THEORY OF COMPUTING, Volume 15 (17), 2019, pp. 1–20
www.theoryofcomputing.org

Separation of $\mathsf{AC}^0[\oplus]$ Formulas and Circuits

Benjamin Rossman* Srikanth Srinivasan†

Received August 11, 2017; Revised October 3, 2019; Published December 17, 2019

Abstract: We give the first separation between the power of formulas and circuits in the $\mathsf{AC}^0[\oplus]$ basis (unbounded fan-in AND, OR, NOT and MOD_2 gates). We show that there exist $\text{poly}(n)$-size depth-d circuits that are not equivalent to any depth-d formula of size $n^{o(d)}$ for all $d \leq O(\log(n)/\log\log(n))$. This result is obtained by a combination of new lower and upper bounds for *Approximate Majorities*, the class of Boolean functions $\{0,1\}^n \to \{0,1\}$ that agree with the Majority function on a $3/4$ fraction of the inputs.

$\mathsf{AC}^0[\oplus]$ *formula lower bound.* We show that every depth-d $\mathsf{AC}^0[\oplus]$ formula of size s has a $1/4$-error polynomial approximation over $\mathbb{F}_2$ of degree $O((1/d)\log s)^{d-1}$. This strengthens a classic $O(\log s)^{d-1}$ degree approximation for circuits due to Razborov (1987). Since any polynomial that approximates the Majority function has degree $\Omega(\sqrt{n})$, this result implies an $\exp(\Omega(dn^{1/2(d-1)}))$ lower bound on the depth-d $\mathsf{AC}^0[\oplus]$ formula size of all Approximate Majority functions for all $d \leq O(\log n)$.

Monotone AC^0 *circuit upper bound.* For all $d \leq O(\log(n)/\log\log(n))$, we give a randomized construction of depth-d monotone AC^0 circuits (without NOT or MOD_2 gates) of size $\exp(O(n^{1/2(d-1)}))$ that compute an Approximate Majority function. This strengthens a construction of formulas of size $\exp(O(dn^{1/2(d-1)}))$ due to Amano (2009).

ACM Classification: F.1.3, F.2.3

AMS Classification: 68Q17, 68Q15

Key words and phrases: formulas, circuits, lower bounds, AC0

A preliminary version of this paper appeared in the Proc. of the 44th International Colloquium on Automata, Languages, and Programming (ICALP 2017).

*Supported by NSERC and a Sloan Foundation Fellowship.
†Supported by MATRICS grant MTR/2017/000958 awarded by SERB, Government of India.

 DOI: 10.4086/toc.2019.v015a017

1 Introduction

The relative power of formulas versus circuits is one of the great mysteries in complexity theory. The central question in this area is whether NC^1 (the class of languages decidable by polynomial-size Boolean formulas) is a proper subclass of P/poly (the class of languages decidable by polynomial-size Boolean circuits). Despite decades of efforts, this question remains wide open.[1] In the meantime, there has been progress on analogues of the NC^1 vs. P/poly question in certain restricted settings. For instance, in the monotone basis (with AND and OR gates only), the power of polynomial-size formulas vs. circuits was separated by the classic lower bound of Karchmer and Wigderson [9] (on the monotone formula size of st-Connectivity).

The bounded-depth setting is another natural venue for investigating the question of formulas vs. circuits. Consider the elementary fact that every depth-d circuit of size s is equivalent to a depth-d formula of size at most s^{d-1}, where we measure *size* by the number of gates. This observation is valid with respect to any basis (i. e., set of gate types). In particular, we may consider the AC^0 basis (unbounded fan-in AND, OR, NOT gates) and the $\mathsf{AC}^0[\oplus]$ basis (unbounded fan-in MOD_2 gates in addition to AND, OR, NOT gates). With respect to either basis, there is a natural depth-d analogue of the NC^1 vs. P/poly question (where $d = d(n)$ is a parameter that may depend on n), namely whether every language decidable by polynomial-size depth-d circuits is decidable by depth-d formulas of size $n^{o(d)}$ (i. e., better than the trivial $n^{O(d)}$ upper bound).

It is reasonable to expect that this question could be resolved in the sublogarithmic-depth regime $d(n) = o(\log n)$ given the powerful lower-bound techniques against AC^0 circuits (Håstad's Switching Lemma [6]) and $\mathsf{AC}^0[\oplus]$ circuits (the Polynomial Method of Razborov [13] and Smolensky [16]). However, because the standard way of applying these techniques does not distinguish between circuits and formulas, it is not clear how to prove quantitatively stronger lower bounds on formula size vis-a-vis circuit size of a given function. Recent work of Rossman [14] developed a new way of applying Håstad's Switching Lemma to AC^0 formulas, in order to prove an $\exp(\Omega(dn^{1/(d-1)}))$ lower bound on the formula size of the Parity function for all $d \le O(\log n)$. Combined with the well-known $\exp(O(n^{1/(d-1)}))$ upper bound on the circuit size of Parity, this yields an asymptotically optimal separation in the power of depth-d AC^0 formulas vs. circuits for all $d(n) \le O(\log(n)/\log\log(n))$, as well as a super-polynomial separation for all $\omega(1) \le d(n) \le o(\log n)$.

In the present paper, we carry out a similar development for formulas vs. circuits in the $\mathsf{AC}^0[\oplus]$ basis, obtaining both an asymptotically optimal separation for all $d(n) \le O(\log(n)/\log\log(n))$ and a super-polynomial separation for all $\omega(1) \le d(n) \le o(\log n)$. Our target functions lie in the class of *Approximate Majorities*, here defined as Boolean functions $\{0,1\}^n \to \{0,1\}$ that agree with the Majority function on a $3/4$ fraction of the inputs. First, we show how to apply the Polynomial Method to obtain better parameters in the approximation of $\mathsf{AC}^0[\oplus]$ formulas by low-degree polynomials over $\mathbb{F}_2$. This leads to an $\exp(\Omega(dn^{1/2(d-1)}))$ lower bound on the $\mathsf{AC}^0[\oplus]$ formula size of all Approximate Majority functions. The other half of our formulas vs. circuits separation comes from an $\exp(O(n^{1/2(d-1)}))$ upper bound on the $\mathsf{AC}^0[\oplus]$ circuit size of some Approximate Majority function. In fact, this upper bound is realized by a randomized construction of monotone AC^0 circuits (AC^0 circuits without NOT or MOD_2 gates).

[1] In this paper we focus on non-uniform complexity classes. The question of uniform-NC^1 vs. P is wide open as well.

Theorem 1.1. *Let $n \in \mathbb{N}$ be a growing parameter.*

1. *For all $2 \le d \le O(\log n)$, the depth-d $\mathsf{AC}^0[\oplus]$ formula size of any n-variable Approximate Majority function is $\exp(\Omega(dn^{1/2(d-1)}))$.*

2. *For all*
$$2 \le d \le \frac{1}{2} \cdot \frac{\log n}{\log\log n + O(1)},$$
there exists an n-variable Approximate Majority function whose depth-d monotone AC^0 circuit size is $\exp(O(n^{1/2(d-1)}))$.

(All asymptotic notation $O(\cdot)$ and $\Omega(\cdot)$ hides constants independent of d.) Denote by $\mathsf{AC}^0[\oplus]$-circuit$[d(n), s(n)]$ and $\mathsf{AC}^0[\oplus]$-formula$[d(n), s(n)]$ the class of languages decidable by $\mathsf{AC}^0[\oplus]$-circuits and formulas, resp., of depth $d(n)$ and size $s(n)$. For all functions $d(n)$, we have the inclusions

$$\mathsf{AC}^0[\oplus]\text{-formula}[d(n), \mathrm{poly}(n)] \subseteq \mathsf{AC}^0[\oplus]\text{-circuit}[d(n), \mathrm{poly}(n)] \subseteq \mathsf{AC}^0[\oplus]\text{-formula}[d(n), n^{O(d(n))}].$$

Theorem 1.1 implies that the above inclusions are tight for small $d(n)$.

Corollary 1.2. *Let $n \in \mathbb{N}$ be a growing parameter.*

(i) *For all functions $2 \le d(n) \le O\big(\frac{\log n}{\log\log n}\big)$,*
$$\mathsf{AC}^0[\oplus]\text{-circuit}[d(n), \mathrm{poly}(n)] \not\subseteq \mathsf{AC}^0[\oplus]\text{-formula}[d(n), n^{o(d(n))}].$$

(ii) *For all functions $\omega(1) \le d(n) \le o(\log n)$,*
$$\mathsf{AC}^0[\oplus]\text{-formula}[d(n), \mathrm{poly}(n)] \subsetneq \mathsf{AC}^0[\oplus]\text{-circuit}[d(n), \mathrm{poly}(n)].$$

Proof. We prove separation (i) using a padding argument. Let $c > 0$ be a constant such that Theorem 1.1(2) holds for all
$$2 \le d \le \frac{1}{2} \cdot \frac{\log n}{\log\log n + c} + 1.$$
And let $\{f_{n,d}\}$ be the family of Approximate Majority functions from Theorem 1.1(2). Now suppose
$$2 \le d \le (1/2)^{c+1} \frac{\log n}{\log\log n} + 1.$$
Letting $m = (\log n)^{2(d-1)}$, we have
$$d - 1 \le (1/2)^{c+1} \frac{\log n}{\log\log n} = (d-1)(1/2)^c \frac{m^{1/2(d-1)}}{\log m}.$$
Dividing both sides by $d-1$ and taking logs, this rearranges to
$$d \le \frac{1}{2} \cdot \frac{\log m}{\log\log m + c} + 1.$$

By Theorem 1.1, $f_{m,d}$ has $\mathsf{AC}^0[\oplus]$ circuit size $\mathrm{poly}(n)$ and $\mathsf{AC}^0[\oplus]$ formula size $n^{\Omega(d)}$. Separation (i) now follows by padding $f_{m,d}$ to a function on n variables (noting that $m < n$).

Separation (ii) follows from (i) when

$$d \leq \frac{1}{2} \cdot \frac{\log n}{\log\log n + c} + 1.$$

For larger depth, separation (ii) follows from Theorem 1.1, noting that the lower bound $\exp(\Omega(dn^{1/2(d-1)}))$ is super-polynomial for $d(n) = o(\log n)$. □

We point out that extending separation (ii) from depth $o(\log n)$ to depth $\log n$ is equivalent to separating complexity classes NC^1 ($= \mathsf{AC}^0[\oplus]$-formula$[\log n, \mathrm{poly}(n)]$) from AC^1 ($= \mathsf{AC}^0[\oplus]$-circuit$[\log n, \mathrm{poly}(n)]$).

1.1 Proof outline

Improved polynomial approximation. The lower bound for $\mathsf{AC}^0[\oplus]$ formulas follows the general template due to Razborov [13] on proving lower bounds for $\mathsf{AC}^0[\oplus]$ circuits using low-degree polynomials over $\mathbb{F}_2$. Razborov showed that for any Boolean function $f : \{0,1\}^n \to \{0,1\}$ that has an $\mathsf{AC}^0[\oplus]$ circuit of size s and depth d, there is a randomized polynomial $\mathbf{P}$ of degree $O(\log s)^{d-1}$ that computes f correctly on each input with probability $7/8$ (we call such polynomials $1/8$-error *probabilistic polynomials*). By showing that some explicit Boolean function f (e. g., the Majority function or the MOD_q function for q odd) on n variables does not have such an approximation of degree less than $\Omega(\sqrt{n})$ [13, 16, 18, 17], we get that any $\mathsf{AC}^0[\oplus]$ circuit of depth d computing f must have size $\exp(\Omega(n^{1/2(d-1)}))$.

In this paper, we improve the parameters of Razborov's polynomial approximation from above for $\mathsf{AC}^0[\oplus]$ *formulas*. More precisely, for $\mathsf{AC}^0[\oplus]$ formulas of size s and depth d, we are able to construct $1/8$-error probabilistic polynomials of degree $O((1/d)\log s)^{d-1}$. (Since every depth-d circuit of size s is equivalent to a depth-d formula of size at most s^{d-1}, this result implies Razborov's original theorem that $\mathsf{AC}^0[\oplus]$ circuits of size s and depth d have $1/8$-error probabilistic polynomials of degree $O(\log s)^{d-1}$.)

We illustrate the idea behind this improved polynomial approximation with the special case of a balanced formula (i. e., all gates have the same fan-in) of fan-in t and depth d. Note that the size of the formula (number of gates) is $\Theta(t^{d-1})$ and hence it suffices in this case to show that it has a $1/8$-error probabilistic polynomial of degree $O(\log t)^{d-1}$. We construct the probabilistic polynomial inductively. Given a balanced formula F of depth d and fan-in t, let $F_1, \ldots, F_t$ be its subformulas of depth $d-1$. Inductively, each F_i has a $1/8$-error probabilistic polynomial $\mathbf{P}_i$ of degree $O(\log t)^{d-2}$ and by a standard error-reduction [11], it has a $(1/16t)$-error probabilistic polynomial of degree $O(\log t)^{d-1}$ (in particular, at any given input $x \in \{0,1\}^n$, the probability that *there exists* an $i \in [t]$ such that $\mathbf{P}_i(x) \neq F_i(x)$ is at most $1/16$). Using Razborov's construction of a $1/16$-error probabilistic polynomial of degree $O(1)$ for the output gate of F and composing this with the probabilistic polynomials $\mathbf{P}_i$, we get the result for balanced formulas. This idea can be extended to general (i. e., not necessarily balanced) formulas with a careful choice of the error parameter for each subformula F_i to obtain the stronger polynomial approximation result.

Improved formula lower bounds. Combining the above approximation result with known lower bounds for polynomial approximation [13, 16, 18, 17], we can already obtain stronger lower bounds for

$\mathsf{AC}^0[\oplus]$ formulas than are known for $\mathsf{AC}^0[\oplus]$ circuits. For instance, it follows that any $\mathsf{AC}^0[\oplus]$ formula of depth d computing the Majority function on n variables must have size $\exp(\Omega(dn^{1/2(d-1)}))$ for all $d \leq O(\log n)$, which is stronger than the corresponding circuit lower bound. Similarly stronger formula lower bounds also follow for the MOD_q function (q odd).

Separation between formulas and circuits. However, the above improved lower bounds do not directly yield the claimed separation between $\mathsf{AC}^0[\oplus]$ formulas and circuits. This is because we do not have circuits computing (say) the Majority function of the required size. To be able to prove our result, we would need to show that the Majority function has $\mathsf{AC}^0[\oplus]$ circuits of depth d and size $\exp(O(n^{1/2(d-1)}))$ (where the constant in the $O(\cdot)$ is independent of d). However, as far as we know, the strongest result in this direction [10] only yields $\mathsf{AC}^0[\oplus]$ circuits of size greater than $\exp(\Omega(n^{1/(d-1)}))$,[2] which is superpolynomially larger than the upper bound.

To circumvent this issue, we change the hard functions to the class of *Approximate Majorities*, which is the class of Boolean functions that agree with the Majority function on a $3/4$ fraction of inputs from $\{0,1\}^n$. (The choice of the constant $3/4$ is not crucial: it can be replaced by any constant in the interval $(1/2,1)$.) While this has the downside that we no longer are dealing with an explicitly defined function, the advantage is that the polynomial approximation method of Razborov yields tight lower bounds for some functions from this class.

Indeed, since the method of Razborov is based on polynomial approximations, it immediately follows that the same proof technique also yields the same lower bound for computing Approximate Majorities. Formally, any $\mathsf{AC}^0[\oplus]$ circuit of depth d computing *any* Approximate Majority must have size $\exp(\Omega(n^{1/2(d-1)}))$. On the upper bound side, it is known from the work of O'Donnell and Wimmer [12] and Amano [1] that there *exist* Approximate Majorities that can be computed by monotone AC^0 *formulas* of depth d and size $\exp(O(dn^{1/2(d-1)}))$. (Note that the double exponent $1/(2(d-1))$ is now the same in the upper and lower bounds.)

In order to separate $\mathsf{AC}^0[\oplus]$ formulas and circuits, we show that both the upper and lower bounds from the previous paragraph are actually tight (up to the constants implicit in the exponents). Plugging in our stronger polynomial approximation for $\mathsf{AC}^0[\oplus]$ formulas, we obtain that any $\mathsf{AC}^0[\oplus]$ formula of depth d computing any Approximate Majority must have size $\exp(\Omega(dn^{1/2(d-1)}))$. In particular, this implies that Amano's construction is tight (up to the universal constant in the exponent) even for $\mathsf{AC}^0[\oplus]$ formulas.

Further, we also modify Amano's construction [1] to obtain better constant-depth *circuits* for Approximate Majorities: we show that there exist Approximate Majorities that are computed by monotone AC^0 circuits of depth d of size $\exp(O(n^{1/2(d-1)}))$ (the constant in the $O(\cdot)$ is independent of d).

Smaller circuits for Approximate Majority. Our construction closely follows Amano's, which in turn is related to Valiant's probabilistic construction [19] of monotone formulas for the Majority function. However, we need to modify the construction in a suitable way that exploits the fact that we are constructing *circuits*. This modification is in a similar spirit to a construction of Hoory, Magen and

[2] Indeed, this is inevitable with all constructions that we are aware of, since they are actually AC^0 circuits and it is known by a result of Håstad [6] that any AC^0 circuit of depth d for the Majority function must have size $\exp(\Omega(n^{1/(d-1)}))$.

Pitassi [7] who modify Valiant's construction to obtain smaller monotone circuits (of depth $\Theta(\log n)$) for computing the Majority function exactly.

At a high level, the difference between Amano's construction and ours is as follows. Amano constructs random formulas F_i of each depth $i \le d$ as follows. The formula F_1 is the AND of a_1 independent and randomly chosen variables. For even $i > 1$, F_i is the OR of a_i independent and random copies of F_{i-1}, while for odd $i > 1$, F_i is the AND of a_i independent and random copies of F_{i-1}. For suitable values of $a_1, \ldots, a_d \in \mathbb{N}$, the random formula F_d computes an Approximate Majority with high probability. In our construction, we build a depth i circuit C_i for each $i \le d$ in a similar way, except that each C_i now has M different outputs. Given such a C_{i-1}, we construct C_i by taking M independent randomly chosen subsets $T_1, \ldots, T_M$ of a_i many outputs of C_{i-1} and adding gates that compute either the OR or AND (depending on whether i is even or odd) of the gates in T_i. Any of the M final gates of C_d now serves as the output gate. By an analysis similar to Amano's (see also [7]) we can show that this computes an Approximate Majority with high probability, which finishes the proof.[3]

2 Preliminaries

Throughout, n will be a growing parameter. We will consider Boolean functions on n variables, i. e., functions of the form $f : \{0,1\}^n \to \{0,1\}$. We will sometimes identify $\{0,1\}$ with the field $\mathbb{F}_2$ in the natural way and consider functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$ instead.

Given a Boolean vector $y \in \{0,1\}^n$, we use $|y|_0$ to denote the number of 0s in y and $|y|_1$ to denote the number of 1s in y.

The Majority function on n variables, denoted MAJ_n is the Boolean function that maps inputs $x \in \{0,1\}^n$ to 1 if and only if $|x|_1 > n/2$.

Definition 2.1. An (ε, n)-*Approximate Majority* is a function $f : \{0,1\}^n \to \{0,1\}$ such that

$$\Pr_{x \in \{0,1\}^n}[f(x) \neq \mathrm{Maj}_n(x)] \le \varepsilon.$$

As far as we know, the study of this class of functions was initiated by O'Donnell and Wimmer [12]. See also [1, 4].

We refer the reader to [2, 8] for standard definitions of Boolean circuits and formulas. We use the terms AC^0 *circuits* and AC^0 *formulas* to describe and formulas of constant depth, made up of AND, OR and NOT gates. Similarly, $\mathsf{AC}^0[\oplus]$ *circuits* and $\mathsf{AC}^0[\oplus]$ *formulas* will be circuits and formulas of constant depth, made up of AND, OR, MOD_2 and NOT gates.

Remark 2.2. In standard terminology, AC^0 and $\mathsf{AC}^0[\oplus]$ circuits refer to circuits of *constant* depth and *polynomial* size (made up of the relevant sets of gates in each case). However, our results hold for more general settings of these parameters. To phrase these results, it will be convenient to abuse notation and refer to AC^0 and $\mathsf{AC}^0[\oplus]$ circuits of size $s(n)$ and depth $d(n)$ where $s(n)$ is possibly superpolynomial and $d(n)$ possibly superconstant, which of course can be defined in the obvious way.

[3] This is a slightly imprecise description of the construction as the final two levels of the circuit are actually defined somewhat differently.

By the *size of a circuit* we mean the number of *gates* in the circuit. (Note that this, too, is a deviation from standard terminology that refers to the number of wires as the "size" of the circuit. However, since the number of wires is polynomially related to the number of gates, Theorem 1.1 and Corollary 1.2 also hold when the size denotes the number of wires.) By the *size of a formula* we mean the number of its *leaves* which is, up to a constant factor, equal to the number of all nodes (including leaves) in the formula.[4]

3 Lower bound

In this section, we show that any $\mathsf{AC}^0[\oplus]$ formulas of depth d computing a $(1/4,n)$-Approximate Majority must have size at least $\exp(\Omega(dn^{1/2(d-1)}))$ for all $d \leq O(\log n)$.

We work over the field $\mathbb{F}_2$ and identify it with $\{0,1\}$ in the natural way. The following concepts are standard in circuit complexity (see, e. g., Beigel's survey [3]).

Definition 3.1. Fix any $\varepsilon \in [0,1]$. A polynomial $P \in \mathbb{F}_2[X_1,\dots,X_n]$ is said to be an *ε-approximating polynomial* for a Boolean function $f : \{0,1\}^n \to \{0,1\}$ if

$$\Pr_{x\in\{0,1\}^n}[f(x) = P(x)] \geq 1-\varepsilon.$$

The following lemma, motivated by the ideas of Smolensky [16] (1987), first appeared explicitly in Szegedy's Ph. D. Thesis [18, Theorem 2.4.7] (1989) and in Smolensky's paper [17, Theorem 5] (1993).

Lemma 3.2 (Smolensky, Szegedy). *Let $\varepsilon \in (0,1/2)$ be any fixed constant. Any $(1/2-\varepsilon)$-approximating polynomial for the Majority function on n variables must have degree $\Omega(\sqrt{n})$.*

Corollary 3.3. *Let f be any $(1/4,n)$-Approximate Majority and $\varepsilon \in (0,1/4)$ an arbitrary constant. Then any $(1/4-\varepsilon)$-approximating polynomial for f must have degree $\Omega(\sqrt{n})$.*

Proof. The proof is immediate from Lemma 3.2 and the triangle inequality. □

Definition 3.4. An *ε-error probabilistic polynomial of degree D* for a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is a random variable $\mathbf{P}$ taking values from polynomials in $\mathbb{F}_2[X_1,\dots,X_n]$ of degree at most D such that for all $x \in \{0,1\}^n$, we have $\Pr[\, f(x) = \mathbf{P}(x)\,] \geq 1-\varepsilon$.

Definition 3.5. Let $D_\varepsilon(f)$ be the minimum degree of an ε-error probabilistic polynomial for f.

We will make use of the following two lemmas concerning $D_\varepsilon(\cdot)$.

Lemma 3.6 (Razborov [13]). *Let OR_n and AND_n be the OR and AND functions in n variables, respectively. Then $D_\varepsilon(\mathrm{OR}_n), D_\varepsilon(\mathrm{AND}_n) \leq \lceil \log(1/\varepsilon) \rceil$.*

Lemma 3.7 (Kopparty and Srinivasan [11]). *There is an absolute constant c_1 such that for any $\varepsilon \in (0,1)$, $D_\varepsilon(f) \leq c_1 \cdot \lceil \log(1/\varepsilon) \rceil \cdot D_{1/8}(f)$ for all Boolean functions f.*

[4] We assume here without loss of generality that the formula does not contain a gate of fan-in 1 feeding into another gate.

We now state our main result, which shows that every $\mathsf{AC}^0[\oplus]$ formula of size s and depth $d+1$ admits a $1/8$-error approximating polynomial of degree $O((1/d)\log s)^d$.

Theorem 3.8. *There is an absolute constant c_2 such that, if f is computed by an $\mathsf{AC}^0[\oplus]$ formula F of size s and depth $d+1$, then*

$$D_{1/8}(f) \le 3\left(c_2\left(\frac{1}{d}\log(s)+1\right)\right)^d.$$

Proof. The proof is an induction on the depth d of the formula.

The base case $d=0$ corresponds to the case when the formula is a single AND, OR or MOD_2 gate and we need to show that $D_{1/8}(f)\le 3$. In the case that the formula is an AND or OR gate, this follows from Lemma 3.6. If the formula is a MOD_2 gate, this follows from the fact that the MOD_2 function is exactly a polynomial of degree 1.

Let $d\ge 1$. We assume that the formula F is the AND/OR/MOD_2 of subformulas $F_1,\ldots,F_m$ computing $f_1,\ldots,f_m$ where F_i has size s_i and depth $d+1$. So F has size $s=s_1+\cdots+s_m$ and depth $d+2$. Assume that

$$D_{1/8}(f_i) \le 3\left(c_2\left(\frac{1}{d}\log(s_i)+1\right)\right)^d$$

for all i. We must show that

$$D_{1/8}(f) \le 3\left(c_2\left(\frac{1}{d+1}\log(s)+1\right)\right)^{d+1}.$$

By Lemma 3.7, each f_i has an $s_i/(16s)$-error probabilistic polynomial $\mathbf{P}_i$ of degree $c_1\cdot\lceil\log(16s/s_i)\rceil\cdot D_{1/8}(f_i)$, which is at most

$$3c_1\cdot 5(\log(s/s_i)+1)\cdot\left(c_2\left(\frac{1}{d}\log(s_i)+1\right)\right)^d.$$

Then $(\mathbf{P}_1,\ldots,\mathbf{P}_m)$ jointly computes $(f_1,\ldots,f_m)$ with error $1/16$ $(=\sum_{i=1}^m(s_i/(16s)))$.

By a reasoning identical to the base case, it follows that there exists a $1/16$-error probabilistic polynomial $\mathbf{Q}$ of degree 4 for the output gate of the formula.

Then $\mathbf{Q}(\mathbf{P}_1,\ldots,\mathbf{P}_m)$ is a $1/8$-error probabilistic polynomial for f of degree

$$60c_1\cdot\max_i\{(\log(s/s_i)+1)\}\cdot\left(c_2\left(\frac{1}{d}\log(s_i)+1\right)\right)^d.$$

So long as $c_2\ge 20c_1$, it suffices to show that for all i,

$$(\log(s/s_i)+1)\cdot\left(\frac{1}{d}\log(s_i)+1\right)^d \le \left(\frac{1}{d+1}\log(s)+1\right)^{d+1}.$$

Consider any i and let $a,b\ge 0$ such that $s_i=2^a$ and $s=2^{a+b}$. We must show

$$(b+1)\left(\frac{a}{d}+1\right)^d \le \left(\frac{a+b}{d+1}+1\right)^{d+1}.$$

For fixed $a \geq 0$, as a polynomial in b, the function

$$p_{a,d}(b) := \left(\frac{a+b}{d+1}+1\right)^{d+1} - (b+1)\left(\frac{a}{d}+1\right)^{d}$$

is nonnegative over $b \geq 0$ with a unique root at $b = a/d$. This follows from

$$\frac{\partial}{\partial b} p_{a,d}(b) = \left(\frac{a+b}{d+1}+1\right)^{d} - \left(\frac{a}{d}+1\right)^{d},$$

which is zero iff $b = a/d$; this value is a minimum of $p_{a,d}$ with $p_{a,d}(a/d) = 0$. □

Corollary 3.9. *Let $n \in \mathbb{N}$ be a growing parameter and $d = d(n)$ (possibly a constant). Let f be any $(1/4,n)$-Approximate Majority. Then any $\mathsf{AC}^0[\oplus]$ formula of depth d computing f must have size $\exp(\Omega(dn^{1/2(d-1)}))$ for all $d \leq O(\log n)$, where the asymptotic notations $O(\cdot)$ and $\Omega(\cdot)$ hide absolute constants (independent of d and n).*

Proof. Say that F is an $\mathsf{AC}^0[\oplus]$ formula of depth d and size s computing f. Then, by Lemma 3.7, we see that F has a $1/8$-error probabilistic polynomial $\mathbf{P}$ of degree $D \leq O((1/d)\log s + 1)^{d-1}$. In particular, by an averaging argument, there is some fixed polynomial $P \in \mathbb{F}_2[X_1,\ldots,X_n]$ of degree at most D such that P is a $1/8$-error approximating polynomial for f.

Corollary 3.3 implies that the degree of P must be $\Omega(\sqrt{n})$. Hence, we obtain $O((1/d)\log s+1)^{d-1} \geq \Omega(\sqrt{n})$. It follows that $s \geq \exp(\Omega(dn^{1/2(d-1)}) - O(d))$. Observe that $\Omega(dn^{1/2(d-1)})$ dominates $O(d)$ so long as $d \leq \varepsilon \log n$ for some absolute constant $\varepsilon > 0$ (depending on the constants in $\Omega(\cdot)$ and $O(\cdot)$). Hence, we get the claimed lower bound $s \geq \exp(\Omega(dn^{1/2(d-1)}))$ for all $d \leq \varepsilon \log n$. □

4 Upper bound

In this section, we show that for any constant ε, there are (ε,n)-Approximate Majorities that can be computed by depth-d AC^0 circuits of size $\exp(O(n^{1/2(d-1)}))$.

4.1 High-level idea

We start with an informal description of our construction, which closely follows the constructions of O'Donnell and Wimmer [12] and Amano [1] (see also [19, 7]).

The first basic observation, from [12], is that to compute an (ε,n)-Approximate majority, it suffices to compute the Majority function on most (say at least a $1-(\varepsilon/4)$ fraction of) inputs of relative Hamming weight bounded away from $1/2$; say those of weight either at most $1/2-(\varepsilon/4\sqrt{n})$ or at least $1/2+(\varepsilon/4\sqrt{n})$. So we assume that the inputs come from one of these sets. Let us call the sets of high-weight inputs and low-weight inputs Y (for "Yes") and N (for "No"), respectively.

The problem of distinguishing inputs from Y and N is easier than computing Majority exactly since there is a gap between the weights of the yes and no instances. In particular, given any inputs $a \in Y$ and $b \in N$, we have that the multiplicative gap between the weights of a and b is $(1+\gamma_0)$, where $\gamma_0 = \Theta(\varepsilon/\sqrt{n})$.

The idea is to iteratively increase this gap at each level of the circuit, so that after $(d-1)$ levels, the multiplicative gap is a large constant, at which point a simple OR or AND determines the answer on most inputs.

The basic observation that allows us to do this (formalized in Lemma 4.1 below) is the following. Say we have an input $x \in Y' \cup N'$ where Y' is the set of inputs of relative weight at least $p(1+\gamma)$ and N' is the set of inputs of relative weight at most $p(1-\gamma)$ (for some known p,γ). Consider the depth-1 circuit C' that chooses M random subsets of the input bits, of size t each, and computes the OR of the variables in each set (C' is thus an M-output circuit). If the input x has (relative) weight at least $p(1+\gamma)$, then the expected fraction of 0s in the output—let us call this the 0-*weight* of the output—is at most

$$(1-p(1+\gamma))^t \approx \exp(-pt - p\gamma t)$$

while if x has weight at most $p(1-\gamma)$, then the output has 0-weight at least $\exp(-pt+p\gamma t)$. By a simple application of the Chernoff bound and averaging, we can easily ensure that we have a fixed circuit C' that has (roughly) this behaviour at every input x of interest. That is, for inputs $a \in Y'$ and $b \in N'$, the 0-weights of $C'(a)$ and $C'(b)$ have 0-weights that are either at most $\exp(-pt)\cdot\exp(-\gamma pt)$ or at least $\exp(-pt)\cdot\exp(\gamma pt)$. In particular, they are multiplicatively separated by $1+\gamma'$ where $\gamma' \approx (pt)\cdot\gamma$,[5] which is much greater than γ if pt is much greater than 1. In an analogous way, we can construct a depth-1 circuit (made up of ANDs this time) that amplifies a multiplicative gap of $(1+\gamma)$ in the 0-weights to a multiplicative gap of $(1+\gamma')$ in the weights.

Given this fact, our way forward is clear. We stack these depth-1 circuits one on top of the other, at each stage amplifying the multiplicative gap from $1+\gamma$ (for some γ) to $1+r\gamma$.[6] Since we want the gap after $d-1$ levels to be a constant, we take $r \approx (1/\gamma_0)^{1/(d-1)} \approx n^{1/2(d-1)}$.

This in turns determines the size of our constructed circuit. Recall that the depth-1 circuit describes above amplifies the gap between the weights from $1+\gamma$ to $1+(pt)\cdot\gamma$, which we would like to be $1+r\gamma$. At the first level of the circuit, we can take $t \approx r$ as $p = 1/2 = \Theta(1)$. At all subsequent levels, however, we have $p \approx \exp(-r)$ and hence $t \approx r/p$ is $\exp(O(r))$. Hence, we obtain a circuit of size $\exp(O(r)) = \exp(O(n^{1/2(d-1)}))$.

4.2 Formal proof

Let $\varepsilon_0 \in (0,1)$ be a small enough constant so that the following inequalities hold for any $\beta \le \varepsilon_0$.

- $\exp(-\beta) \le 1-\beta\exp(-\beta)$, and
- $1-\beta \ge \exp(-\beta-\beta^2) \ge \exp(-2\beta)$.

(It suffices to take $\varepsilon_0 = 1/2$.)

We need the following technical lemma.

[5] We use the approximation that $\exp(z) \approx (1+z)$ when z is small.

[6] One could also amplify the gaps by varying amounts at different levels of the circuit. This turns out to be below optimal in terms of the size of the circuit constructed.

Lemma 4.1. *Let A,s be positive reals, $M,n \in \mathbb{N}$, and $\gamma \in (1/n, 1/10)$ be such that $\mathrm{e}^A \geq n^3$, $n \geq 1/\varepsilon_0$, and $1 \leq s \leq n$. Define*

$$I_0(\gamma) := \{y \in \{0,1\}^M \mid |y|_1 \leq M\mathrm{e}^{-A}(1-\gamma)\} \quad \textit{and} \quad I_1(\gamma) := \{y \in \{0,1\}^M \mid |y|_1 \geq M\mathrm{e}^{-A}(1+\gamma)\}.$$

If we choose $S \subseteq [M]$ of size $t := \lceil \mathrm{e}^A \cdot s \rceil$ by picking t random elements from M with replacement, then

$$x \in I_0(\gamma) \Rightarrow \Pr_S\left[\bigvee_{j\in S} x_j = 0\right] \geq \exp(-s)\cdot\exp(s\gamma/2),$$
$$x \in I_1(\gamma) \Rightarrow \Pr_S\left[\bigvee_{j\in S} x_j = 0\right] \leq \exp(-s)\cdot\exp(-s\gamma).$$

Further, if $s\gamma \leq \varepsilon_0$, then the above probabilities can be bounded from below and from above by $\exp(-s)\cdot(1+s\gamma\exp(-s\gamma))$ and $\exp(-s)\cdot(1-s\gamma\exp(-s\gamma))$, respectively.

A similar statement can be obtained above for the sets

$$J_1(\gamma) := \{y \in \{0,1\}^M \mid |y|_0 \leq M\mathrm{e}^{-A}(1-\gamma)\} \quad \textit{and} \quad J_0(\gamma) := \{y \in \{0,1\}^M \mid |y|_0 \geq M\mathrm{e}^{-A}(1+\gamma)\},$$

with the event "$\bigvee_{j\in S} x_j = 0$" being replaced by the event "$\bigwedge_{j\in S} x_j = 1$."

Proof. We give the proof only for $I_0(\gamma)$ and $I_1(\gamma)$. The proof for $J_0(\gamma)$ and $J_1(\gamma)$ is similar.

Consider first the case that $x \in I_1(\gamma)$. In this case, we have the following computation.

$$\Pr_S[\bigvee_{j\in S} x_j = 0] \leq \left(1 - \frac{1+\gamma}{\mathrm{e}^A}\right)^{\mathrm{e}^A\cdot s} \leq \exp(-(1+\gamma)\cdot s) \leq \exp(-s)\cdot\exp(-s\gamma). \tag{4.1}$$

The above implies the first upper bound on $\Pr_S[\bigvee_{j\in S} x_j = 0]$ from the lemma statement. When $s\gamma \leq \varepsilon_0$, we further have $\exp(-s\gamma) \leq 1 - s\gamma\exp(-s\gamma)$, which implies the second upper bound. This proves the lemma when $x \in I_1(\gamma)$.

Now consider the case that $x \in I_0(\gamma)$. We have

$$\begin{aligned}
\Pr_S[\bigvee_{j\in S} x_j = 0] &\geq \left(1 - \frac{1-\gamma}{\mathrm{e}^A}\right)^{\mathrm{e}^A\cdot s+1} \\
&\geq \exp\left(\left(-\frac{1-\gamma}{\mathrm{e}^A} - \frac{1}{\mathrm{e}^{2A}}\right)\cdot(\mathrm{e}^A\cdot s+1)\right) \\
&= \exp\left(-s + s\gamma - \frac{s}{\mathrm{e}^A} - \frac{1-\gamma}{\mathrm{e}^A} - \frac{1}{\mathrm{e}^{2A}}\right) \\
&\geq \exp\left(-s + s\gamma - \frac{2s}{\mathrm{e}^A}\right) \\
&= \exp(-s)\cdot\exp\left(s\gamma\left(1 - \frac{2\mathrm{e}^{-A}}{\gamma}\right)\right)
\end{aligned} \tag{4.2}$$

where for the second inequality we have used the fact that since $e^{-A} \leq 1/n^3 \leq \varepsilon_0$, we have

$$1 - \frac{1-\gamma}{e^A} \geq \exp\left(-\frac{1-\gamma}{e^A} - \frac{1}{e^{2A}}\right).$$

Since $e^{-A} \leq 1/n^3 \leq 1/(4n) \leq \gamma/4$, we can bound from below the right hand side of inequality (4.2) by $\exp(-s) \cdot \exp(s\gamma/2)$. Also, note that

$$1 - \frac{2e^{-A}}{\gamma} \geq 1 - \frac{2/n^3}{1/n} = 1 - \frac{2}{n^2} \geq \exp(-1/n) \geq \exp(-\gamma) \geq \exp(-s\gamma).$$

This implies that the righthand side of inequality (4.2) can also be bounded from below by

$$\exp(-s)\exp(s\gamma\exp(-s\gamma)) \geq \exp(-s) \cdot (1 + s\gamma\exp(-s\gamma)),$$

which implies the claim about $\Pr_S[\bigvee_{j \in S} x_j = 0]$ assuming that $x \in I_0(\gamma)$. □

We now prove the main result of this section.

Theorem 4.2. *For any growing parameter $n \in \mathbb{N}$ and*

$$2 \leq d \leq \frac{1}{2} \cdot \frac{\log n}{\log\log n + O(1)}$$

and $\varepsilon > 0$, there is an (ε, n)-Approximate Majority f_n computable by a monotone AC^0 circuit with at most $\exp(O(n^{1/2(d-1)} \log(1/\varepsilon)/\varepsilon))$ many gates, where both $O(\cdot)$'s hide absolute constants (independent of d, ε).

Proof. We assume throughout that ε is a small enough constant and that n is large enough for various inequalities to hold. We will actually construct a monotone circuit of depth d and size

$$\exp\left(O\left(n^{1/2(d-1)} \frac{\log(1/\varepsilon)}{\varepsilon}\right)\right)$$

computing a $(4\varepsilon, n)$-Approximate Majority, which also implies the theorem.

Fix parameter $A = n^{1/2(d-1)} - \alpha$ where $\alpha \in [0, \ln 2]$ is chosen so that $A = A_1 \cdot \ln 2$ for some positive integer A_1. We assume that $A \geq 10 \log n$ (which holds as long as

$$2d \leq \frac{\log n}{\log\log n + c}$$

for a large enough absolute constant $c > 0$) and that $\varepsilon \leq \varepsilon_0$. Let $M = \lceil e^{10A} \rceil$.

Define a sequence of real numbers $\gamma_0, \gamma_1, \ldots, \gamma_{d-2}$ as follows.

$$\gamma_0 = \frac{\varepsilon}{\sqrt{n}},$$
$$\gamma_i = A\gamma_{i-1}\exp(-2A\gamma_{i-1}), \text{ for each } i \in [d-2].$$

It is clear that $\gamma_i \le A^i\gamma_0$ for each $i \in [d-2]$. As a result we also obtain

$$\begin{aligned}\gamma_i &= A^i\gamma_0 \exp(-2A(\gamma_0+\gamma_1+\cdots+\gamma_{i-1}))\\ &\ge A^i\gamma_0\exp(-2\gamma_0 A(1+A+A^2+\cdots+A^{i-1})) \ge A^i\gamma_0\exp(-3A^i\gamma_0).\end{aligned} \tag{4.3}$$

Let

$$Y_\varepsilon = \left\{x \in \{0,1\}^n \;\middle|\; |x|_1 \ge \left(\frac{1}{2}+\frac{\varepsilon}{\sqrt{n}}\right)n\right\},$$
$$N_\varepsilon = \left\{x \in \{0,1\}^n \;\middle|\; |x|_1 \le \left(\frac{1}{2}-\frac{\varepsilon}{\sqrt{n}}\right)n\right\}.$$

The idea is to define a sequence of circuits $C_1, C_2, \ldots, C_{d-2}$ with n inputs and M outputs such that C_i has depth i and iM many (non-input) gates. Further, for odd i,

$$x \in N_\varepsilon \Rightarrow C_i(x) \in I_0(\gamma_i), \quad \text{and} \quad x \in Y_\varepsilon \Rightarrow C_i(x) \in I_1(\gamma_i), \tag{4.4}$$

and similarly for even i,

$$x \in N_\varepsilon \Rightarrow C_i(x) \in J_0(\gamma_i), \quad \text{and} \quad x \in Y_\varepsilon \Rightarrow C_i(x) \in J_1(\gamma_i). \tag{4.5}$$

After this is done, we will add on the top a depth-2 circuit that will reject most inputs from $I_0(\gamma_{d-2})$ or $J_0(\gamma_{d-2})$—depending on whether $(d-2)$ is odd or even—and accept most inputs from $I_1(\gamma_{d-2})$ or $J_1(\gamma_{d-2})$.

We begin with the construction of $C_1, \ldots, C_{d-2}$ which is done by induction.

Construction of C_1. The base case of the induction is the construction of C_1, which is done as follows. We choose M i.i.d. random subsets $T_1, \ldots, T_M \subseteq [n]$ in the following way: for each $i \in [M]$, we sample A_1 random elements of $[n]$ with replacement. Let $b_i^x = \bigwedge_{j \in T_i} x_j$.

If $x \in N_\varepsilon$, then the probability that $b_i^x = 1$ is given by

$$\Pr[b_i^x = 1] \le \left(\frac{1}{2}-\gamma_0\right)^{A_1} = \frac{1}{2^{A_1}}(1-2\gamma_0)^{A_1} \le \frac{1}{e^A}(1-\gamma_0 A_1) \le \frac{1}{e^A}(1-\gamma_0 A),$$

where the second-last inequality follows from the fact that

$$(1-z)^{A_1} \le \left(1 - zA_1 + \frac{A_1^2 z^2}{2}\right)$$

for $z \in [0,1]$.

Let $\delta = 1/n^3$. Note in particular that $2\delta/\gamma_0 A \le \varepsilon_0$ for large enough n.

By a Chernoff bound, the probability that

$$\frac{1}{M}\sum_i b_i^x \ge \frac{1}{e^A}(1-\gamma_0 A)(1+\delta)$$

is bounded by $\exp(-\Omega(\delta^2 M/\mathrm{e}^A)) \le \exp(-\Omega(\mathrm{e}^{9A}/n^6)) \le \exp(-n)$, since $\mathrm{e}^A \ge n^{10}$. Thus, with probability at least $1-\exp(-n)$, we have

$$\begin{aligned}\frac{\sum_i b_i^x}{M} &\le \frac{1}{\mathrm{e}^A}(1-\gamma_0 A)(1+\delta)\\ &\le \frac{1}{\mathrm{e}^A}(1-\gamma_0 A+\delta) = \frac{1}{\mathrm{e}^A}\left(1-\gamma_0 A\left(1-\frac{\delta}{\gamma_0 A}\right)\right)\\ &\le \frac{1}{\mathrm{e}^A}\left(1-\gamma_0 A\exp\left(-\frac{2\delta}{\gamma_0 A}\right)\right) \le \frac{1}{\mathrm{e}^A}(1-\gamma_0 A\exp(-\gamma_0 A))\\ &\le \frac{1}{\mathrm{e}^A}(1-\gamma_1). \end{aligned} \tag{4.6}$$

Above, we have used the fact that

$$\left(1-\frac{\delta}{\gamma_0 A}\right) \ge \exp\left(\frac{-2\delta}{\gamma_0 A}\right)$$

since $\delta/\gamma_0 A \le \varepsilon_0$.

If $x \in Y_\varepsilon$, then the probability that $b_i^x = 1$ is given by

$$\Pr[b_i^x = 1] \ge \left(\frac{1}{2}+\gamma_0\right)^{A_1} \ge \frac{1}{2^{A_1}}(1+2\gamma_0)^{A_1} \ge \frac{1}{\mathrm{e}^A}(1+\gamma_0 A_1) \ge \frac{1}{\mathrm{e}^A}(1+\gamma_0 A).$$

As above, we can argue that the probability that

$$\frac{1}{M}\sum_i b_i^x \le \frac{1}{\mathrm{e}^A}(1+\gamma_0 A)(1-\delta)$$

is at most $\exp(-n)$. Thus, with probability $1-\exp(-n)$,

$$\begin{aligned}\frac{\sum_i b_i^x}{M} &\ge \frac{1}{\mathrm{e}^A}(1+\gamma_0 A)(1-\delta)\\ &\ge \frac{1}{\mathrm{e}^A}(1+\gamma_0 A-2\delta) = \frac{1}{\mathrm{e}^A}\left(1+\gamma_0 A\left(1-\frac{2\delta}{\gamma_0 A}\right)\right)\\ &\ge \frac{1}{\mathrm{e}^A}\left(1+\gamma_0 A\exp\left(-\frac{4\delta}{\gamma_0 A}\right)\right) \ge \frac{1}{\mathrm{e}^A}(1+\gamma_0 A\exp(-\gamma_0 A))\\ &\ge \frac{1}{\mathrm{e}^A}(1+\gamma_1). \end{aligned} \tag{4.7}$$

Thus, by a union bound over x, we can fix a choice of $T_1,\dots,T_M$ so that inequality (4.6) holds for all $x \in N_\varepsilon$ and inequality (4.7) holds for all $x \in Y_\varepsilon$. Hence, condition (4.4) holds for $i=1$ as required. This concludes the construction of C_1, which just outputs the values of $\bigwedge_{j\in T_i} x_j$ for each i.

Construction of C_{i+1}. For the inductive case, we proceed as follows. We assume that i is odd (the case that i is even is similar). So by the inductive hypothesis, we know that condition (4.4) holds and hence that $C_i(x) \in I_0(\gamma_i)$ or $I_1(\gamma_i)$ depending on whether $x \in N_\varepsilon$ or Y_ε. Let $\gamma := \gamma_i$. Let the output gates of C_i be $g_1, \ldots, g_M$.

We choose $T_1, \ldots, T_M \subseteq [M]$ randomly as in the statement of Lemma 4.1 with $s = A$. Note that the chosen parameters satisfy all the hypotheses of Lemma 4.1. Further we also have

$$s\gamma \leq A \cdot A^i \gamma_0 \leq A^{d-2} \cdot \frac{\varepsilon}{\sqrt{n}} \leq \varepsilon_0.$$

The random circuit C' is defined to be the circuit obtained by adding M OR gates to C_i such that the jth OR gate computes $\bigvee_{k \in T_j} g_k$. Let b_j^x be the output of the jth OR gate on $C_i(x)$.

By Lemma 4.1, we have

$$\begin{aligned} x \in N_\varepsilon &\Rightarrow \Pr_S[b_j^x = 0] \geq \exp(-A) \cdot (1 + A\gamma \exp(-A\gamma)), \text{ and} \\ x \in Y_\varepsilon &\Rightarrow \Pr_S[b_j^x = 0] \leq \exp(-A) \cdot (1 - A\gamma \exp(-A\gamma)). \end{aligned} \tag{4.8}$$

Let $\delta = 1/n^3$. Note that

$$A\gamma \in \left[\frac{1}{\sqrt{n}}, \frac{1}{n^{1/2(d-1)}}\right]$$

and hence for large enough n,

$$\frac{2\delta}{A\gamma \exp(-A\gamma)} \leq \varepsilon_0.$$

Assume $x \in N_\varepsilon$. In this case, the Chernoff bound implies that the probability that

$$\sum_{j \in [M]} b_j^x \leq M \exp(-A) \cdot (1 + A\gamma \exp(-A\gamma))(1 - \delta)$$

is at most $\exp(-\Omega(\delta^2 M / \mathrm{e}^A)) \leq \exp(-n)$. When this event does not occur, we have

$$\begin{aligned} \frac{\sum_i b_i^x}{M} &\geq \frac{1}{\mathrm{e}^A}(1 + A\gamma \exp(-A\gamma))(1 - \delta) \\ &\geq \frac{1}{\mathrm{e}^A}(1 + A\gamma \exp(-A\gamma) - 2\delta) = \frac{1}{\mathrm{e}^A}\left(1 + A\gamma \exp(-A\gamma)\left(1 - \frac{2\delta}{A\gamma \exp(-A\gamma)}\right)\right) \\ &\geq \frac{1}{\mathrm{e}^A}\left(1 + A\gamma \exp(-A\gamma) \cdot \exp\left(-\frac{4\delta}{A\gamma \exp(-A\gamma)}\right)\right) \\ &\geq \frac{1}{\mathrm{e}^A}(1 + A\gamma \exp(-A\gamma) \cdot \exp(-A\gamma)) \\ &\geq \frac{1}{\mathrm{e}^A}(1 + A\gamma \exp(-2A\gamma)) \geq \frac{1}{\mathrm{e}^A}(1 + \gamma_{i+1}). \end{aligned} \tag{4.9}$$

We have used above that for large enough n,

$$\frac{2\delta}{A\gamma \exp(-A\gamma)} \leq \varepsilon_0$$

and hence

$$1-\frac{2\delta}{A\gamma\exp(-A\gamma)} \geq \exp\left(\frac{-4\delta}{A\gamma\exp(-A\gamma)}\right).$$

Similarly when $x \in Y_\varepsilon$, the Chernoff bound tells us that the probability that

$$\sum_{j\in[M]} b_j^x \geq M\exp(-A)\cdot(1-A\gamma\exp(-A\gamma))(1+\delta)$$

is at most $\exp(-n)$. In this case, we get

$$\begin{aligned}
\frac{\sum_i b_i^x}{M} &\leq \frac{1}{\mathrm{e}^A}(1-A\gamma\exp(-A\gamma))(1+\delta)\\
&\leq \frac{1}{\mathrm{e}^A}(1-A\gamma\exp(-A\gamma)+\delta) = \frac{1}{\mathrm{e}^A}\left(1-A\gamma\exp(-A\gamma)\left(1-\frac{\delta}{A\gamma\exp(-A\gamma)}\right)\right)\\
&\leq \frac{1}{\mathrm{e}^A}\left(1-A\gamma\exp(-A\gamma)\cdot\exp\left(-\frac{2\delta}{A\gamma\exp(-A\gamma)}\right)\right)\\
&\leq \frac{1}{\mathrm{e}^A}(1-A\gamma\exp(-A\gamma)\cdot\exp(-A\gamma))\\
&= \frac{1}{\mathrm{e}^A}(1-A\gamma\exp(-2A\gamma)) \geq \frac{1}{\mathrm{e}^A}(1-\gamma_{i+1}).
\end{aligned} \tag{4.10}$$

By a union bound, we can fix $T_1,\dots,T_M$ so that inequality (4.9) and inequality (4.10) are true for all $x \in N_\varepsilon$ and $x \in Y_\varepsilon$, respectively. This gives us the circuit C_{i+1} which satisfies all the required properties.

The top two levels of the circuit. At the end of the above procedure we have a circuit C_{d-2} of depth $d-2$ and at most $(d-2)M$ gates that satisfies one of condition (4.4) or condition (4.5) depending on whether $d-2$ is odd or even, respectively. We assume that $d-2$ is even (the other case is similar).

Define $\gamma := \gamma_{d-2}$. Recall from inequality (4.3) that $\gamma \geq A^{d-2}\gamma_0\exp(-3A^{d-2}\gamma_0) \geq A^{d-2}\gamma_0/2$.

Let

$$M' = \left\lceil \exp\left(\frac{10A\log(1/\varepsilon)}{\varepsilon}+10A\right)\right\rceil.$$

We choose M' many subsets $T_1,\dots,T_{M'} \subseteq [M]$ i.i.d. so that each T_j is picked as in Lemma 4.1 with $s = 10A\log(1/\varepsilon)/\varepsilon$. Note that

$$s\gamma \geq s\frac{A^{d-2}\gamma_0}{2} = \frac{10A\log(1/\varepsilon)}{\varepsilon}\cdot\frac{A^{d-2}}{2}\cdot\frac{\varepsilon}{\sqrt{n}} \geq 4\log(1/\varepsilon).$$

where the final inequality uses the fact that $A^{d-1} = (n^{1/2(d-1)}-\alpha)^{d-1} \geq \sqrt{n}\cdot(1-o(1))$ as long as $2d \leq \log n/(\log\log n + c)$ for a large enough absolute constant $c > 0$.

Say $g_1,\dots,g_M$ are the output gates of C_{d-2}. We define the random circuit C' (with n inputs and M' outputs) to be the circuit obtained by adding M' AND gates such that the jth AND gate computes $\bigwedge_{k\in T_j} g_k$. Let b_j^x be the output of the jth AND gate on $C_{d-2}(x)$.

By Lemma 4.1, we have

$$x \in N_\varepsilon \Rightarrow \Pr_S[b_j^x = 1] \le \exp(-s)\cdot\exp(-s\gamma) \le \varepsilon^2\cdot\exp(-s), \text{ and}$$
$$x \in Y_\varepsilon \Rightarrow \Pr_S[b_j^x = 1] \ge \exp(-s)\cdot\exp(s\gamma/2) \ge \frac{\exp(-s)}{\varepsilon^2}. \tag{4.11}$$

Say $x \in N_\varepsilon$. By a Chernoff bound, the probability that $\sum_j b_j^x \ge 2\varepsilon^2 M' \exp(-s)$ is at most

$$\exp(-\Omega(\varepsilon^2 M' \exp(-s))) \le \exp(-\Omega(\varepsilon^2 \mathrm{e}^{10A})) \le \exp(-n).$$

Similarly, when $x \in Y_\varepsilon$, the probability that

$$\sum_j b_j^x \le \frac{M' \exp(-s)}{2\varepsilon^2}$$

is also bounded by $\exp(-n)$. By a union bound, we can fix a $T_1,\dots,T_{M'}$ to get a circuit C_{d-1} such that

$$x \in N_\varepsilon \Rightarrow |C_{d-1}(x)|_1 \le 2\varepsilon^2 \exp(-s) M', \text{ and}$$
$$x \in Y_\varepsilon \Rightarrow |C_{d-1}(x)|_1 \ge \frac{1}{2\varepsilon^2}\exp(-s)M'. \tag{4.12}$$

This gives us the depth-$(d-1)$ circuit C_{d-1}. Note that C_{d-1} has $M' + O(dM) = O(M')$ gates.

To get the depth-d circuit, we choose a random subset $T \subseteq [M']$ by sampling exactly $\lceil \exp(s) \rceil$ many elements of $[M']$ with replacement. We construct a random depth-d circuit C'_d by taking the OR of the output gates of C_{d-1} indexed by the subset T.

From implication (4.12) it follows that

$$x \in N_\varepsilon \Rightarrow \Pr_T[C'_d(x) = 1] \le |T|\cdot 2\varepsilon^2 \exp(-s) \le 4\varepsilon^2 < \varepsilon, \text{ and}$$
$$x \in Y_\varepsilon \Rightarrow \Pr_T[C'_d(x) = 0] \le \left(1 - \frac{\exp(-s)}{2\varepsilon^2}\right)^{\exp(s)} \le \exp(-1/2\varepsilon^2) < \varepsilon.$$

The final inequalities in each case above hold as long as ε is a small enough constant.

It follows from the above that there is a choice for T such that C'_d makes an error—i. e., $C'_d(x) = 1$ for $x \in N_\varepsilon$ or $C'_d(x) = 0$ for $x \in Y_\varepsilon$—on at most a 2ε fraction of inputs from $N_\varepsilon \cup Y_\varepsilon$. We fix such a choice for T and the corresponding circuit C.

We have

$$\Pr_{x\in\{0,1\}^n}[C(x) \ne \mathrm{Maj}_n(x)] \le \Pr_{x\in Y_\varepsilon \cup N_\varepsilon}[C(x) \ne \mathrm{Maj}_n(x)] + \Pr_{x\in\{0,1\}^n}[x \notin Y_\varepsilon \cup N_\varepsilon]$$
$$\le 2\varepsilon + \Pr_{x\in\{0,1\}^n}[x \notin Y_\varepsilon \cup N_\varepsilon].$$

Finally by Stirling's approximation we get

$$\Pr_{x\in\{0,1\}^n}[x \notin Y_\varepsilon \cup N_\varepsilon] = \frac{1}{2^n} \sum_{m \in [\frac{n}{2} - \varepsilon\sqrt{n}, \frac{n}{2} + \varepsilon\sqrt{n}]} \binom{n}{m} \le \frac{1}{2^n} \sum_{m \in [\frac{n}{2} - \varepsilon\sqrt{n}, \frac{n}{2} + \varepsilon\sqrt{n}]} \binom{n}{n/2} \le 2\varepsilon.$$

Hence we see that the circuit C computes a $(4\varepsilon, n)$-Approximate Majority, which proves Theorem 4.2. The circuit has depth d and size $O(M') = \exp(O(n^{1/2(d-1)} \log(1/\varepsilon)/\varepsilon))$. $\square$

5 Conclusions

It is straightforward to extend our main results to $\mathsf{AC}^0[\mathrm{MOD}_p]$ for any fixed prime p to yield a lower bound of $\exp(\Omega(dn^{1/2(d-1)}/p))$ (the $\Omega(\cdot)$ hides an absolute constant.). To prove this, we construct approximating polynomials as in Section 3 for size-s depth-d $\mathsf{AC}^0[\mathrm{MOD}_p]$ formulas of degree $O((1/d)p\log s)^{d-1}$. The construction is almost exactly the same, except that we use $\mathbb{F}_p$-analogs of Lemma 3.6 (which is true with the weaker degree bound of $(p-1)\cdot\lceil\log(1/\varepsilon)\rceil$ [16]) and Lemma 3.7 (which holds as stated over all fields [5]). The Smolensky–Szegedy degree lower bound for approximating the Majority function (Lemma 3.2) is true over all fields.

The improved construction of approximating polynomials for $\mathsf{AC}^0[\mathrm{MOD}_p]$ formulas also implies another result. Using the fact [16] that any $(1/4)$-approximating polynomial over $\mathbb{F}_p$ (p odd) for the Parity function on n variables must have degree $\Omega(\sqrt{n})$, we see that any polynomial-sized $\mathsf{AC}^0[\mathrm{MOD}_p]$ formula computing the Parity function on n variables must have depth $\Omega(\log n)$ for constant primes p. This strengthens a result of Rossman [14] which gives this statement for AC^0 formulas.

Acknowledgements. We thank Rahul Santhanam for valuable discussions, and the organizers of the 2016 Complexity Semester at St. Petersburg, where this collaboration began. We are also grateful to the anonymous referees who reviewed this paper for *Theory of Computing* for their corrections and feedback.

References

[1] KAZUYUKI AMANO: Bounds on the size of small depth circuits for approximating majority. In *Proc. 36th Internat. Colloq. on Automata, Languages and Programming (ICALP'09)*, pp. 59–70. Springer, 2009. [doi:10.1007/978-3-642-02927-1_7] 5, 6, 9

[2] SANJEEV ARORA AND BOAZ BARAK: *Computational Complexity - A Modern Approach*. Cambridge Univ. Press, 2009. [doi:10.1017/CBO9780511804090] 6

[3] RICHARD BEIGEL: The polynomial method in circuit complexity. In *Proc. 8th IEEE Conf. on Structure in Complexity Theory (SCT'93)*, pp. 82–95. IEEE Comp. Soc. Press, 1993. [doi:10.1109/SCT.1993.336538] 7

[4] ERIC BLAIS AND LI-YANG TAN: Approximating Boolean functions with depth-2 circuits. *SIAM J. Comput.*, 44(6):1583–1600, 2015. Preliminary version in CCC'13. [doi:10.1137/14097402X] 6

[5] PRAHLADH HARSHA AND SRIKANTH SRINIVASAN: On polynomial approximations to AC^0. *Random Structures Algorithms*, 54(2):289–303, 2019. Preliminary version in RANDOM'16. [doi:10.1002/rsa.20786, arXiv:1604.08121] 18

[6] JOHAN HÅSTAD: Almost optimal lower bounds for small depth circuits. In *Proc. 18th STOC*, pp. 6–20. ACM Press, 1986. [doi:10.1145/12130.12132] 2, 5

[7] SHLOMO HOORY, AVNER MAGEN, AND TONIANN PITASSI: Monotone circuits for the majority function. In *Proc. Internat. Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (RANDOM'06)*, pp. 410–425. Springer, 2006. [doi:10.1007/11830924_38] 6, 9

[8] STASYS JUKNA: *Boolean Function Complexity: Advances and Frontiers*. Springer, 2012. [doi:10.1007/978-3-642-24508-4] 6

[9] MAURICIO KARCHMER AND AVI WIGDERSON: Monotone circuits for connectivity require super-logarithmic depth. *SIAM J. Discrete Math.*, 3(2):255–265, 1990. Preliminary version in STOC'88. [doi:10.1137/0403021] 2

[10] MARIA M. KLAWE, WOLFGANG J. PAUL, NICHOLAS PIPPENGER, AND MIHALIS YANNAKAKIS: On monotone formulae with restricted depth. In *Proc. 16th STOC*, pp. 480–487. ACM Press, 1984. [doi:10.1145/800057.808717] 5

[11] SWASTIK KOPPARTY AND SRIKANTH SRINIVASAN: Certifying polynomials for $AC^0[\oplus]$ circuits, with applications to lower bounds and circuit compression. *Theory of Computing*, 14(12):1–24, 2018. Preliminary version in FSTTCS'12. [doi:10.4086/toc.2018.v014a012] 4, 7

[12] RYAN O'DONNELL AND KARL WIMMER: Approximation by DNF: Examples and counterexamples. In *Proc. 34th Internat. Colloq. on Automata, Languages and Programming (ICALP'07)*, pp. 195–206. Springer, 2007. [doi:10.1007/978-3-540-73420-8_19] 5, 6, 9

[13] ALEXANDER ALEXANDROVICH RAZBOROV: Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987. Russian original in *Mat. Zametki* 41(4), 1987, 598–607. [doi:10.1007/BF01137685] 2, 4, 7

[14] BENJAMIN ROSSMAN: The average sensitivity of bounded-depth formulas. *Comput. Complexity*, 27(2):209–223, 2018. Preliminary version in FOCS'15. [doi:10.1007/s00037-017-0156-0, arXiv:1508.07677] 2, 18

[15] PETR SAVICKÝ AND ALAN R. WOODS: The number of Boolean functions computed by formulas of a given size. *Random Structures Algorithms*, 13(3–4):349–382, 2000. [doi:10.1002/(SICI)1098-2418(199810/12)13:3/4<349::AID-RSA9>3.0.CO;2-V]

[16] ROMAN SMOLENSKY: Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proc. 19th STOC*, pp. 77–82. ACM Press, 1987. [doi:10.1145/28395.28404] 2, 4, 7, 18

[17] ROMAN SMOLENSKY: On representations by low-degree polynomials. In *Proc. 34th FOCS*, pp. 130–138. IEEE Comp. Soc. Press, 1993. [doi:10.1109/SFCS.1993.366874] 4, 7

[18] MARIO SZEGEDY: *Algebraic Methods in Lower Bounds for Computational Models with Limited Communication*. Ph. D. thesis, University of Chicago, 1989. Available at advisor's home page. 4, 7

[19] LESLIE G. VALIANT: Short monotone formulae for the majority function. *J. Algorithms*, 5(3):363–366, 1984. [doi:10.1016/0196-6774(84)90016-6] 5, 9

AUTHORS

Benjamin Rossman
Assistant professor
Department of Mathematics and
Department of Computer Science
University of Toronto
Toronto, ON, Canada
ben.Rossman@utoronto.ca
http://math.toronto.edu/~Rossman

Srikanth Srinivasan
Assistant professor
Department of Mathematics
IIT Bombay
Powai, Mumbai, India
srikanth@math.iitb.ac.in
http://www.math.iitb.ac.in/~srikanth/

ABOUT THE AUTHORS

BENJAMIN ROSSMAN found his way to complexity theory after studying mathematical logic as an undergraduate at the University of Pennsylvania and subsequently as an intern with Yuri Gurevich at Microsoft Research. He received his Ph. D. from the Massachusetts Institute of Technology in 2010, under the supervision of Madhu Sudan and under the distraction of the ballroom dance team.

SRIKANTH SRINIVASAN got his undergraduate degree from the Indian Institute of Technology Madras, where his interest in the theory side of CS was piqued under the tutelage of N. S. Narayanswamy. Subsequently, he obtained his Ph. D. from The Institute of Mathematical Sciences in 2011, where his advisor was V. Arvind. His research interests span all of TCS (in theory), but in practice are limited to circuit complexity, derandomization, and related areas of mathematics. He looks back fondly at the days when he enjoyed running and pretending to play badminton.