Parity Helps to Compute Majority

Igor Carboni Oliveira
Department of Computer Science, University of Oxford, UK
igor.carboni.oliveira@cs.ox.ac.uk

Rahul Santhanam
Department of Computer Science, University of Oxford, UK
rahul.santhanam@cs.ox.ac.uk

Srikanth Srinivasan
Department of Mathematics, IIT Bombay, India
srikanth@math.iitb.ac.in

Abstract

We study the complexity of computing symmetric and threshold functions by constant-depth circuits with Parity gates, also known as $\mathsf{AC}^0[\oplus]$ circuits. Razborov [23] and Smolensky [25, 26] showed that Majority requires depth-d $\mathsf{AC}^0[\oplus]$ circuits of size $2^{\Omega(n^{1/2(d-1)})}$. By using a divide-and-conquer approach, it is easy to show that Majority can be computed with depth-d $\mathsf{AC}^0[\oplus]$ circuits of size $2^{\widetilde{O}(n^{1/(d-1)})}$. This gap between upper and lower bounds has stood for nearly three decades.

Somewhat surprisingly, we show that *neither* the upper bound nor the lower bound above is tight for large d. We show for $d \geq 5$ that any symmetric function can be computed with depth-d $\mathsf{AC}^0[\oplus]$ circuits of size $\exp(\widetilde{O}(n^{\frac{2}{3}\cdot\frac{1}{(d-4)}}))$. Our upper bound extends to threshold functions (with a constant additive loss in the denominator of the double exponent). We improve the Razborov-Smolensky lower bound to show that for $d \geq 3$ Majority requires depth-d $\mathsf{AC}^0[\oplus]$ circuits of size $2^{\Omega(n^{1/(2d-4)})}$. For depths $d \leq 4$, we are able to refine our techniques to get almost-optimal bounds: the depth-3 $\mathsf{AC}^0[\oplus]$ circuit size of Majority is $2^{\widetilde{\Theta}(n^{1/2})}$, while its depth-4 $\mathsf{AC}^0[\oplus]$ circuit size is $2^{\widetilde{\Theta}(n^{1/4})}$.

2012 ACM Subject Classification Theory of computation

Keywords and phrases Computational Complexity, Boolean Circuits, Lower Bounds, Parity, Majority

Digital Object Identifier 10.4230/LIPIcs.CCC.2019.23

Funding This work was supported in part by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2014)/ERC Grant Agreement no. 615075.

Acknowledgements This paper is the result of a collaboration that happened at the Simons Institute program on Lower Bounds in Computational Complexity, where all three authors were long-term visitors. We are grateful to the Simons Institute for their support.

1 Introduction

Given the difficulty of proving lower bounds for general Boolean circuits, much work in circuit complexity has focused on restricted classes, and in particular on bounded-depth classes. Super-polynomial lower bounds are known for explicit Boolean functions against various classes of bounded-depth circuit classes, including AC^0 (constant-depth circuits with unbounded fan-in AND and OR gates) and $\mathsf{AC}^0[\oplus]$ (constant-depth circuits with unbounded fan-in AND, OR and Parity gates).

In the case of AC^0, we have almost optimal size bounds [1, 11, 14] for the Parity function. A simple divide-and-conquer argument shows that Parity on n variables can be computed by depth-d AC^0 circuits of size $\widetilde{O}(2^{n^{1/(d-1)}})$. The classic lower bound of Håstad [14] shows that Parity requires depth-d AC^0 circuits of size $2^{\Omega(n^{1/(d-1)})}$. Thus the upper bound is tight up to the constant factor in the exponent. The same lower bound holds for the Majority

34th Computational Complexity Conference (CCC 2019).
Editor: Amir Shpilka; Article No. 23; pp. 23:1–23:17
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

function [14], but the upper bound given by the divide-and-conquer argument [8] weakens slightly to $2^{\widetilde{O}(n^{1/(d-1)})}$, meaning that the upper bound is tight up to a logarithmic factor in the exponent.

For $\mathsf{AC}^0[\oplus]$, however, we do not have optimal bounds. The celebrated polynomial approximation method of Razborov and Smolensky [23, 25, 26] yields a lower bound of $2^{n^{1/2(d-1)}}$ on the size of depth-d $\mathsf{AC}^0[\oplus]$ circuits computing Majority. The best known upper bound thus far for Majority was the one mentioned in the previous paragraph, which in fact gives constant-depth circuits that don't use Parity gates.

Note that there is a significant gap between upper and lower bounds for Majority – the exponent in the upper bound is quadratically larger than the exponent in the lower bound. This gap between upper and lower bounds has stood for almost three decades.

Since the best known upper bound for Majority can be implemented without Parity gates, a natural question arises. Do Parity gates help when computing Majority using bounded-depth circuits? It is easy to see that Majority gates help to compute Parity – indeed, Parity can be easily written as a small DNF of Majorities. However, it is far from clear how to take advantage of Parity to compute Majority. Indeed, we ourselves believed until recently that the upper bound was close to optimal for $\mathsf{AC}^0[\oplus]$ circuits computing Majority.

1.1 Our results

Our main result in this paper is that *neither* the upper bound using divide-and-conquer nor the lower bound given by the Razborov-Smolensky method is tight for Majority, when the depth is large enough. We first describe our new upper bound, and then our lower bound that slightly improves Razborov-Smolensky.

First, we show how to save a constant factor in the double exponent when computing Majority.

▶ **Theorem 1.** *Let $d \geq 5$ be an integer. Majority on n bits can be computed by depth-d $\mathsf{AC}^0[\oplus]$ circuits of size $2^{\widetilde{O}\left(n^{\frac{2}{3}\cdot\frac{1}{(d-4)}}\right)}$.*

Theorem 1 follows from a result giving the same upper bound for the $\mathsf{AC}^0[\oplus]$ size complexity of any *symmetric* function. Similar techniques combined with another idea and a more careful implementation allow us to obtain an improved upper bound at depth $d = 4$ (stated in Corollary 3 below). We also show how to extend the upper bound in Theorem 1 to any *linear threshold function*, though we lose a small additive term in the denominator of the double exponent. We refer to the body of the paper for more details about the latter result.

Next, we show how to improve the Razborov-Smolensky lower bound slightly to achieve a better double exponent.

▶ **Theorem 2.** *For any integer $d \geq 3$, Majority on n bits requires depth-d $\mathsf{AC}^0[\oplus]$ circuits of size $2^{\Omega(n^{1/(2d-4)})}$.*

Note that there is still a gap between our new upper bound for Majority given by Theorem 1 and our new lower bound given by Theorem 2. We do not have a clear belief at this point about what the optimal size bound should be at large depths.

For depths $d = 3$ and $d = 4$, our results do provide nearly optimal bounds. For $d = 3$, the new lower bound is close to the upper bound given by the divide-and-conquer strategy, showing that parity gates do no significantly help when $d = 3$. On the other hand, for depth $d = 4$ our improved upper bound construction essentially matches the new lower bound from Theorem 2.

▶ **Corollary 3.** *The following results hold:*

(i) *The depth-3* $\mathsf{AC}^0[\oplus]$ *circuit size complexity of Majority is* $2^{\widetilde{\Theta}(n^{1/2})}$.

(ii) *The depth-4* $\mathsf{AC}^0[\oplus]$ *circuit size complexity of Majority is* $2^{\widetilde{\Theta}(n^{1/4})}$.

Note the contrast with the known size bounds for AC^0 circuits, as a result of which we have a significantly stronger $\mathsf{AC}^0[\oplus]$ upper bound than an AC^0 lower bound at depth $d = 4$, but not at smaller depths.

Our results indicate that even for simple circuit models, naive upper bound strategies might not be optimal, and surprising savings can be achieved in circuit size. It might be worthwhile to look for other examples of this phenomenon.

1.2 Proof ideas

Upper bounds. We describe the upper bound for the Majority function (the same idea works for any symmetric function.) We follow the same basic high-level strategy as a construction of better approximating polynomials for the Majority function due to Alman and Williams [3]. They observed that while the Majority function on n variables seems hardest to compute when the Hamming weight is close to $n/2$, by polynomial interpolation, it is easy to obtain a low-degree polynomial that computes Majority on such inputs. Conversely, when the input has weight far from $n/2$, one can use sampling to reduce the input size and recurse.

A similar idea works in our setting as well. For inputs of weight within distance t of $n/2$, we use a degree-t polynomial to compute the Majority function. This polynomial is over $\mathbb{F}_2$ and moreover *symmetric*, and thus by standard techniques can be represented as a $\mathsf{AC}^0[\oplus]$ circuit of depth d and size roughly $\exp(t^{2/d} \log n)$. When the weight is t-far from $n/2$, we use sampling not to recurse but to solve the problem directly. In fact, using standard results on the complexity of the *Coin Problem* from the literature [21, 4], it follows that there are AC^0 circuits that solve Majority on inputs that are t-far from $n/2$ in depth d and size roughly $\exp((n/t)^{1/d})$. Putting these strategies together and optimizing the value of t yields the upper bound.

The above strategy yields a constant factor improvement in the double exponent of known upper bounds for large enough d. Using these ideas and a bit more work, we are also able to obtain a similar improvement at depth 4. In particular, all these upper bounds are stronger than known AC^0 *lower bounds* for symmetric functions [14], proving that parity gates indeed help in computing arbitrary symmetric functions.

It is worth understanding what these upper bounds mean at a higher level. A possible comparison can be made with the well-known result of Barrington, Beigel and Rudich [6], which showed that the OR function on n variables can be represented by a polynomial modulo 6 of degree just $O(\sqrt{n})$. The crucial observation there was that given any two distinct integers $i, j \in \{0, \ldots, n\}$, their difference $i - j$ cannot simultaneously be divisible by a large power of 2 and a large power of 3 (here, "large" means more than $\sqrt{n}$). This can be stated in the language of *p-adic norms*: recall that for a prime p, the p-adic distance between i and j is inversely related to the largest power of p that divides $i - j$. Thus, the result of [6] uses the fact that no i, j as above can be at small 2-adic as well as 3-adic distances. Our result leverages a similar contrast between the 2-adic norm and the standard Euclidean norm.

Lower Bounds. First we describe a new but weaker circuit size lower bound of $2^{\Omega(n^{1/(2d-3)})}$. Our proof follows the general polynomial approximation framework of Razborov [23]. The high-level idea is to show that any $\mathsf{AC}^0[\oplus]$ circuit of small size can be approximated by low-degree polynomials from $\mathbb{F}_2[x_1, \ldots, x_n]$: this is done by approximating each AND/OR gate[1] in the circuit by a low-degree polynomial and composing these approximations together. The second step is to show that the hard function (the Majority function in our scenario) does not have low-degree polynomial approximations of this form; here, the proofs that yield the best known parameters are due to Smolensky [25, 26].

To improve on known lower bounds, we need two new ingredients.

1. The first is the observation that the standard Razborov approximations for the OR and AND functions are *one-sided.* While this is obvious from the construction, we do not know of a previous lower-bound application of this. In our setting, we use this to show that any $\mathsf{AC}^0[\oplus]$ circuit C has a low-degree polynomial approximation P where the approximation is much better on one of $C^{-1}(0)$ or $C^{-1}(1)$.
2. To use these improved polynomial approximations, we need an improved lower bound for approximating the Majority function in the sense described above. It follows from Smolensky's work [25] that any polynomial that computes the Majority function on all but an ε-fraction of inputs must have degree $\Omega(\sqrt{n \log(1/\varepsilon)})$. In our setting, however, we need to lower bound the degree of a polynomial computing the Majority function on all but an ε-fraction of the 0-inputs but may err on a constant (say $1/10$) fraction of the 1-inputs. We are able to recover the lower bound of $\Omega(\sqrt{n \log(1/\varepsilon)})$ even under these weaker assumptions, which finishes the proof. This extension of Smolensky's lower bound uses results on the combinatorics of Hilbert functions [29, 15, 20].

Using some of the above ideas in conjunction with standard AC^0 lower bound techniques [1, 11, 14] based on random restrictions, we show how to get a lower bound of $\exp(\Omega(\sqrt{n}))$ on the size of depth-3 circuits for the Majority function, matching the known AC^0 lower bound [14] and nearly matching the AC^0 upper bound of $\exp(\tilde{O}(\sqrt{n}))$ [8].

Finally, we observe that the method of random restrictions employed in the depth-3 lower bound allows us to further improve the lower bound in the general case. To achieve that, we combine the refined analysis of approximate degree from items 1. and 2. above with the effects of a random restriction on the approximate degree of depth-2 subcircuits. This gives a $2^{\Omega(n^{1/(2d-4)})}$ lower bound on the depth-d $\mathsf{AC}^0[\oplus]$ circuit size of Majority, completing the proof of Theorem 2.

2 The Upper Bounds

2.1 An improved upper bound for all large depths

▶ **Theorem 4.** *For every integer $d \geq 5$, if $f_n\colon \{0,1\}^n \to \{0,1\}$ is a symmetric boolean function then it can be computed by an $\mathsf{AC}^0[\oplus]$ circuit of depth d and of size $2^{\widetilde{O}\left(n^{\frac{2}{3}\cdot\frac{1}{(d-4)}}\right)}$.*

Proof. For convenience, given a string $x \in \{0,1\}^*$, we let $|x|_1 \stackrel{\text{def}}{=} \sum_i x_i$ denote its hamming weight. For $0 \leq i, j \leq n$ and $i \neq j$, let $D_{i,j}$ and E_i be boolean functions on n-bit inputs satisfying

$$D_{i,j}(y) = \begin{cases} 1 & \text{if } |y|_1 = i, \\ 0 & \text{if } |y|_1 = j. \end{cases} \qquad E_i(y) = \begin{cases} 1 & \text{if } |y|_1 = i, \\ 0 & \text{otherwise.} \end{cases}$$

[1] The parity gates are already low-degree polynomials and hence trivially approximable in this sense.

(The behaviour of $D_{i,j}$ on inputs of different hamming weight is not relevant in our construction.) Notice that, for every $0 \le i \le n$,

$$E_i(y) = \bigwedge_{\substack{0 \le j \le n \\ j \neq i}} D_{i,j}(y).$$

Clearly, if $f_n \colon \{0,1\}^n \to \{0,1\}$ is a symmetric boolean function, then it can be written as a disjunction of at most $n+1$ functions E_i. (In other words, separating two layers of the hypercube can be as difficult as computing the hardest symmetric function.) Consequently, our task is reduced to the construction of $\mathsf{AC}^0[\oplus]$ circuits for each (partial) boolean function $D_{i,j}$.

Given $0 \le i, j \le n$ with $i \neq j$, we describe two circuits that agree with $D_{i,j}$ over the relevant input strings. The first circuit relies on an "algebraic" construction, while the second circuit is of a more "combinatorial" nature. Before going into further details, let us informally describe the main properties of the circuits (the discussion omits polylogarithmic factors in exponents).

1. Algebraic construction. If $|i-j| \le n^{1/4}$, there is an $\mathsf{AC}^0[\oplus]$ circuit for $D_{i,j}$ of depth d and size roughly $2^{n^{1/2d}}$. This circuit explores parity gates in a crucial way.
2. Combinatorial construction. If $|i-j| \ge n^{1/2}$, there is an AC^0 circuit for $D_{i,j}$ of depth d and size roughly $2^{n^{1/2d}}$. This circuit is obtained from AND/OR circuits solving the coin problem.
3. In the "critical" interval $n^{1/4} \le |i-j| \le n^{1/2}$, we still don't know if there are circuits computing $D_{i,j}$ of size roughly $2^{n^{1/2d}}$. Jumping ahead, we will rely on the algebraic construction when $n^{1/4} \le |i-j| \le n^{1/3}$, and on the combinatorial construction when $n^{1/3} \le |i-j| \le n^{1/2}$. The maximum circuit complexity peaks at $|i-j| = n^{1/3}$, where both constructions provide depth-d circuits of size roughly $2^{n^{2/3d}}$.

We now present the technical details.

$\mathsf{AC}^0[\oplus]$ circuits for small $|i-j|$. We need the following lemma.

▶ **Lemma 5** ([3, Proof of Lemma 3.1]). *For integers $n \ge 1$, $k \ge 0$, and $\ell \ge 1$ such that $n \ge k+\ell-1$, and for every $c_0, \ldots, c_{\ell-1} \in \mathbb{Z}$, there is a multivariate polynomial $Q \colon \{0,1\}^n \to \mathbb{Z}$ of degree at most $\ell - 1$ and with integer coefficients such that $p(x) = c_t$ for every $x \in \{0,1\}^n$ for which $|x|_1 = k + t$, where $0 \le t \le \ell - 1$. Moreover,*

$$Q(x) = \sum_{t=0}^{\ell-1} a_t \cdot Q_t(x),$$

where each $a_t \in \mathbb{Z}$, and $Q_t(x) = \sum_{S \in \binom{[n]}{t}} \prod_{j \in S} x_j$ denotes the t-th elementary symmetric polynomial.

Since the function $\psi \colon \mathbb{Z} \to \mathbb{F}_2$ that maps each integer to its parity is a ring homomorphism, it follows from Lemma 5 that there is a polynomial $P \in \mathbb{F}_2[y_1, \ldots, y_n]$ of degree $\ell \le |i-j|$ that agrees with $D_{i,j}$ on every input $y \in \{0,1\}^n$ such that $|y|_1 \in \{i, j\}$. Moreover,

$$P(y) = \sum_{t=0}^{\ell} b_t \cdot P_t(y),$$

where each $b_t \in \{0,1\}$, $P_t \in \mathbb{F}_2[y_1, \ldots, y_n]$ is the t-th elementary symmetric polynomial (over $\mathbb{F}_2$), and $t \le \ell$.

It is well known that each polynomial P_t can be computed by an *algebraic branching program* of width n and length $t+1$ (the j-th layer contains nodes $1, \ldots, n$ that store the largest coordinate in $[n]$ that has been read so far). Using a standard divide-and-conquer approach which is analogous to the construction of bounded-depth circuits for distance-k connectivity (see e.g. [10]), it follows that for every even depth $d' \geq 2$, P_t can be computed by a layered $\mathsf{AC}^0[\oplus]$ circuit of depth d' (consisting of $\oplus$ and $\wedge$ gates) and of size at most $n^{O(t^{2/d'})}$. Moreover, the top gate of this circuit is a parity gate. Using the definition of P as a sum of polynomials P_t over $\mathbb{F}_2$ (which allows us to collapse two layers of parities), and the fact that $t \leq \ell \leq |i-j|$, it follows that for every even integer $d' \geq 2$, each $D_{i,j}$ can be computed by a depth-d' $\mathsf{AC}^0[\oplus]$ circuit of size at most $n^{O(|i-j|^{2/d'})}$. Consequently, there are circuits for $D_{i,j}$ of size $n^{O(|i-j|^{2/(d'-1)})}$ and depth d' for each integer $d' \geq 3$.

AC^0 circuits for large $|i-j|$. We assume without loss of generality that $i > j$, since negating the output of the circuit will handle the other case. First, we note that the computation of $D_{i,j}$ can be reduced to the case where i and j are near the middle layer. For $0 \leq a < b \leq m$, let $\mathsf{Promise\text{-}Th}^m_{a,b}$ be an m-bit boolean function such that

$$\mathsf{Promise\text{-}Th}^m_{a,b}(y) = \begin{cases} 0 & \text{if } |y|_1 \leq a, \\ 1 & \text{if } |y|_1 \geq b. \end{cases}$$

Let $r \stackrel{\text{def}}{=} i - j$. Then $D_{i,j}$ can be obtained as a projection of $\mathsf{Promise\text{-}Th}^{10n}_{5n-\lceil r/10\rceil, 5n+\lceil r/10\rceil}$. More precisely, it is easy to check that, over the inputs of interest for $D_{i,j}$,

$$D_{i,j}(y) \;=\; \mathsf{Promise\text{-}Th}^{10n}_{5n-\lceil r/10\rceil, 5n+\lceil r/10\rceil}(1^{5n-\lceil r/10\rceil - j}\, y\, 0^{10n-(5n-\lceil r/10\rceil - j + n)}).$$

It follows from the work of [24] (see also [18]) that this promise threshold function can be computed by *randomized* AC^0 circuits of depth $d' \geq 2$ and of size $\exp(O(1/\delta)^{1/(d'-1)})$, where $\delta \stackrel{\text{def}}{=} \Theta(r/n)$. By a standard derandomization argument (see e.g. [2]) that increases the number of layers by at most 2, and by collapsing adjacent layers during this derandomization, it follows that for every $d' \geq 3$, $\mathsf{Promise\text{-}Th}^{10n}_{5n-\lceil r/10\rceil, 5n+\lceil r/10\rceil}$ can be computed by a (deterministic) depth-d' circuit of size $\exp(O(n/r)^{1/(d'-2)})$. Therefore, for every integer $d' \geq 3$, each $D_{i,j}$ can be computed by a depth-d' AC^0 circuit of size at most $\exp(O(n/|i-j|)^{1/(d'-2)})$.

Let $d \geq 5$ be given. In order to compute the symmetric function f_n, we proceed as described above. Two layers are employed to combine the sub-circuits $D_{i,j}$ in the appropriate way (via the functions E_i). In the remaining $d' \stackrel{\text{def}}{=} d-2$ layers, we pick the best construction for $D_{i,j}$ depending on the value $|i-j|$. If $|i-j| \leq n^{1/3}$, we employ the algebraic construction. It provides for each integer $d' \geq 3$ a depth-d' $\mathsf{AC}^0[\oplus]$ circuit of size at most $\exp(O(\log n \cdot |i-j|^{2/(d'-1)})) = \exp(\widetilde{O}(n^{\frac{2}{3}\cdot\frac{1}{(d'-1)}}))$. On the other hand, if $|i-j| > n^{1/3}$ the combinatorial construction gives for each integer $d' \geq 3$ a depth-d' AC^0 circuit of size at most $\exp(O(n/|i-j|)^{1/(d'-2)}) = \exp(O(n^{\frac{2}{3}\cdot\frac{1}{(d'-2)}}))$. Overall, we obtain a depth-d $\mathsf{AC}^0[\oplus]$ circuit for f_n of size at most $2^{\widetilde{O}(n^{\frac{2}{3}\cdot\frac{1}{(d-4)}})}$. ◀

2.2 An upper bound for linear threshold functions

Recall that an *exact threshold function* is a boolean function $f(x_1, \ldots, x_n)$ that evaluates to 1 if and only if $\sum_i w_i x_i = t$, where $w_1, \ldots, w_n, t \in \mathbb{R}$.

▶ **Lemma 6.** *Any threshold function on n variables can be computed by a polynomial-size constant-depth circuit with unbounded fan-in AND and OR gates, and a single layer of exact threshold gates each of fan-in n.*

Proof. The lemma is implicit in the work of Hansen and Podolskii [13]. Indeed, the proof of Theorem 7 in their paper shows that every threshold function on n variables can be written as a polynomial-sized OR of exact threshold functions, each of which is also on n variables. ◀

▶ **Lemma 7.** *Any exact threshold function on n variables can be computed by a polynomial-size constant-depth circuit with unbounded fan-in AND and OR gates, and a single layer of symmetric gates each of fan-in n.*

Proof. Our proof proceeds via Chinese remaindering, which is a common technique in the study of threshold functions.

Suppose the exact threshold function is $\sum_i w_i x_i = t$, where we can assume w.l.o.g. that each w_i as well as t are integers that are $n^{O(n)}$ (we refer to [13] for more information about exact threshold gates). Let $p_1, p_2 \ldots p_{n^2}$ be the first n^2 primes. Using the upper bounds on each w_i, we have that for any input x, $\sum_i w_i x_i$ is characterized by its sequence of remainders modulo $\{p_j\}, j = 1 \ldots n^2$; the same holds for t. We design a circuit that is an AND of n^2 circuits C_j, one for each p_j, where circuit C_j verifies that $(\sum_i w_i x_i) \mod p_j = t \mod p_j$.

We now describe how to construct any fixed C_j. Note that $t \mod p_j$ is a fixed quantity independent of the input, so our task reduces to computing $(\sum_i w_i x_i) \mod p_j$ and taking an AND of the output bits or their negations as appropriate.

Let w_{ij} be $w_i \mod p_j$ for each $i \in [1, n], j \in [1, n^2]$. We need to compute $\sum_i w_{ij} x_i$ using a polynomial-size constant-depth circuit with unbounded fan-in AND and OR gates, and a single layer of symmetric gates each of fan-in n.

It follows from the Prime Number Theorem that each w_{ij} has at most $3\log(n)$ bits in its binary representation, for large enough n. We write each w_{ij} as $\sum_k w_{ijk} 2^k$, where w_{ijk} is the kth bit in the binary representation of w_{ij}, for $k \leq 3\log(n)$.

For each j and k, consider the following circuit B_{jk}. It has n inputs, where the i'th input bit is the AND of w_{ijk} (which is a fixed bit independent of the input) and x_i. B_{jk} computes the sum of these n inputs – this can be done by using at most $\lceil \log(n) \rceil$ symmetric gates in parallel, each of fan-in n.

Let y_{jk} be the output of each circuit B_{jk}. C_j computes $(\sum_k y_{jk} 2^k) \mod p_j$ using a constant-depth circuit of polynomial size. This can be done because there are only $O(\log(n))^2$ input bits and the function we are computing is in NC^1 (see e.g. [19]); it is folklore that any NC^1 function on polylogarithmically many bits can be computed by polynomial-size AC^0 circuits.

Summing up, our circuit has poly(n) size and $O(1)$ depth, and has a single layer of symmetric gates with fan-in n, as promised. ◀

▶ **Theorem 8.** *There is an integer c such that for every integer $d > c$, if $f_n : \{0,1\}^n \to \{0,1\}$ is a threshold function, then it can be computed by an $\mathsf{AC}^0[\oplus]$ circuit of depth d and of size $2^{\widetilde{O}(n^{\frac{2}{3} \cdot \frac{1}{(d-c)}})}$.*

Proof. We combine Lemmas 6 and 7 with Theorem 4. From Lemma 6 and Lemma 7, it follows that every threshold function on n variables can be computed by a polynomial-size constant-depth circuit with unbounded fan-in AND and OR gates, and a single layer of symmetric gates each of fan-in n. Suppose that the depth of this circuit is k. We set $c = k + 4$.

By using Theorem 4, we can replace each symmetric gate by a $\mathsf{AC}^0[\oplus]$ circuit of depth $d-k$ and of size $2^{\widetilde{O}(n^{\frac{2}{3}\cdot\frac{1}{(d-c)}})}$. The total size of the resulting circuit is $\mathrm{poly}(n)2^{\widetilde{O}(n^{\frac{2}{3}\cdot\frac{1}{(d-c)}})}$ and its depth is d. We can absorb the polynomial factor in the circuit size into the exponential term, to yield the result stated in the theorem. ◀

We have made no attempt to optimize the integer c in the statement of Theorem 8.

Note that the same proof as for Theorem 8 yields that any Boolean function truth-table reducible to linear threshold functions using a polynomial-size AC^0 reduction, where the size of any query is at most n, is also computable by $\mathsf{AC}^0[\oplus]$ circuits of the same size and depth as in the statement of the theorem. In particular, this is the case for polytopes, since any polytope over n variables is simply an AND of linear threshold functions over n variables.

2.3 A depth-4 upper bound

For any $n \geq 1$ and $i \in \{0, \ldots, n\}$, let $E_{n,i}$ denote the n-variable Boolean function that accepts inputs of Hamming weight i and rejects all other inputs.

▶ **Theorem 9.** *Any symmetric function on n variables has a depth-4 $\mathsf{AC}^0[\oplus]$ circuit of size $\exp(O(n^{1/4} \cdot (\log n)^{3/4}))$.*

This improves on the AC^0 upper bound of $\exp(\widetilde{O}(n^{1/3}))$ [8], which is tight up to log factors in the exponent [14]. To prove the above theorem, it suffices to show the following upper bound for exact majorities.

▶ **Lemma 10.** *Assume n is even. Then $E_{n,n/2}$ has a depth-4 $\mathsf{AC}^0[\oplus]$ circuit C of size $\exp(O(n^{1/4} \cdot (\log n)^{3/4}))$ with the output gate being an OR gate.*

We first prove Theorem 9 assuming Lemma 10.

Proof of Theorem 9. By Lemma 10, we have a depth-4 $\mathsf{AC}^0[\oplus]$ circuit C of size $\exp(O(n^{1/4} \cdot (\log n)^{3/4}))$ with an OR output gate that computes $E_{2n,n}$. Note that this yields a circuit C_i for $E_{n,i}$ via the substitution $C_i(x_1, \ldots, x_n) = C(x0^i1^{n-i})$; observe that C_i is also a depth-4 $\mathsf{AC}^0[\oplus]$ circuit of size $\exp(O(n^{1/4}(\log n)^{3/4}))$ with an OR output gate.

Since any symmetric function on n variables is an OR of a subset of the $E_{n,i}$, this yields the theorem. ◀

We now discuss the proof of Lemma 10. Let r, s be growing functions of n with $s = o(n/\log n)$. We will design a *random* depth-3 circuit $\boldsymbol{C}'_{n,r,s}$ such that

1. For any input a of Hamming weight $k \neq n/2$, $\Pr_{\boldsymbol{C}'_{n,r,s}}[\boldsymbol{C}'_{n,r,s}(a) = 1] = 0$.
2. For any input a of Hamming weight $n/2$, $\Pr_{\boldsymbol{C}'_{n,r,s}}[\boldsymbol{C}'_{n,r,s}(a) = 1] \geq p \overset{\mathrm{def}}{=} n^{-r}$.

(The parameter s will be used to optimize the size of the final circuit.)

The construction of C will easily follow from that of $\boldsymbol{C}'_{n,r,s}$. The latter, which we now describe, uses a modification of Amano's construction [4] of random formulas for approximating the Majority function (which itself builds upon [2, 28, 21]), some basic facts about polynomial interpolation, and well-known ideas for computing Exact majorities [8, 22].

The lemma below is Amano's construction with a few parameters modified.

▶ **Lemma 11.** *Let m be a growing parameter and $\delta = o(1/\log m)$. There exists a random $\bigwedge\bigvee\bigwedge$ formula $\boldsymbol{F}_3$ of size $\exp(O(\sqrt{(\log m)/\delta}))$ such that*

1. *For any input a of Hamming weight $i \leq m((1/2) - \delta)$, $\Pr_{\boldsymbol{F}_3}[\boldsymbol{F}_3(a) = 1] = 0$.*
2. *For any input a of Hamming weight $i \geq m/2$, $\Pr_{\boldsymbol{F}_3}[\boldsymbol{F}_3(a) = 1] \geq (3/4)$.*

A proof sketch is given later in this section. The construction of the random circuit $\boldsymbol{C}'_{n,r,s}$ now proceeds as follows.

1. Divide the n input variables $x_1, \ldots, x_n$ randomly into r buckets $B_1, \ldots, B_r$ of size n/r each. We assume $r|n$ and that $m \overset{\text{def}}{=} n/r$ is even for simplicity.
2. Let $\delta = s/m$. For each bucket B_i, use Lemma 11 to construct a random $\bigwedge\bigvee\bigwedge$ formula $\boldsymbol{F}^{(i)}$ of size $\exp(O(\sqrt{(1/\delta)\log m}))$ on the variables in B_i that accepts no input of Hamming weight at most $m((1/2) - \delta)$ and accepts each input of Hamming weight at least $m/2$ with probability at least $3/4$.

   Define $\boldsymbol{G}^{(i)}$ to be $\boldsymbol{F}^{(i)}(\neg x : x \in B_i)$. Note that $\boldsymbol{G}^{(i)}$ accepts no input of Hamming weight at least $m((1/2) + \delta)$ and accepts each input of Hamming weight at most $m/2$ with probability at least $3/4$.

   Let $\boldsymbol{H}^{(i)} = \boldsymbol{F}^{(i)} \wedge \boldsymbol{G}^{(i)}$. By a union bound, $\boldsymbol{H}^{(i)}$ accepts each input of Hamming weight *exactly* $m/2$ with probability at least $1/2$ and no input of Hamming weight k such that $|k - (m/2)| \geq \delta m$.
3. For each bucket B_i, let $P^{(i)} \in \mathbb{F}_2[x : x \in B_i]$ be a multilinear polynomial of degree at most $2s$ that accepts inputs of Hamming weight $m/2$ but no input of Hamming weight k such that $|k - (m/2)| < s$. Such a polynomial exists by standard interpolation arguments (cf. Lemma 5; for a proof see e.g. Alman and Williams [3, Proof of Lemma 3.1]).

   We think of $P^{(i)}$ as a depth-2 $\bigoplus\bigwedge$ formula of size $n^{O(s)}$.
4. Finally, we define $\boldsymbol{C}'_{n,r,s} = \bigwedge_{i\in[r]} \left(\boldsymbol{H}^{(i)} \wedge P^{(i)}\right)$.

   By construction, $\boldsymbol{C}'_{n,r,s}$ is a depth-3 $\mathsf{AC}^0[\oplus]$ circuit of size $\mathrm{poly}(n) \cdot \exp(O(s \log n + \sqrt{(m/s)\log n}))$.

   Given any input a of Hamming weight $k \neq n/2$, there is a bucket B_i such that the restriction $a^{(i)}$ to B_i has weight $k_i \neq m/2$. In this case, either $\boldsymbol{H}^{(i)}$ or $P^{(i)}$ rejects $a^{(i)}$ (depending on whether $|k_i - m| \geq s$ or $|k_i - m| < s$ respectively). Hence, $\boldsymbol{C}'_{n,r,s}$ rejects a (with probability 1).

   Conversely, given an input a of Hamming weight $n/2$, $\boldsymbol{C}'_{n,r,s}$ accepts a if its restriction $a^{(i)}$ of a to each bucket B_i has weight exactly $m/2$ and we have a good choice for the randomness of each $\boldsymbol{H}^{(i)}$. The probability of this is at least

$$\left(\frac{3}{4}\right)^r \cdot \frac{\binom{m}{m/2}^r}{\binom{n}{n/2}} \geq \frac{(2^m/10\sqrt{m})^r}{2^n} \geq \frac{1}{n^r} = p.$$

So we have constructed $\boldsymbol{C}'_{n,r,s}$ as required. In order to convert this to a circuit for $E_{n,n/2}$, we use a standard covering argument. Let $t = n/p$. We choose independent random circuits $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_t$ where each $\boldsymbol{C}_i$ has the same distribution as $\boldsymbol{C}'_{n,r,s}$. Define $\boldsymbol{C}_{n,r,s} = \bigvee_i \boldsymbol{C}_i$.

Clearly, $\boldsymbol{C}_{n,r,s}$ accepts no input a of Hamming weight $k \neq n/2$. On the other hand, the probability that $\boldsymbol{C}_{n,r,s}$ rejects an input a of weight $n/2$ can be upper bounded by $(1-p)^t \leq \exp(-pt) = \exp(-n)$. By a union bound, the probability that $\boldsymbol{C}_{n,r,s}$ rejects *some* input of weight $n/2$ is at most $\binom{n}{n/2} \cdot \exp(-n) < 1$.

In particular, by averaging, there is a fixed circuit $C_{n,r,s}$ in the support of the distribution of $\boldsymbol{C}_{n,r,s}$ that computes $E_{n,n/2}$ correctly on all inputs.

By construction, the circuit $C_{n,r,s}$ has size $\mathrm{poly}(n) \cdot \exp(O((r+s)\log n + \sqrt{(m/s)\log n}))$. Setting $r = s = \Theta((n/\log n)^{1/4})$, we get a circuit C of the claimed size. This completes the proof of Lemma 10.

Proof Sketch of Lemma 11. We provide a sketch of the proof, omitting calculations. The reader is invited to consult Amano's paper [4] for more details.

Set $\ell = \lceil \sqrt{(\log m)/\delta} \rceil$ and define the random formulas $\boldsymbol{F}_i$ ($i \in [3]$) of depth i as follows.

1. $\boldsymbol{F}_1$ is simply an AND of size ℓ, where each input is chosen i.u.a.r. from among the input variables $\{x_1, \ldots, x_m\}$.
2. $\boldsymbol{F}_2$ is an OR of $L \stackrel{\text{def}}{=} \lceil 2^\ell \cdot (100\ell \ln 2) \rceil$ independent copies of $\boldsymbol{F}_1$.
3. $\boldsymbol{F}'_3$ is an AND of $M \stackrel{\text{def}}{=} 2^{100\ell - 3}$ independent copies of $\boldsymbol{F}_2$.
4. We define $\boldsymbol{F}_3$ to be $\boldsymbol{F}'_3$ conditioned on the event that $\boldsymbol{F}'_3$ does not accept any input of Hamming weight $i \leq m((1/2) - \delta)$.

Clearly, $\boldsymbol{F}_3$ is a random formula of the required size.

To argue that $\boldsymbol{F}_3$ has the required input-output behaviour, we proceed as follows.

1. Say a is an input of Hamming weight at least $m/2$.
   a. A uniformly random co-ordinate of a is 1 with probability at least $1/2$. Hence, $\boldsymbol{F}_1(a) = 1$ with probability at least $p_1 \stackrel{\text{def}}{=} 2^{-\ell}$.
   b. Hence, the probability that $\boldsymbol{F}_2$ rejects a is at most $(1 - p_1)^L \leq p_2 \stackrel{\text{def}}{=} 2^{-100\ell}$.
   c. Therefore, the probability that $\boldsymbol{F}'_3$ rejects a is at most $Mp_2 \leq 1/8$.
2. Now assume a' is an input of Hamming weight at most $m((1/2) - \delta)$.
   a. A uniformly random co-ordinate of a' is 1 with probability at most $1/2 - \delta$. Hence, $\boldsymbol{F}_1(a') = 1$ with probability at most $q_1 \stackrel{\text{def}}{=} p_1 \cdot (1 - \delta\ell)$.
   b. Hence, the probability that $\boldsymbol{F}_2$ rejects a' is at least $(1-q_1)^L \geq q_2 \stackrel{\text{def}}{=} p_2 \cdot \exp(10\delta\ell^2 \ln 2) \geq p_2 \cdot m^{10}$.
   c. Therefore, the probability that $\boldsymbol{F}'_3$ accepts a' is at most $(1-q_2)^M \leq \exp(-M \cdot p_2 \cdot m^{10}) \leq \exp(-m^9)$.
3. Thus, the probability that $\boldsymbol{F}_3$ accepts an input a of Hamming weight at least $m/2$ is at least $(7/8) - 2^m \cdot \exp(-m^9) \geq (3/4)$. This concludes the proof.

3 The Lower Bounds

3.1 A refined analysis of approximate-degree bounds

The main theorem of this section is the following result.

▶ **Theorem 12.** *Fix any $d \geq 2$. Let C be a depth-d $\mathsf{AC}^0[\oplus]$ circuit computing the n-bit majority function Maj_n. Then, C has size at least $\exp(\Omega(n^{1/(2d-3)}))$.*

We follow the lower bound approach of Razborov [23], who showed that any small $\mathsf{AC}^0[\oplus]$ circuit C can be suitably approximated by a low-degree polynomial. This is proved by iteratively constructing low-degree polynomials for the OR and AND gates of C (parity gates are low-degree by definition, and hence trivial to approximate), and then composing the polynomials together to obtain a low-degree approximation to C. We follow a similar idea, but make the (crucial) observation that the approximations for the AND and OR gates are one-sided (on opposite sides). This means that the construction of Razborov is slightly better than normally advertised: the error is much lower on $C^{-1}(b)$ than $C^{-1}(1-b)$ for some $b \in \{0,1\}$.

▶ **Definition 13.** *Let $f : \{0,1\}^n \to \{0,1\}$ be any Boolean function. For parameters $\varepsilon_0, \varepsilon_1 \in (0,1)$, an $(\varepsilon_0, \varepsilon_1)$-error* Probabilistic polynomial *for f is a random multilinear polynomial $\boldsymbol{P}$ chosen from $\mathbb{F}_2[x_1, \ldots, x_n]$ such that for $b \in \{0,1\}$ and any $a \in f^{-1}(b)$,*

$$\Pr_{\boldsymbol{P}}[\boldsymbol{P}(a) \neq f(a)] \leq \varepsilon_b.$$

*We say that $\boldsymbol{P}$ has degree at most d (denoted $\deg(\boldsymbol{P}) \leq d$) if the underlying distribution is supported on multilinear polynomials of degree at most d. We define the $(\varepsilon_0, \varepsilon_1)$-*error probabilistic degree *of f (denoted $\mathrm{pdeg}_{\varepsilon_0,\varepsilon_1}(f)$) to be the least d such that there is a $\boldsymbol{P}$ as above of degree at most d.*

Typically, the above is stated for $\varepsilon_0 = \varepsilon_1$, but it will be important for us to track the errors on the 0 and 1 inputs of f separately. For example, it allows us to observe the following feature of a construction due to Razborov [23]. (See also Kopparty's lecture notes [16] for a proof.)

▶ **Lemma 14** (Razborov [23]). *Let AND_m and OR_m denote the AND and OR functions on m inputs respectively. Then, for any $\varepsilon > 0$, $\mathrm{pdeg}_{\varepsilon,0}(\mathrm{AND}_m)$ and $\mathrm{pdeg}_{0,\varepsilon}(\mathrm{OR}_m)$ are both at most $\lceil \log(1/\varepsilon) \rceil$.*

From this, we get the following corollary.

▶ **Corollary 15.** *Let C be an $\mathsf{AC}^0[\oplus]$ circuit of size s and depth $d \geq 1$. Then, for any $c \geq 1$ and large enough s, we have*

$$\min\{\mathrm{pdeg}_{(1/10),(1/s^c)}(C), \mathrm{pdeg}_{(1/s^c),(1/10)}(C)\} \leq O(c \log s)^{d-1}$$

where the $O(\cdot)$ hides an absolute constant.

Proof Sketch. We assume that the output gate of C is either a parity gate or an OR gate (the case of the AND gate is similar to the case of the OR gate).

For each non-output gate g in the circuit (viewed as a function of its input wires), we first construct a $(1/s^{c+2}, 1/s^{c+2})$-error probabilistic polynomial $\boldsymbol{P}_g$ of degree $O(c \log s)$ for g. Note that the existence of $\boldsymbol{P}_g$ is trivial if g is a parity gate (since the parity function is a polynomial of degree 1) and otherwise, Lemma 14 gives us such a probabilistic polynomial.

For the output gate g_0, we construct a $(0, 1/20)$-error probabilistic polynomial $\boldsymbol{P}_{g_0}$ of degree $O(1)$: again, this is trivial if g_0 is a parity gate and follows from Lemma 14 if g_0 is an OR gate.

Composing these polynomials together, we get a probabilistic polynomial $\boldsymbol{P}$ of degree $O(c \log s)^{d-1} \cdot O(1) = O(c \log s)^{d-1}$. Further for any input $a \in \{0,1\}^n$, we have $\boldsymbol{P}(a) = C(a)$ unless there is some gate g of C such that $\boldsymbol{P}_g$ does not simulate g faithfully on the corresponding setting of its inputs. For non-output gates, this probability is at most $1/s^{c+2}$. For the output gate, this probability is either 0 or at most $1/20$ depending on whether $a \in C^{-1}(0)$ or $C^{-1}(1)$ respectively. A union bound now implies that $\mathrm{pdeg}_{1/s^c,1/10}(C) = O(c \log s)^{d-1}$. ◀

The rest of the proof follows the lower bound of Smolensky [25] on the probabilistic degree of the Majority function. More precisely, we prove the following.

▶ **Lemma 16.** *Let n be a growing parameter. There exist absolute constants $\alpha, \beta > 0$ such that for all large enough n and all $\varepsilon \in (1/2^{\alpha n}, \beta)$, we have*

$$\min\{\mathrm{pdeg}_{1/10,\varepsilon}(\mathrm{Maj}_n), \mathrm{pdeg}_{\varepsilon,1/10}(\mathrm{Maj}_n)\} = \Omega(\sqrt{n \log(1/\varepsilon)}).$$

Smolensky[2] proved the above for $\mathrm{pdeg}_{\varepsilon,\varepsilon}(\mathrm{Maj}_n)$. Note that the above statement in conjunction with Corollary 15 immediately implies Theorem 12. Putting the upper bound in Corollary 15 for $c = 1$ together with the lower bound in Lemma 16, we get

$$O(\log s)^{d-1} \geq \Omega(\sqrt{n \log s}),$$

which yields $s = \exp(\Omega(n^{1/(2d-3)}))$.

To prove Lemma 16, we follow a "dual" version of Smolensky's proof that appears in a result of Kopparty and Srinivasan [17], which itself follows the closely-related ideas of Aspnes, Beigel, Furst and Rudich [5] and Green [12]. It is not clear that this dual reformulation is necessary for the proof below, but the language of this formulation makes it easier to use some other results from the literature in this context.

We start with the notion of a certifying polynomial for a Boolean function.

▶ **Definition 17** (Certifying polynomials). *A non-zero multilinear polynomial* $R \in \mathbb{F}_2[x_1, \ldots, x_n]$ *is a* certifying polynomial *for* $f : \{0,1\}^n \to \{0,1\}$ *if* f *is constant on* $\mathrm{Supp}(R) \overset{\mathrm{def}}{=} \{a \in \mathbb{F}_2^n \mid R(a) \neq 0\}$.

The following is an easy corollary of standard properties of multilinear polynomials (see e.g. [17, Lemma 3.3]).

▶ **Fact 18.** *If* R *is a certifying polynomial for* Maj_n*, then* $\deg(R) \geq \lceil n/2 \rceil$.

We now return to Lemma 16. Let $\boldsymbol{P}$ be an $(\varepsilon, 1/10)$-error[3] probabilistic polynomial for Maj_n, and let us assume $\deg(\boldsymbol{P}) \leq d$. We need to lower bound d. By a union bound and averaging, we can find a (deterministic) polynomial $P \in \mathbb{F}_2[x_1, \ldots, x_n]$ of degree at most d such that

$$\Pr_{x \in \mathrm{Maj}_n^{-1}(0)}[P(x) \neq 0] \leq 2\varepsilon \quad \text{and} \quad \Pr_{x \in \mathrm{Maj}_n^{-1}(1)}[P(x) \neq 1] \leq \frac{1}{4}. \tag{1}$$

It suffices to lower bound $\deg(P)$. To do so, we show that there is a non-zero polynomial $Q \in \mathbb{F}_2[x_1, \ldots, x_n]$ of low degree such that Q vanishes on all points in $E_0 \overset{\mathrm{def}}{=} \{x \in \mathrm{Maj}_n^{-1}(0) \mid P(x) \neq 0\}$. We then consider the multilinear polynomial $R = P \cdot Q$ (we use the identity $x_i^2 = x_i$ to ensure that R is multilinear). Note that R vanishes on all points of Hamming weight less than $n/2$: given a of weight less than $n/2$, either $a \in E_0$, in which case $Q(a) = 0$, or $a \notin E_0$, which implies that $P(a) = 0$. If we could argue that R is a non-zero polynomial, then it follows that R is a certifying polynomial for Maj_n and hence has degree at least $\lceil n/2 \rceil$ (by Fact 18). On the other hand, $\deg(R) \leq \deg(P) + \deg(Q)$ which implies a lower bound on $\deg(P)$.

The main part of the above argument is arguing the non-zeroness of R. To do this, we would like to show that there is a low-degree polynomial Q such that Q vanishes on E_0, but there is an $a \in \mathrm{Supp}(P)$ such that $Q(a) \neq 0$. To argue the existence of a suitable such Q, we use a result of Nie and Wang [20]. Informally, the result says that if a parameter D is chosen so that the number of multilinear monomials of degree at most D is much larger than $|E_0|$, then constraining a polynomial of degree at most D to be zero on E_0 does not constrain it at too many other points.

[2] Smolensky actually proves lower bounds for mod functions, which we don't consider here. However, it is clear that his proof works also for Majority. As far as we know, this proof first appeared in Szegedy's PhD thesis [27]. See [9] for a more recent exposition. A different proof, also due to Smolensky, appears in [26].

[3] A symmetric argument can be used to argue about $\mathrm{pdeg}_{1/10,\varepsilon}(\mathrm{Maj}_n)$.

To make this precise, we define the *degree-D Closure of E_0*, denoted $\mathrm{cl}_D(E_0)$, to be the set of all $a \in \mathbb{F}_2^n$ such that $Q(a) = 0$ for each Q of degree at most D such that Q vanishes at all points in E_0. Clearly, $\mathrm{cl}_D(E_0) \supseteq E_0$ but could potentially be much larger. The result of Nie and Wang [20] bounds the closure of small sets in $\mathbb{F}_2^n$ (see also the earlier results of Wei [29] and Keevash and Sudakov [15] which prove similar or stronger statements in different language).

▶ **Theorem 19** (Nie and Wang [20]). *Fix any $E \subseteq \mathbb{F}_2^n$ and any $D \geq 1$. Let N_D denote the number of multilinear monomials of degree at most D. Then, we have*

$$\frac{|\mathrm{cl}_D(E)|}{2^n} \leq \frac{|E|}{N_D}.$$

▶ Remark 20. Note that the above theorem generalizes the following standard fact (which follows easily from linear algebra): if $|E| < N_D$, then there is a non-zero polynomial of degree at most D that vanishes on E. Smolensky's proof (as formulated in [17]) can be seen as using only this special case of Theorem 19. Using the theorem in its full generality is what yields the stronger result below.

We are now ready to prove Lemma 16.

Proof of Lemma 16. It suffices to lower bound $\deg(P)$ where P is as in (1). Let E_0 be as defined above; by (1), we have $|E_0| \leq 2\varepsilon \cdot 2^n$. Also define $S = \mathrm{Supp}(P)$. Note that S contains all those $a \in \mathrm{Maj}_n^{-1}(1)$ such that $P(a) = 1$ which, by (1), has size at least $2^n \cdot (3/8 - o(1))$.

Now, choose the least D such that $N_D \geq (20\varepsilon) \cdot 2^n$. As long as α and β are small enough constants, we have $D = (n/2) - \Omega(\sqrt{n \log(1/\varepsilon)})$. Also, by Theorem 19, we know that $|\mathrm{cl}_D(E_0)| \leq 2^n/10 < |S|$. In particular, $S \not\subseteq \mathrm{cl}_D(E_0)$. This means that there is some $a_0 \in S$ and some Q of degree at most D such that Q vanishes on E_0 but not at a_0.

Let $R = P \cdot Q$ (we assume R is multilinear by using the identity $x_i^2 = x_i$). As argued above, R vanishes at all points in $\mathrm{Maj}_n^{-1}(0)$ and further, $R(a_0) = P(a_0)Q(a_0) \neq 0$. Hence, R is a non-zero polynomial such that Maj_n is the constant function 1 on inputs from $\mathrm{Supp}(R)$. By Fact 18, we have $\deg(R) \geq n/2$.

This implies that $\deg(P) \geq n/2 - D = \Omega(\sqrt{n \log(1/\varepsilon)})$. ◀

3.2 A depth-3 lower bound

In this section, we show how to use the ideas from the proof of Theorem 12 in conjunction with standard AC^0 lower bound techniques to get a near optimal lower bound of $\exp(\Omega(\sqrt{n}))$ for depth-3 circuits (there is an AC^0 circuit of size $\exp(\widetilde{O}(\sqrt{n}))$ computing the Majority function [8]).

▶ **Theorem 21.** *Let C be any depth-3 $\mathsf{AC}^0[\oplus]$ circuit computing the n-bit Majority function Maj_n. Then, C has size $\exp(\Omega(\sqrt{n}))$.*

The proof requires random restriction arguments [11, 1]. Recall that a *restriction* on n variables $x_1, \ldots, x_n$ is a function $\rho\colon \{x_1, \ldots, x_n\} \to \{0, 1, *\}$. A *Random restriction with $*$-probability $p \in [0, 1]$* is a random function $\boldsymbol{\rho}\colon \{x_1, \ldots, x_n\} \to \{0, 1, *\}$ where $\boldsymbol{\rho}^{-1}(*)$ is chosen to be a random subset $S \subseteq [n]$ of size $\lfloor pn \rfloor$ and each $\boldsymbol{\rho}(x_i)$ ($x_i \notin S$) is set to 0 or 1 independently with probability $(1-p)/2$ each. We use $\boldsymbol{\rho} \sim \mathcal{R}_p^n$ to denote the fact $\boldsymbol{\rho}$ is a random restriction on n variables with $*$-probability p.

Given a Boolean function $f : \{0,1\}^n \to \{0,1\}$, and a restriction $\rho : \{x_1, \ldots, x_n\} \to \{0, 1, *\}$, we use $f|_\rho$ to denote the restriction of f obtained by substituting variables as dictated by ρ (variables in $\rho^{-1}(*)$ are left as is).

We recall the Switching Lemma of Håstad [14] (this version is from Beame's survey [7]).

▶ **Lemma 22** (Håstad's Switching Lemma). *Let φ be a k-CNF or k-DNF. Then for $p \leq 1/10k$, we have*

$$\Pr_{\rho\sim\mathcal{R}_p}[\varphi|_\rho \textit{ has no decision tree of depth at most } t] \leq (7pk)^t.$$

We say that a restriction ρ is *balanced* if $|\rho^{-1}(1)| = |\rho^{-1}(0)|$. Balanced restrictions will be useful for us since for a balanced restriction ρ, we have $\mathrm{Maj}_n|_\rho = \mathrm{Maj}_{|\rho^{-1}(*)|}$. Lemma 22 easily implies a similar corollary for random balanced restrictions.

▶ **Corollary 23.** *Let φ be a k-CNF or k-DNF on n variables. Then for $p \leq 1/10k$ such that $(n - \lfloor pn \rfloor)$ is even, we have*

$$\Pr_{\rho\sim\mathcal{R}_p^n}[\varphi|_\rho \textit{ has no decision tree of depth at most } t \mid \rho \textit{ is balanced}] \leq O(\sqrt{n}) \cdot (7pk)^t.$$

Proof. Follows directly from Lemma 22 and Bayes' rule since the probability that a random $\rho \sim \mathcal{R}_p^n$ is balanced is at least $\Omega(1/\sqrt{n})$. ◀

Using Corollary 23, we can derive the following simplification lemma for general depth-2 $\mathsf{AC}^0[\oplus]$ circuits.

▶ **Lemma 24.** *Let n, s be growing parameters with $s \geq n^2$. Let C' be any $\mathsf{AC}^0[\oplus]$ circuit on n variables of depth 2 and size at most s. Assume $p \leq 1/(500 \log s)$ is chosen so that $(n - \lfloor pn \rfloor)$ is even. Then, for large enough n, s, we have*

$$\Pr_{\boldsymbol{\rho}\sim\mathcal{R}_p^n}[\mathrm{pdeg}_{1/s^2,1/s^2}(C'|_{\boldsymbol{\rho}}) > 10\log s \mid \boldsymbol{\rho} \textit{ is balanced}] < \frac{1}{10s}.$$

Proof. The proof is a routine application of the switching lemma. We provide details for completeness.

To avoid some technicalities, we assume that n and $n/10$ are even integers. The proof can easily be extended to the other cases.

We use $\boldsymbol{\rho} \sim \mathcal{R}_{p,bal}^n$ to denote that ρ is a random restriction on n variables with $*$-probability p conditioned on being balanced. We sample $\boldsymbol{\rho}$ in two steps: we sample random restrictions $\boldsymbol{\rho}_1 \sim \mathcal{R}_{1/10,bal}^n$ and $\boldsymbol{\rho}_2 \sim \mathcal{R}_{10p,bal}^{n/10}$ and set $\boldsymbol{\rho}$ to be their composition $\boldsymbol{\rho}_2 \circ \boldsymbol{\rho}_1$ (i.e. we apply $\boldsymbol{\rho}_2$ to the variables in $\boldsymbol{\rho}_1^{-1}(*)$).

We first analyze the effect of applying $\boldsymbol{\rho}_1$. Consider any OR or AND gate g at depth 1 in C'. Say that g is *bad* for $\boldsymbol{\rho}_1$ if $g|_{\boldsymbol{\rho}_1}$ has fan-in at least $5 \log s$. Applying Corollary 23 with $k = 1$, we get for any gate g at depth 1,

$$\Pr_{\boldsymbol{\rho}_1}[g \text{ bad for } \boldsymbol{\rho}_1] \leq O(\sqrt{n}) \cdot (7/10)^{5\log s} < 1/(20s^2)$$

for large enough s. Union bounding over all gates g at depth 1 (there are at most s of them), we see that with probability at least $1 - 1/(20s)$, all gates at depth 1 are good for $\boldsymbol{\rho}_1$. Condition on such a setting ρ_1 of the random restriction $\boldsymbol{\rho}_1$. By definition of ρ_1, the circuit $C_1' \overset{\text{def}}{=} C'|_{\rho_1}$ has the property that all the AND and OR gates of depth 1 in C_1' have fan-in at most $5 \log s$: in particular, they are polynomials of degree at most $5 \log s$.

Now we analyze the effect of $\boldsymbol{\rho}_2$ on C_1'. This is by a case analysis on the output gate g_0 of C_1'. Since the statement of the lemma is true for C' if and only if it is true for $\neg C'$, we can assume w.l.o.g. that the output gate of C' is either a parity gate or an OR gate.

1. **g_0 is a parity gate**: In this case, since the OR and AND gates at depth 1 compute polynomials of degree at most $5 \log s$, the entire circuit C'_1 already computes a polynomial of degree at most $5 \log s$. In particular, $C'_1|_{\boldsymbol{\rho}_2}$ has degree at most $5 \log s$ with probability 1.
2. **g_0 is an OR gate**: Then, we can write $C'_1 = C'_{1,1} \vee C'_{1,2}$, where $C'_{1,1}$ is an OR of parity gates and $C'_{1,2}$ is a $(5 \log s)$-DNF.

   $C'_{1,2}|_{\boldsymbol{\rho}_2}$ continues to be an OR of parities. By Lemma 14, any OR (and hence any OR of parities) has $(0, 1/s^2)$-probabilistic degree at most $\lceil 2 \log s \rceil \leq 3 \log s$. In particular, we have $\mathrm{pdeg}_{0,1/s^2}(C'_{1,2}|_{\boldsymbol{\rho}_2}) \leq 5 \log s$ with probability 1.

   For $C'_{1,1}$, we apply Corollary 23 (the Switching lemma) with $k = 5 \log s$. The random restriction $\boldsymbol{\rho}_2$ has $*$-probability $10p \leq 1/50 \log s = 1/10k$. Corollary 23 implies that the probability that $C'_{1,1}$ does not have a decision tree of height $5 \log s$ is at most $O(\sqrt{n}) \cdot (1/10)^{5 \log s} < 1/(20s)$. As a decision tree of height t can be represented as a polynomial of degree at most t, we see that with probability $1 - 1/(20s)$, the restricted $C'_{1,1}$ has degree at most $5 \log s$.

   Consequently, we see that with probability $1 - 1/(20s)$, we have both $\mathrm{pdeg}_{0,1/s^2}(C'_{1,2}|_{\boldsymbol{\rho}_2}) \leq 5 \log s$ and $\deg(C'_{1,1}) \leq 5 \log s$. When this happens, we also have $\mathrm{pdeg}_{0,1/s^2}(C'_1|_{\boldsymbol{\rho}_2}) \leq 10 \log s$ (the probabilistic polynomial for C'_1 can be obtained by composing the polynomial for the 2-bit OR function with polynomials for $C'_{1,1}$ and $C'_{1,2}$).

In both cases above, we have shown that

$$\Pr_{\boldsymbol{\rho}_2}[\mathrm{pdeg}_{(1/s^2,1/s^2)}(C'_1|_{\boldsymbol{\rho}_2}) > 10 \log s] < \frac{1}{20s}.$$

Along with our analysis of $\boldsymbol{\rho}_1$, this implies

$$\Pr_{\boldsymbol{\rho}}[\mathrm{pdeg}_{(1/s^2,1/s^2)}(C'|_{\boldsymbol{\rho}}) > 10 \log s] < \frac{1}{20s} + \frac{1}{20s} = \frac{1}{10s}. \qquad ◀$$

We are now ready to prove Theorem 21.

Proof of Theorem 21. Assume that C has size $s \leq \exp(\sqrt{n}/100)$, since otherwise we are done. Let $C'_1, \ldots, C'_s$ be the depth-2 subcircuits of C. Fix $p = \Theta(1/\log s)$ so that Lemma 24 is applicable.

Using Lemma 24 and applying a union bound over $i \in [s]$, we get

$$\Pr_{\boldsymbol{\rho} \sim \mathcal{R}_p^n}[\exists i \in [s],\ \mathrm{pdeg}_{1/s^2,1/s^2}(C'_i|_{\boldsymbol{\rho}}) > 10 \log s \mid \boldsymbol{\rho} \text{ is balanced}] < \frac{1}{10}.$$

In particular, there is a balanced restriction ρ on $\{x_1, \ldots, x_n\}$ such that $|\rho^{-1}(*)| = m = \Theta(n/\log s)$, and further, $\mathrm{pdeg}(C'_i|_\rho) \leq 10 \log s$ for each $i \in [s]$. Fix such a restriction ρ. W.l.o.g. we assume $\rho^{-1}(*) = \{x_1, \ldots, x_m\}$.

Fix $(1/s^2, 1/s^2)$-error probabilistic polynomials $\boldsymbol{P}_i(x_1, \ldots, x_m)$ of degree at most $10 \log s$ for C'_i ($i \in [s]$). We assume that the output gate g of C is either an OR gate or a parity gate (the case when the output is an AND gate is similar). In either case, Lemma 14 implies that g has a $(0, 1/10)$-error probabilistic polynomial $\boldsymbol{P}$ of constant degree.

Define $\boldsymbol{Q}(x_1, \ldots, x_m) = \boldsymbol{P}(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_s)$. Clearly, $\deg(\boldsymbol{Q}) \leq O(\max_i \deg(\boldsymbol{P}_i)) = O(\log s)$. Also, it is easy to see that $\boldsymbol{Q}$ is a $(1/s, 1/5)$-error probabilistic polynomial for $C|_\rho = \mathrm{Maj}_n|_\rho = \mathrm{Maj}_m$. Lemma 16 therefore implies that $\deg(\boldsymbol{Q}) \geq \Omega(\sqrt{m \cdot \log s}) = \Omega(\sqrt{n})$, which implies that $s = \exp(\Omega(\sqrt{n}))$. ◀

3.3 An improved lower bound for all depths

In this section, we complete the proof of Theorem 2. The proof extends the ideas employed in the preceding sections in a natural way. The difference here is that the argument below employs the construction from Corollary 15 as an intermediate step, while the proof of Theorem 21 is slightly simpler and only requires Lemma 14.

Proof of Theorem 2. Let C be a depth-d $\mathsf{AC}^0[\oplus]$ circuit of size s that computes Majority over n input bits, where $d \geq 3$. Proceeding as in the proof of Theorem 21, we fix $p \overset{\text{def}}{=} \Theta(1/\log s)$, and apply Lemma 24 to the depth-2 subcircuits of C. By the same argument and after renaming input variables, this provides a balanced restriction ρ on $\{x_1, \ldots, x_m\}$ with $m \overset{\text{def}}{=} |\rho^{-1}(*)| = \Theta(n/\log s)$ and $(1/s^2, 1/s^2)$-error probabilistic polynomials $\boldsymbol{P}_i(x_1, \ldots, x_m)$ of degree $O(\log s)$ for each (ρ-restricted) depth-2 subcircuit C'_i of C.

We apply now the construction in Corollary 15 to the top $d-2$ layers of $C|_\rho$, replacing its depth-2 subcircuits by the probabilistic polynomials $\boldsymbol{P}_i(x_1, \ldots, x_m)$ obtained above. Adapting parameters in a straightforward way, this argument shows that $C|_\rho$ satisfies

$$\zeta \overset{\text{def}}{=} \min\{\mathrm{pdeg}_{(1/10),(1/s)}(C|_\rho), \mathrm{pdeg}_{(1/s),(1/10)}(C|_\rho)\} \;\leq\; O(c\log s)^{d-2}.$$

Moreover, since ρ is a balanced restriction the function computed by $C|_\rho$ is precisely Majority on m input bits.

We can assume w.l.o.g. that $s \leq 2^{\gamma\sqrt{n}}$ for a small enough (universal) constant $\gamma > 0$ independent of n and d. This allows us to invoke Lemma 16, which implies that $\zeta = \Omega(\sqrt{m \cdot \log s})$. Using the previously obtained upper bound on ζ and the value of m completes the proof of Theorem 2. ◀

References

1 Miklós Ajtai. $\sum_1^1$-formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1–48, 1983.

2 Miklós Ajtai and Michael Ben-Or. A Theorem on Probabilistic Constant Depth Computations. In *Symposium on Theory of Computing* (STOC), pages 471–474, 1984.

3 Josh Alman and Ryan Williams. Probabilistic Polynomials and Hamming Nearest Neighbors. In *Symposium on Foundations of Computer Science* (FOCS), pages 136–150, 2015.

4 Kazuyuki Amano. Bounds on the Size of Small Depth Circuits for Approximating Majority. In *International Colloquium on Automata, Languages and Programming* (ICALP), pages 59–70, 2009.

5 James Aspnes, Richard Beigel, Merrick L. Furst, and Steven Rudich. The Expressive Power of Voting Polynomials. *Combinatorica*, 14(2):135–148, 1994.

6 David A. Mix Barrington, Richard Beigel, and Steven Rudich. Representing Boolean Functions as Polynomials Modulo Composite Numbers. *Computational Complexity*, 4:367–382, 1994. doi:10.1007/BF01263424.

7 Paul Beame. A switching lemma primer. Technical report, UW-CSE-95-07-01, 1994.

8 Ravi B. Boppana. Threshold Functions and Bounded Depth Monotone Circuits. In *Symposium on Theory of Computing* (STOC), pages 475–479, 1984. doi:10.1145/800057.808716.

9 Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay Tal. Pseudorandom Generators from the Second Fourier Level and Applications to AC^0 with Parity Gates. In *Innovations in Theoretical Computer Science Conference* (ITCS), pages 22:1–22:15, 2019.

10 Xi Chen, Igor Carboni Oliveira, Rocco A. Servedio, and Li-Yang Tan. Near-optimal small-depth lower bounds for small distance connectivity. In *Symposium on Theory of Computing* (STOC), pages 612–625, 2016.

11 Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, Circuits, and the Polynomial-Time Hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984.

12 Frederic Green. A complex-number Fourier technique for lower bounds on the Mod-m degree. *Computational Complexity*, 9(1):16–38, 2000.

13 Kristoffer Hansen and Vladimir Podolskii. Exact Threshold Circuits. In *Conference on Computational Complexity* (CCC), pages 270–279, 2010.

14 Johan Håstad. Almost Optimal Lower Bounds for Small Depth Circuits. In *Symposium on Theory of Computing* (STOC), pages 6–20, 1986.

15 Peter Keevash and Benny Sudakov. Set Systems with Restricted Cross-Intersections and the Minimum Rank of Inclusion Matrices. *SIAM J. Discrete Math.*, 18(4):713–727, 2005.

16 Swastik Kopparty. AC^0 lower bounds and pseudorandomness, 2013. Lecture notes on 'Topics in Complexity Theory and Pseudorandomness'. Can be found at: `http://sites.math.rutgers.edu/~sk1233/courses/topics-S13/lec4.pdf`.

17 Swastik Kopparty and Srikanth Srinivasan. Certifying Polynomials for $\mathrm{AC}^0[\oplus]$ Circuits, with Applications to Lower Bounds and Circuit Compression. *Theory of Computing*, 14(1):1–24, 2018.

18 Nutan Limaye, Karteek Sreenivasiah, Srikanth Srinivasan, Utkarsh Tripathi, and S. Venkitesh. The Coin Problem in Constant Depth: Sample Complexity and Parity gates. *Electronic Colloquium on Computational Complexity* (ECCC), 157, 2018.

19 Alexis Maciel and Denis Thérien. Threshold Circuits of Small Majority-Depth. *Inf. Comput.*, 146(1):55–83, 1998. `doi:10.1006/inco.1998.2732`.

20 Zipei Nie and Anthony Y Wang. Hilbert functions and the finite degree Zariski closure in finite field combinatorial geometry. *Journal of Combinatorial Theory, Series A*, 134:196–220, 2015.

21 Ryan O'Donnell and Karl Wimmer. Approximation by DNF: Examples and Counterexamples. In *International Colloquium on Automata, Languages and Programming* (ICALP), pages 195–206, 2007.

22 Ramamohan Paturi, Michael E. Saks, and Francis Zane. Exponential lower bounds for depth three Boolean circuits. *Computational Complexity*, 9(1):1–15, 2000. `doi:10.1007/PL00001598`.

23 Alexander A. Razborov. Lower bounds on the size of constant-depth networks over a complete basis with logical addition. *Mathematicheskie Zametki*, 41(4):598–607, 1987.

24 Benjamin Rossman and Srikanth Srinivasan. Separation of $\mathrm{AC}^0[\oplus]$ Formulas and Circuits. In *International Colloquium on Automata, Languages, and Programming* (ICALP), pages 50:1–50:13, 2017.

25 Roman Smolensky. Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity. In *Symposium on Theory of Computing* (STOC), pages 77–82, 1987.

26 Roman Smolensky. On Representations by Low-Degree Polynomials. In *Symposium on Foundations of Computer Science* (FOCS), pages 130–138, 1993.

27 Mario Szegedy. *Algebraic Methods in Lower Bounds for Computational Models.* PhD thesis, University of Chicago, 1989.

28 Leslie G. Valiant. Short Monotone Formulae for the Majority Function. *J. Algorithms*, 5(3):363–366, 1984.

29 Victor K. Wei. Generalized Hamming weights for linear codes. *IEEE Transactions on Information Theory*, 37(5):1412–1418, 1991.