'}(x)$ . Note that since each  $f'_j(x)$  is the AND of at most  $w$  parities from  $\{\chi_1, \dots, \chi_{s'}\}$ ,  $f'_j(x)$  can be expressed as a polynomial of degree at most  $w$  with  $\chi_1, \dots, \chi_{s'}$  as variables (indeed, if  $f'_1(x) = \text{AND}(\chi_1(x), \dots, \chi_w(x))$ , then  $f'_1 = (\frac{1-\chi_1}{2}) \dots (\frac{1-\chi_w}{2})$ ). Consequently,  $N_\rho^i$  is a polynomial of degree at most  $i \cdot w$  in  $\chi_1, \dots, \chi_{s'}$ . Now because  $\ell_\rho$  is of  $S$ -degree larger than  $k \geq i \cdot w$  for  $i = 1, 2, \dots, k/w$ , we have that  $\ell_\rho$  is not in the support of the polynomial representation of  $N_\rho^i$ . Finally, by the orthogonality of parities, letting  $\chi_\rho(x) := (-1)^{\ell_\rho(x)}$ , we have

$$0 = \langle N_\rho^i, \ell_\rho \rangle = \mathbf{E}_x(N_\rho^i(x) \cdot \ell_\rho(x)) = \frac{1}{2} (\mathbf{E}_x(N_\rho^i(x) \mid \chi_\rho(x) = 0) - \mathbf{E}_x(N_\rho^i(x) \mid \chi_\rho(x) = 1)).$$

### B.3 Omitted details of proof of Lemma 12

Denote by  $P_\rho$  the value of the primal LP in Fig. 1. The dual linear program is

$$\begin{aligned} \text{minimize} \quad & z_{d+1} + z_{d+2} + \xi_\rho z_{d+3} \\ \text{such that} \quad & z_{d+1} + z_{d+3} \geq 0 \\ & z_{d+2} \geq 1 \\ & (\sum_{j=1}^d i^j z_j) + z_{d+1} \geq 0 \quad i = 1, \dots, s' \\ & (\sum_{j=1}^d -i^j z_j) + z_{d+2} \geq 0 \quad i = 1, \dots, s' \end{aligned}$$

We can interpret the dual as a problem involving polynomials. The feasible solutions correspond to coefficients of degree- $d$  polynomials  $p(x) = \sum_{j=1}^d z_j x^j$  with  $p(0) = 0$ . By duality, the objective value of the dual is nonnegative for any feasible solution. Thus, by scaling, we can assume  $z_{d+2} = 1$ . Further, since  $z_{d+3}$  only appears in the first constraint in this minimization problem, we can always take  $z_{d+3} = -z_{d+1}$ .

Rearranging the last two constraints of this problem yields that the values  $\{p(1), p(2), \dots, p(s')\}$  must all lie in the interval  $[-z_{d+1}, z_{d+2}]$ . Setting  $z = -z_{d+1}$ , the dual problem can be rephrased as the final Dual LP showed in Fig. 1.

Denote by  $D_\rho$  the value of this dual LP. By the Strong Duality Theorem,  $P_\rho = D_\rho$ , and therefore if  $V(p)$  is the value of any feasible solution corresponding to a polynomial  $p$  to the dual LP, we have

$$\Pr(Y_\rho = 0) \leq P_\rho = D_\rho \leq V(p).$$

The above modified problem about polynomials is strikingly similar to the problem of approximating OR functions by low-degree polynomials, for which Nisan and Szegedy gave an optimal solution based on Chebyshev polynomials [17]. Recall that Chebyshev polynomial (of the first kind)  $T_k(x)$  is a degree  $k$  polynomial defined by  $T_k(x) = \cos(k \arccos(x))$ , or more explicitly

$$T_k(x) = \frac{1}{2} \left[ \left( x + \sqrt{x^2 - 1} \right)^k + \left( x - \sqrt{x^2 - 1} \right)^k \right].$$

It is well-known that  $-1 \leq T_k(x) \leq 1$  for all  $x \in [-1, 1]$  and  $T_k(x) > 1$  when  $x > 1$ . For a detailed treatment of Chebyshev polynomials see e.g. [22].

We now construct a dual feasible polynomial  $p$  based on Chebyshev polynomials. Define

$$q(x) = 1 - \frac{T_d(\frac{s'-x}{s'-1})}{T_d(\frac{s'-1}{s'-1})},$$

**23:14**  $AC^0 \circ MOD_2$  lower bounds for the Boolean IP

and let

$$p(x) = \frac{q(x)}{\max_{i \in \{1, \dots, s'\}} q(i)}$$

Clearly  $p(x)$  is a degree  $d$  polynomial,  $p(0) = 0$  and  $p(i) \leq 1$  for  $i = 1, \dots, s'$ , hence a feasible solution to the dual LP.

► **Claim 17.** The value of  $p(x)$  with respect to the dual LP satisfies that  $D(p) \leq \xi_\rho + \frac{2(1-\xi_\rho)}{T_d(1+\frac{1}{s'})}$ .

**Proof.** Since  $-1 \leq T_d(w) \leq 1$  for all  $-1 \leq w \leq 1$ , then for  $i = 1, \dots, s'$ ,

$$\begin{aligned} p(i) &= \frac{q(x)}{\max_{i \in \{1, \dots, s'\}} q(i)} \\ &= \frac{T_d(1 + \frac{1}{s'-1}) - T_d(\frac{s'-i}{s'-1})}{T_d(1 + \frac{1}{s'-1}) - \min_{j \in [s']} T_d(\frac{s'-j}{s'-1})} \\ &\geq \frac{T_d(1 + \frac{1}{s'-1}) - 1}{T_d(1 + \frac{1}{s'-1}) + 1} = \frac{1 - 1/T_d(1 + \frac{1}{s'-1})}{1 + 1/T_d(1 + \frac{1}{s'-1})} \\ &\geq 1 - \frac{2}{T_d(1 + \frac{1}{s'-1})} \geq 1 - \frac{2}{T_d(1 + \frac{1}{s'})}. \end{aligned}$$

Therefore the value  $z$  in the objective function of dual LP is at least  $z \geq 1 - \frac{2}{T_d(1+\frac{1}{s'})}$  and the claim follows. ◀

We will need the following two inequalities bounding  $T_k(x)$ 's growth when  $x \geq 1$ .

► **Claim 18.** For any nonnegative integer  $k$ , we have<sup>5</sup>

1.  $T_k(1 + \mu) \geq \frac{1}{2}e^{(\sqrt{2\mu+\mu^2})k/2}$  for all real number  $0 \leq \mu \leq 1$ .
2.  $T_k(1 + \mu) \leq e^{2(\sqrt{2\mu+\mu^2})k}$  for all  $\mu \geq 0$ .

**Proof.** For the first part, using that  $1 + x \geq e^{x/2}$  for  $0 \leq x \leq 2$ , we obtain

$$\begin{aligned} T_k(1 + \mu) &\geq \frac{1}{2}(1 + \mu + \sqrt{2\mu + \mu^2})^k \\ &\geq \frac{1}{2}(1 + \sqrt{2\mu + \mu^2})^k \\ &\geq \frac{1}{2}e^{(\sqrt{2\mu+\mu^2})k/2}, \end{aligned}$$

for all  $0 \leq \mu \leq 1$ .

For the second part, by the standard inequality  $(1 + t/n)^n \leq e^t$  for all nonnegative  $t$  and  $n$ ,

$$\begin{aligned} T_k(1 + \mu) &\leq (1 + \mu + \sqrt{2\mu + \mu^2})^k \\ &\leq (1 + 2\sqrt{2\mu + \mu^2})^k \leq e^{2(\sqrt{2\mu+\mu^2})k}. \end{aligned}$$

◀

Finally, by setting  $\mu = 1/s'$  in the first inequality of Claim 18, we have  $T_d(1 + 1/s') \geq \frac{1}{2}e^{(\sqrt{2/s'+1/s'^2})d/2} \geq \frac{1}{2}e^{\sqrt{d^2/2s'}}$ . Combining this with Claim 17, we get

$$\Pr(Y_\rho = 0) \leq \xi_\rho + 4(1 - \xi_\rho)e^{-d/\sqrt{2s'}}$$

which completes the proof of Lemma 12.

---

<sup>5</sup> The second inequality also appeared in [19].

## B.4 Omitted details of the proof of Lemma 7

Recall that  $X_\rho$  (resp.  $Y_\rho$ ) is the (conditional) random variable of  $N_\rho(x) \mid (\ell_\rho(x) = 1)$  (resp.,  $N_\rho(x) \mid (\ell_\rho(x) = 0)$ ), where  $N_\rho(x)$  counts the number of AND gates in  $C''|_\rho$  that evaluate to 1. Since  $C''|_\rho$  provides a one-sided approximation of the function  $\ell_\rho$ , we have that

$$\xi_\rho := \Pr_x(N_\rho(x) = 0 \mid \ell_\rho(x) = 1) = 0.$$

Taking  $d = \frac{n}{8 \log n \log s}$  and  $s' < s$  into Proposition 12, we have that for any good  $y$ -restriction  $\rho$ ,

$$\begin{aligned} \Pr_x(C''|_\rho(x) = 0 \mid \ell_\rho(x) = 0) &= \Pr(Y_\rho = 0) \\ &\leq \xi_\rho + 4(1 - \xi_\rho) \exp(-d/2\sqrt{s}) \\ &= 4 \exp(-d/2\sqrt{s}). \end{aligned} \tag{1}$$

Taking into account bad  $\rho$ 's, which happens with probability at most  $2^{-n/2}$  (according to Claim 10), the discriminator parameter  $\epsilon'$  for  $C''(x, y)$  can now be upper bounded as

$$\begin{aligned} \epsilon' &= \Pr_{x, \rho}(C''|_\rho(x) = 0 \mid \ell_\rho(x) = 0) \\ &= \mathbf{E}_\rho(\Pr_x(C''|_\rho(x) = 0 \mid \ell_\rho(x) = 0)) = \mathbf{E}_\rho(Y_\rho) \\ &= \mathbf{E}_{\text{good } \rho}(Y_\rho) \Pr(\rho \text{ is good}) + \mathbf{E}_{\text{bad } \rho}(Y_\rho) \Pr(\rho \text{ is bad}) \\ &\leq 4 \exp(-d/2\sqrt{s}) + 2^{-n/2}. \end{aligned}$$

Finally, since  $\epsilon' \geq \epsilon - \tau$ , where we recall that  $\tau = \Pr_{x, y}(C'(x, y) \neq C''(x, y)) \leq s/n^4$ , the proof of Lemma 7 is complete.

## C Comparison to Linial-Nisan

In this appendix, we describe two connections between our work and the work by Linial and Nisan on approximate inclusion exclusion. First, we show that the bounds are actually equivalent: each bound can be recovered from the other. Second, we show that approximate inclusion-exclusion can be used *in place of* moment matching in the proof of our lower bound for depth-4 circuits. But, we stress that we believe that moment-matching is more natural in many contexts (including the proof of our lower bound), and is therefore of independent interest. Indeed, one may view our moment-matching bounds as giving a new interpretation of approximate inclusion-exclusion.

### C.1 Equivalence of moment matching and approximate inclusion-exclusion

Here, we show that our moment-matching technique can be recovered from Theorem 1 of [15].

► **Theorem 19.** *Let  $d$  and  $s'$  be integers and let  $A_1, A_2, \dots, A_{s'}$  and  $B_1, B_2, \dots, B_{s'}$  be two collections of arbitrary events in two probability spaces, where  $\Pr(B_i) > 0$  for at least one  $i$ . Further, assume that*

$$\Pr\left(\bigcap_{i \in S} A_i\right) = \Pr\left(\bigcap_{i \in S} B_i\right)$$



23:16  $AC^0 \circ MOD_2$  lower bounds for the Boolean IP

for every subset  $S \subset [s']$  with  $|S| \leq d$ . Then for  $d \geq \Omega(\sqrt{s'})$ , we have  $\Pr(\bigcup_{i=1}^{s'} A_i) / \Pr(\bigcup_{i=1}^{s'} B_i) = 1 + O(\exp(-2d/\sqrt{s'}))$ .

We show that our moment matching bound follows from the above theorem.

*Proof of Lemma 12 from Theorem 19:* Let  $X$  and  $Y$  be random variables supported on  $\{0, 1, 2, \dots, s'\}$  such that  $\mathbf{E}(X^j) = \mathbf{E}(Y^j)$  for  $1 \leq j \leq d$ . For  $0 \leq i \leq s'$ , let  $p_i := \Pr(X = i)$  and  $q_i := \Pr(Y = i)$ . Define two distributions  $P$  and  $Q$  over  $\{0, 1\}^{s'}$  such that

$$P(z) = \frac{p_{|z|}}{\binom{s'}{|z|}}, \text{ and } Q(z) = \frac{q_{|z|}}{\binom{s'}{|z|}},$$

where  $|z|$  is the Hamming weight of  $z$ .

Finally, for  $1 \leq i \leq s'$ , define the event  $A_i(z)$  (resp.  $B_i(z)$ ) to be the event that the  $i^{\text{th}}$  bit of a random string  $z$  drawn from  $\{0, 1\}^{s'}$  according to distribution  $P$  (resp.  $Q$ ) is 1.

Now the moment matching condition implies that

$$\sum_{w=1}^{s'} p_w w^j = \sum_{w=1}^{s'} q_w w^j,$$

for  $1 \leq j \leq d$ ; or if we let  $r_w := p_w - q_w$ , then

$$\sum_{w=1}^{s'} r_w w^j = 0,$$

for all  $1 \leq j \leq d$ . Viewing as vectors in the univariate polynomial vector space, the two sets of polynomials  $\{1, w, w^2, \dots, w^d\}$  and the linear span of polynomials  $\{\binom{w}{0}, \binom{w}{1}, \dots, \binom{w}{d}\}$  both form a basis for the linear space of polynomials of degree at most  $d$ . In particular, each of degree- $j$  polynomial  $\binom{w}{j}$  can be expressed as a linear combination of  $\{1, w, w^2, \dots, w^j\}$  for every  $1 \leq j \leq d$ , and therefore

$$\sum_{w=1}^{s'} r_w \binom{w}{j} = 0, \text{ or equivalently, } \sum_{w=1}^{s'} p_w \binom{w}{j} = \sum_{w=1}^{s'} q_w \binom{w}{j},$$

for every  $1 \leq j \leq d$ .

► **Claim 20.**  $\sum_{w=1}^{s'} p_w \binom{w}{j} = \sum_{S:|S|=j} \Pr(\bigcap_{i \in S} A_i) = \binom{s'}{j} \Pr(\bigcap_{i \in [j]} A_i)$ .

Notice, it follows from Claim 20 that for all  $S \subset [s']$  with  $|S| \leq d$ ,  $\Pr(\bigcap_{i \in S} A_i) = \Pr(\bigcap_{i \in S} B_i)$ .

*Proof of Claim 20:* Notice, the first quantity is

$$\sum_{w=1}^{s'} \binom{w}{j} \Pr_z(|z| = w) = \mathbf{E}_z \left( \sum_{S:|S|=j} I \left[ \bigcap_{i \in S} A_i(z) \right] \right) = \sum_{S:|S|=j} \mathbf{E}_z \left( I \left[ \bigcap_{i \in S} A_i(z) \right] \right)$$

by linearity of expectation. The first equality is now immediate. The second equality is because the  $A_i$ 's are symmetric events; given that exactly  $j$  of the  $A_i$ 's happen, all collections of  $j$  events that happened are equally likely. ◀

We can now invoke Theorem 19 to find that

$$\Pr\left(\bigcup_{i=1}^{s'} A_i\right) / \Pr\left(\bigcup_{i=1}^{s'} B_i\right) = 1 + O(\exp(-2d/\sqrt{s'})),$$

where, since by construction  $\Pr(\bigcup_{i=1}^{s'} A_i) = 1 - \Pr(X = 0)$  and  $\Pr(\bigcup_{i=1}^{s'} B_i) = 1 - \Pr(Y = 0)$ , we find that since

$$\frac{1}{1 + O(\exp(-2d/\sqrt{s'}))} \geq 1 - O(\exp(-2d/\sqrt{s'}))$$

$$(1 - O(\exp(-2d/\sqrt{s'})))(1 - \Pr(X = 0)) \leq 1 - \Pr(Y = 0)$$

$$\Pr(Y = 0) \leq \Pr(X = 0) + (1 - \Pr(X = 0)) \cdot O(\exp(-2d/\sqrt{s'}))$$

This is essentially the claimed form of our moment matching bound (cf. Lemma 12). ◀

We further note that it is also possible to derive a slightly weaker version of Theorem 19 from our moment matching bound; thus, the two are essentially equivalent.

*Proof of Approximate Inclusion-Exclusion from Lemma 12:* Consider  $X = \sum_{i=1}^{s'} I[A_i]$  and  $Y = \sum_{i=1}^{s'} I[B_i]$ . Then, for all  $t \leq d$ , we have

$$\mathbf{E}(X^t) = \mathbf{E} \left( \sum_{z \in [s']^t} \prod_{i=1}^t I[A_{z_i}] \right) = \sum_{z \in [s']^t} \Pr \left( \bigcap_{j: \exists i z_i = j} A_j \right) = \sum_{z \in [s']^t} \Pr \left( \bigcap_{j: \exists i z_i = j} B_j \right) = \mathbf{E}(Y^t)$$

since each  $\Pr \left( \bigcap_{j: \exists i z_i = j} A_j \right) = \Pr \left( \bigcap_{j: \exists i z_i = j} B_j \right)$  by the set-intersection conditions. Therefore, it follows from Lemma 12 that

$$\Pr(Y = 0) \leq \Pr(X = 0) + 4(1 - \Pr(X = 0)) \exp(-(d/\sqrt{2s'}))$$

By construction,  $1 - \Pr(X = 0) = \Pr \left( \bigcup_{i=1}^{s'} A_i \right)$  and  $1 - \Pr(Y = 0) = \Pr \left( \bigcup_{i=1}^{s'} B_i \right)$ , so we immediately have

$$\Pr \left( \bigcup_{i=1}^{s'} A_i \right) (1 - 4 \exp(-d/\sqrt{2s'})) \leq \Pr \left( \bigcup_{i=1}^{s'} B_i \right)$$

Noting that (when  $d/\sqrt{s'}$  is not too small)

$$\frac{1}{1 - 4 \exp(-d/\sqrt{2s'})} \leq 1 + O(\exp(-d/\sqrt{2s'}))$$

we find

$$\frac{\Pr \left( \bigcup_{i=1}^{s'} A_i \right)}{\Pr \left( \bigcup_{i=1}^{s'} B_i \right)} \leq 1 + O(\exp(-d/\sqrt{2s'}))$$

◀

## C.2 Alternative proof of the depth-4 bound via approximate inclusion-exclusion

In addition to the use of approximate inclusion-exclusion as an alternative proof of the moment-matching bound, we note that one of the corollaries to approximate inclusion-exclusion proved by Linial and Nisan [15] can be used to give an alternative proof of our almost-quadratic lower bound for depth-4 circuits. Specifically, Linial and Nisan obtained the following application of approximate inclusion-exclusion to Boolean circuits:

► **Theorem 21** (Theorem 5 of Linial and Nisan [15]). *Let  $f_1, f_2, \dots, f_s$  and  $g$  be Boolean functions such that for every  $S \subseteq \{1, \dots, s\}$ ,*

$$\left| \Pr \left[ \bigwedge_{i \in S} f_i(x) = g(x) \right] - \Pr \left[ \bigwedge_{i \in S} f_i(x) \neq g(x) \right] \right| \leq 2^{-t}$$

for  $t \geq \Omega(\sqrt{s} \log s)$ . Then

$$\left| \Pr \left[ \bigvee_{i=1}^s f_i(x) = g(x) \right] - \Pr \left[ \bigvee_{i=1}^s f_i(x) \neq g(x) \right] \right| \leq 2^{-\Omega(t/\sqrt{s} \log s)}$$

Let the functions  $f_1, \dots, f_s$  of Theorem 21 be the functions computed by the AND gates in the depth-3 approximator we obtain from the Discriminator Lemma (i.e., feeding in into the output-layer OR gate), and let  $g$  be the inner product function. Then ANDs of any subset of  $f_1, \dots, f_s$  is simply another AND gate, and we can obtain a sufficiently small bound on the advantage of and AND-of-parities at computing IP in order to apply Theorem 21 with  $t \approx \sqrt{s}$  poly  $\log(s)$ . The conclusion of Theorem 21 then establishes that the depth-3 approximator is sufficiently poorly correlated (has agreement less than  $1/\text{poly}(n)$ ) with IP to complete our argument when  $s = \tilde{O}(n^2)$ .

## D Omitted details of Section 3

### D.1 Linear restrictions of the Inner Product function

We now prove the folklore claim that IP cannot be made constant by imposing less than  $n/2$  linear restrictions on its inputs. Actually, we prove this claim for more general bent functions, thus proving a stronger statement. We also observe one could similarly show that if a Boolean function has all its Fourier coefficients bounded in magnitude by  $2^{-\Omega(n)}$ , then the function is not constant on a subspace of  $\mathbb{F}_2^n$  of co-dimension  $o(n)$ . Such a condition would also suffice to prove superlinear lower bounds in our framework.

#### Linear restrictions.

Let  $f: \mathbb{F}_2^n \rightarrow \{0, 1\}$  be a Boolean function. Define two sub-functions  $f_0$  (resp.,  $f_1$ ) mapping  $\mathbb{F}_2^{n-1}$  to  $\{0, 1\}$  as  $f_0(y) := f(0, y)$  (resp.,  $f_1(y) := f(1, y)$ ), where  $(z, y)$  denotes string concatenation of  $z$  and  $y$ . The function  $f_0$  (resp.,  $f_1$ ) is the result of the *restriction*  $x_1 = 0$  (resp.,  $x_1 = 1$ ). In other words, truth table of  $f_0$  and  $f_1$  are each restriction of truth table of  $f$  to an affine subspace of co-dimension 1. Such restrictions can be naturally generalized with respect to arbitrary affine constraints (rather than  $x_i = 0$  or  $x_i = 1$ ).

Let  $S$  be an affine subspace defined by a set of linearly independent affine constraints  $\ell_{\alpha_1}(x) = b_1, \dots, \ell_{\alpha_t}(x) = b_t$ , where  $\alpha_1, \dots, \alpha_t \in \mathbb{F}_2^n$  and  $b_1, \dots, b_t \in \{0, 1\}$ . Then, the sub-function resulting from these affine restrictions,  $f|_S(x)$  is a partial function defined by  $f|_S(x) := f(x)$  for all  $x \in S$ . Usually it is convenient to map the domain of such sub-functions to the Boolean hypercube. To this end, one can define an invertible linear transformation  $L: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$  such that  $L(\alpha_i) = e_i$  for  $i \in [t]$  (where  $e_i$  is the  $i$ th standard basis vector), and let  $(Lf)(x) := f(L(x))$  be the function  $f$  under the basis change defined by  $L$ . Under this change of basis, we see that an ordinary restriction of variable of the function  $Lf$  (i.e.,  $x_1 = b_1, \dots, x_t = b_t$ ) corresponds to the restriction of the original  $f$  to  $S$ . That is, the truth table of  $Lf$  under the restriction  $x_1 = b_1, \dots, x_t = b_t$  (that we denote by  $(Lf)_{b_1, \dots, b_t}$ ) would be the same as the truth table of  $f$  on  $S$ . Therefore, we can conveniently study sub-functions

resulting from linear restrictions by first applying a linear transformation to the input space  $x_1, \dots, x_n$  and then applying restriction in the ordinary sense (i.e., setting individual input bits) to the resulting function.

► **Lemma 22** (generalization of Folk Lemma). *Let  $f$  be a bent function and let  $\ell_{\alpha_1} = b_1, \dots, \ell_{\alpha_t} = b_t$  be a set of  $t < n/2$  linearly independent restrictions. If  $f|_S$  is the subfunction resulting from these linear restrictions, then  $f|_S$  is not a constant function.*

**Proof.** In an intuitive sense, the proof can be described as follows. We write down the polynomial representation of  $f$  in the  $\{-1, +1\}$  representation as defined by its Fourier transform. We note that a function is constant if and only if the Fourier coefficient corresponding to the empty set is either  $-1$  or  $+1$ . The restrictions can be applied to the polynomial representation and collapse some of the monomials into the constant term. Since the coefficient of each monomial is equal to  $2^{-n/2}$  in absolute value, and noting that the given restrictions collapse at most  $2^t$  of the monomials, the function can not accumulate enough mass on the Fourier coefficient for the empty set to reduce to the constant function, as long as  $t < n/2$ .

More formally, let  $L$  be any invertible linear transformation as defined above satisfying  $L(\alpha_1) = e_1, \dots, L(\alpha_t) = e_t$ . It suffices to show that  $g := (Lf)_{b_1 \dots b_t}$  is not constant, where  $g: \mathbb{F}_2^{n-t} \rightarrow \{0, 1\}$ . Since the Fourier spectrum of  $Lf$  satisfies  $\widehat{Lf}(\alpha) = \widehat{f}(L^{-1}\alpha)$  and  $L$  is invertible, hence  $Lf$  is also a bent function (the Fourier coefficients of  $Lf$  are simply a reordering of the Fourier coefficients of  $f$ ). We now switch to the  $\{-1, 1\}$ -representation of Boolean functions (letting  $\widehat{g}^\pm(x) := (-1)^{g(x)}$  and  $(\widehat{Lf})^\pm(x) := (-1)^{(Lf)(x)}$ ), and recall the well-known fact that for any  $\gamma \in \mathbb{F}_2^{n-t}$ ,

$$\widehat{g}^\pm(\gamma) = \sum_{\beta \in \mathbb{F}_2^t} (\widehat{Lf})^\pm(\beta, \gamma) \chi_\beta(b_1, \dots, b_t).$$

Since  $Lf$  is bent,  $|(\widehat{Lf})^\pm(\beta, \gamma)| = 2^{-n/2}$  for every  $\beta$  and  $\gamma$ ; and since  $t < n/2$ , it follows that  $|\widehat{g}^\pm(\gamma)| < 1$  for all  $\gamma \in \mathbb{F}_2^{n-t}$ . In particular,  $|\widehat{g}^\pm(0)| < 1$ . But if  $g$  is a constant Boolean function,  $\widehat{g}^\pm(0) = 1$  or  $-1$ . Therefore,  $g$  is not constant. ◀

## D.2 Omitted details of the proof of Lemma 14

The proof fixes the gates in the circuit from bottom up. Let  $d = S^{2 \cdot 4^{-k}}$ , and  $M = S^{4^{-k}}$ . Define a sequence of degrees  $d_0 = 0$ ,  $d_1 = d \geq 2$ , and  $d_{i+1} = d_i^4$ , for  $i \in [k]$ . Let  $\delta(g)$  be the number of input variables or parities that “influence” a free<sup>6</sup> gate  $g$ ; that is, the minimum set of variables or parity gates whose fixing suffices to set the value of  $g$  to either 0 or 1. As the proof proceeds and fixes various input variables or parity gates, the value of  $\delta(g)$  may reduce for each gate (for gates that are already fixed,  $\delta(g)$  is defined to be zero).

To fix a gate  $g$  at level  $i$ , we perform either a `FIXINDEGREE`( $g, i$ ) or a `FIXOUTDEGREE`( $g, i$ ) operation. If the indegree of gate  $g$  at level  $i$  is larger than  $d_i$  then `FIXINDEGREE`( $g, i$ ) fixes the gate as follows: If  $g$  is an AND gate it fixes one of the free gates feeding into it to 0. If  $g$  is an OR gate we fix a free gate feeding into it to 1. Note that at most  $\delta(g)$  input variables or parities are fixed this way.

If the outdegree of gate  $g$  at level  $i$  is larger than  $M \cdot \delta(g)$ , then `FIXOUTDEGREE`( $g, i$ ) fixes the gate as follows: If at least half the gates that  $g$  feeds into are OR gates it fixes the gate to 1, otherwise it fixes the gate to 0. This fixes a number of gates lower bounded by  $M/2$  times the number of bottom gates set.

<sup>6</sup> A free gate is one whose output is not fixed to a constant.

Note that in order to fix gates at higher levels we need to fix the free gates at the bottom level, and propagate them upwards. To ensure consistency, we maintain a set of bottom parities  $P$  (we view input variables  $x_1, \dots, x_n$  as parities as well) that have been already set (for example, if a bottom gate  $x_1 + x_3$  needs to be set to 0, we update the set  $P \leftarrow P \cup \{x_1 + x_3 = 0\}$ .) Once a new parity is added to the set  $P$ , we accordingly fix the values of bottom parities that are linear combinations of the parities in  $P$ . We then propagate the new gate values up the circuit, and then continue fixing gates in this consistent manner, using `FIXINDEGREE` and `FIXOUTDEGREE` of gates at increasing levels. Note that we only add to  $P$  linear constraints that are linearly independent, and thus consistency is always maintained.

We fix gates by sequentially fixing gates from the bottom level (i.e., the parity gates) to the output level (i.e., the output gate). First, the outdegrees of all gates at level  $i$  are fixed, then indegrees of all gates at level  $i + 1$ , and then outdegrees of all gates at level  $i + 1$ , and so forth until the output gate is reached.

We may now show that this procedure fixes only  $5S^{1-4^{-k}}$  bottom parity inputs, by repeating the computation from [3]. Let  $\sigma$  be the partial assignment to the variables (and parities) after completing the gate fixing steps and reaching the output gate (essentially this assignment is saved in the set of linear restrictions  $P$ ).

Note that the partial circuit obtained in the end has every gate  $g$  at level  $i$  of indegree at most  $d_i$ , and outdegree of each gate is at most  $M\delta(g)$ . Also note that the total number of bottom gates fixed in calls to `FIXOUTDEGREE`( $g, i$ ), for all  $i$ , is at most  $2S/M$ ; since the number of gates fixed is at least  $M/2$  times the number of bottom inputs fixed, which in turn is at most  $S$ .

Now we show that the number of bottom gates set during calls to `FIXINDEGREE`( $g, i$ ), for all  $i$ , is at most  $2SM/d$ . Note that since we first fix indegrees at level  $i - 1$  before fixing indegrees at level  $i$ , the number of variables and parities set while fixing a gate at level  $i$  is at most  $d_1 d_2 \cdots d_{i-1}$ . Similarly, since we fix outdegrees at level  $i - 1$  before fixing outdegrees at level  $i$ , the outdegree of any gate at level  $j < i$  is at most  $M(d_1 \cdots d_{i-1})$ . So the total number of gates at level  $i$  of degree larger than  $d_i$  is at most  $SM(d_1 \cdots d_{i-1})/d_i$ . Summing over all levels, the number of bottom gates set during calls to `FIXINDEGREE` is at most  $SM \sum_{i=1}^k (d_1 \cdots d_{i-1})^2/d_i$ . It can be verified that  $(d_1 \cdots d_{i-1})^2/d_i \leq 1/(2^{i-1}d)$ , which is at most  $SM \sum_{i=1}^k (d_1 \cdots d_{i-1})^2/d_i \leq 2SM/d$ .

To fix the output gate, we might need to fix at most an additional  $d_1 \cdots d_{k-1} < S^{1/2} < S^{1-4^{-k}}$  bottom gates or parities (since in the end, the indegree of each gate at level  $i$  is at most  $d_i$ ). Thus, overall it is enough to fix a total of  $2S/M + 2SM/d + S^{1-4^{-k}} = 5S^{1-4^{-k}}$  bottom gates in order to fix the final output of the circuit, and those are the inputs or parities collected in the set  $P$ .

### D.3 Generalization of Theorem 13 to bent functions

Using Lemma 22 and Lemma 14, the following theorem follows by the same argument from the proof of Theorem 13.

► **Theorem 23.** *If  $C$  is an  $AC^0 \circ MOD_2$  circuit of depth  $k$  and size  $S$  that computes a bent function  $f : \{0, 1\}^n \rightarrow \{0, 1\}$ , then  $S = \Omega(n^{1+4^{-k}})$ .*