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

Benjamin Rossman$^1$ and Srikanth Srinivasan$^2$

1 Departments of Mathematics and Computer Science, University of Toronto, Toronto, Canada
rossman@utoronto.ca
2 Department of Mathematics, IIT Bombay, Bombay, India
srikanth@math.iitb.ac.in

Abstract

This paper gives the first separation between the power of formulas and circuits of equal depth in the $\text{AC}^0[\oplus]$ basis (unbounded fan-in AND, OR, NOT and MOD gates). We show, for all $d(n) \leq O(\log \log n)$, that there exist polynomial-size depth-$d$ circuits that are not equivalent to depth-$d$ formulas of size $n^{o(d)}$ (moreover, this is optimal in that $n^{o(d)}$ cannot be improved to $n^{O(d)}$). This result is obtained by a combination of new lower and upper bounds for Approximate Majorities, the class of Boolean functions $\{0, 1\}^n \rightarrow \{0, 1\}$ that agree with the Majority function on $\frac{3}{4}$ fraction of inputs.

$\text{AC}^0[\oplus]$ formula lower bound. We show that every depth-$d$ $\text{AC}^0[\oplus]$ formula of size $s$ has a $1/8$-error polynomial approximation over $\mathbb{F}_2$ of degree $O(\log s)^{d-1}$. This strengthens a classic $O(\log s)^{d-1}$ degree approximation for circuits due to Razborov [12]. Since the Majority function has approximate degree $\Theta(\sqrt{n})$, this result implies an $\exp(\Omega(dn^{1/2(d-1)}))$ lower bound on the depth-$d$ $\text{AC}^0[\oplus]$ formula size of all Approximate Majority functions for all $d(n) \leq O(\log n)$.

Monotone $\text{AC}^0$ circuit upper bound. For all $d(n) \leq O(\log n)$, we give a randomized construction of depth-$d$ monotone $\text{AC}^0$ circuits (without NOT or MOD$_2$ gates) of size $\exp(\Theta(dn^{1/2(d-1)}))$ that compute an Approximate Majority function. This strengthens a construction of formulas of size $\exp(\Theta(dn^{1/2(d-1)}))$ due to Amano [1].

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases circuit complexity, lower bounds, approximate majority, polynomial method

Digital Object Identifier 10.4230/LIPIcs.ICALP.2017.50

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 $\text{NC}^1$ (the class of languages decidable by polynomial-size Boolean formulas) is a proper subclass of $\text{P/poly}$ (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 $\text{NC}^1$ vs. $\text{P/poly}$ 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

$^1$ In this paper we focus on non-uniform complexity classes. The question of uniform-$\text{NC}^1$ vs. $\text{P}$ is wide open as well.
by the classic lower bound of Karchmer and Wigderson [8] (on the monotone formula size of st-Connectivity).

The bounded-depth setting is another natural venue for investigating the question of formula 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 $AC^0[⊕]$ 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 sub-logarithmic depth regime ($d(n) \ll \log n$), given the powerful lower bound techniques against $AC^0$ circuits (Håstad’s Switching Lemma [5]) and $AC^0[⊕]$ circuits (the Polynomial Method of Razborov [12] and Smolensky [14]). 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 [13] developed a new way of applying Håstad’s Switching Lemma to $AC^0$ formulas, in order to prove an $\exp(\Omega(d(n^{1/(d-1)})))$ lower bound on the formula size of the Parity function for all $d \leq O(\log n)$. 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 $d(n) \leq O(\log n^{\log n \log \log n})$, as well as a super-polynomial separation for all $\omega(1) \leq d(n) \leq o(\log n)$.

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

**Theorem 1.**

(i) For all $2 \leq d(n) \leq O(\log n / \log \log n)$, there exist $AC^0[⊕]$ circuits (in fact, monotone $AC^0$ circuits) of depth $d$ and size poly$(n)$ that are not equivalent to any $AC^0[⊕]$ formulas of depth $d$ and size $n^{o(d)}$.

(ii) For all $\omega(1) \leq d(n) \leq o(\log n)$, the class of languages decidable by polynomial-size depth-$d$ $AC^0[⊕]$ formulas is a proper subclass of the class of languages decidable by polynomial-size depth-$d$ $AC^0[⊕]$ circuits.

Separation (i) is asymptotically optimal, in view of the aforementioned simulation of poly$(n)$-size depth-$d$ circuits by depth-$d$ formulas of size $n^{O(d)}$. Separation (ii) resembles an analogue of $NC^1 \neq P/poly$ (or rather $NC^1 \neq AC^1$) within the class $AC^0[⊕]$. In fact, extending separation (ii) from depth $o(\log n)$ to depth $\log n$ is equivalent the separation of $NC^1$ and $AC^1$. 

Formulas and Circuits
1.1 Proof outline

Improved polynomial approximation

The lower bound for $AC^0[⊕]$ formulas follows the general template due to Razborov [12] on proving lower bounds for $AC^0[⊕]$ 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 $AC^0[⊕]$ circuit of size $s$ and depth $d$, there is a randomized polynomial $P$ of degree $O(\log s)^d$ that computes $f$ correctly on each input with probability $\frac{1}{2}$ (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 function for $q$ odd) on $n$ variables does not have such an approximation of degree less than $\Omega(\sqrt{n})$ [12, 14, 15], we get that any $AC^0[⊕]$ 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 $AC^0[⊕]$ formulas. More precisely, for $AC^0[⊕]$ formulas of size $s$ and depth $d$, we are able to construct 1/8-error probabilistic polynomials of degree $O(\frac{1}{2} \log s)^{d-1}$. (Since every depth-$d$ circuit of size $s$ is equivalent to a depth-$d$ formula of size at most $s^{2^{d-1}}$, this result implies Razborov’s original theorem that $AC^0[⊕]$ 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 $P_i$ of degree $O(\log t)^{d-2}$ and by a standard error-reduction [10], 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 $P_i(x) \neq F_i(x)$ 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 $P_i$, 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 [12, 14, 15], we can already obtain stronger lower bounds for $AC^c[⊕]$ formulas than are known for $AC^0[⊕]$ circuits. For instance, it follows that any $AC^0[⊕]$ formula of depth $d$ computing the Majority function on $n$ variables must have size $\exp(\Omega(n^{1/2(d-1)}))$ 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 $AC^0[⊕]$ 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 $AC^c[⊕]$ circuits of depth $d$ and size $\exp(O(n^{1/2(d-1)}))$ (where the constant in the $O(·)$ is independent of $d$). However, as far as we know, the strongest
result in this direction [9] only yields $\text{AC}^0[\oplus]$ 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 Majority function on most inputs. 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 $\text{AC}^0[\oplus]$ 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 [11] and Amano [1] that there exist Approximate Majorities that can be computed by monotone $\text{AC}^0$ formulas of depth $d$ and size $\exp(O(dn^{1/2(d-1)}))$. (Note that the double exponent $\frac{1}{2(d-1)}$ is now the same in the upper and lower bounds.)

We use the above ideas for our separation between $\text{AC}^0[\oplus]$ formulas and circuits. Plugging our stronger polynomial approximation for $\text{AC}^0[\oplus]$ formulas, we obtain that any $\text{AC}^0[\oplus]$ 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 $\text{AC}^0[\oplus]$ 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 $\text{AC}^0$ circuits of depth $d$ of size $\exp(O(n^{1/2(d-1)}))$ (the constant in the $O(\cdot)$ is a constant independent of $d$).

### Smaller circuits for Approximate Majority

Our construction closely follows Amano’s, which in turn is related to Valiant’s probabilistic construction [17] 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 Pitassi [6] 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 $i \leq d$ as follows. The formula $F_1$ is the AND of $a_1$ independent and randomly chosen variables. For even (respectively odd) $i > 1$, $F_i$ is the OR (respectively 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 $i \leq d$ 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 [6]) we can show that this computes an Approximate Majority with high probability, which finishes the proof.\(^3\)

---

\(^2\) Indeed, this is inevitable with all constructions that we are aware of, since they are actually $\text{AC}^0$ circuits and it is known by a result of Håstad [5] that any $\text{AC}^0$ circuit of depth $d$ for the Majority function must have size $\exp(\Omega(n^{1/(d-1)}))$.

\(^3\) This is a slightly imprecise description of the construction as the final two levels of the circuit are actually defined somewhat differently.
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 $|y|_0$ and $|y|_1$ to denote the number of 0s and number of 1s respectively in $y$.

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

Definition 2. An $(\varepsilon, n)$-Approximate Majority is a function $f : \{0, 1\}^n \to \{0, 1\}$ such that

$$\Pr_{x \in \{0, 1\}^n}[f(x) \neq \text{MAJ}_n(x)] \leq \varepsilon.$$ 

As far as we know, the study of this class of functions was initiated by O’Donnell and Wimmer [11]. See also [1, 4].

We refer the reader to [2, 7] for standard definitions of Boolean circuits and formulas. We use $\text{AC}^0$ circuits (respectively formulas) to denote circuits (respectively formulas) of constant depth made up of AND, OR and NOT gates. Similarly, $\text{AC}^0[2]$ circuits (respectively formulas) will be circuits (respectively formulas) of constant depth made up of AND, OR, MOD2 and NOT gates.

The size of a circuit will denote the number of gates in the circuit and the size of a formula will denote the number of its leaves which is within a constant multiplicative factor of the number of gates in the formula.\(^4\)

3 Lower Bound

In this section, we show that any $\text{AC}^0[2]$ formulas of depth $d$ computing a $(1/4, n)$-Approximate Majority must have size at least $\exp(O(dn^{1/2(d - 1)}))$ 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. Fix any $\varepsilon \in [0, 1]$. A polynomial $P \in \mathbb{F}_2[X_1, \ldots, X_n]$ is said to be an $\varepsilon$-approximating polynomial for a Boolean function $f : \{0, 1\}^n \to \{0, 1\}$ if

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

We will use the following result of Smolensky [15] (see also Szegedy’s PhD thesis [16]).

Lemma 4 (Smolensky [15]). Let $\varepsilon \in (0, \frac{1}{2})$ be any fixed constant. Any $(\frac{1}{4} - \varepsilon)$-approximating polynomial for the Majority function on $n$ variables must have degree $\Omega(\sqrt{n})$.

Corollary 5. Let $f$ be any $(1/4, n)$-Approximate Majority and $\varepsilon \in (0, 1/4)$ an arbitrary constant. Then any $(\frac{1}{4} - \varepsilon)$-approximating polynomial for $f$ must have degree $\Omega(\sqrt{n})$.

Proof. The proof is immediate from Lemma 4 and the triangle inequality. \(\blacktriangleleft\)

Definition 6. An $\varepsilon$-error probabilistic polynomial of degree $D$ for a Boolean function $f : \{0, 1\}^n \to \{0, 1\}$ is a random variable $P$ taking values from polynomials in $\mathbb{F}_2[X_1, \ldots, X_n]$ of degree at most $D$ such that for all $x \in \{0, 1\}^n$, we have $\Pr[f(x) = P(x)] \geq 1 - \varepsilon$.

\(^4\) We assume here without loss of generality that the formula does not contain a gate of fan-in 1 feeding into another.
Definition 7. Let $D_\varepsilon(f)$ be the minimum degree of an $\varepsilon$-error probabilistic polynomial for $f$.

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

Lemma 8 (Razborov [12]). Let $\text{OR}_n$ and $\text{AND}_n$ be the OR and AND functions on $n$ variables respectively. Then $D_\varepsilon(\text{OR}_n), D_\varepsilon(\text{AND}_n) \leq \lceil \log(1/\varepsilon) \rceil$.

Lemma 9 (Kopparty and Srinivasan [10] (implicit in proof of Lemma 10)). There is an absolute constant $c_2$ such that for any $\varepsilon \in (0,1)$, $D_\varepsilon(f) \leq c_2(1/\varepsilon) \cdot D_{1/8}(f)$ for all Boolean functions $f$.

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

Theorem 10. There is an absolute constant $c_2$ such that, if $f$ is computed by an $\text{AC}^0[\oplus]$ formula $F$ of size $s$ and depth $d+1$, then $D_{1/8}(f) \leq 3(c_2(1/\varepsilon) \log(s+1))^{d+1}$.

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 $D_{1/8}(f) \leq 3$. In the case that the formula is an AND or OR gate, this follows from Lemma 8. 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 $d \geq 1$. We assume that the formula $F$ is the AND/OR/ MOD$_2$ of sub-formulas $F_1, \ldots, F_m$ 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 $D_{1/8}(f_i) \leq 3(c_2(1/\varepsilon) \log(s_i+1))^{d+1}$ for all $i$.

We must show that $D_{1/8}(f) \leq 3(c_2(1/\varepsilon) \log(s+1))^{d+1}$.

By Lemma 9, each $f_i$ has an $s_i/(16s)$-error probabilistic polynomial $P_i$ of degree $c_1 \cdot [\log(16s/s_i)] \cdot D_{1/8}(f_i)$, which is at most

$$3c_1 \cdot 5(\log(s/s_i) + 1) \cdot (c_2(\frac{1}{\varepsilon} \log(s_i + 1)))^d.$$ Then $(P_1, \ldots, P_m)$ 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 $Q$ of degree 4 for the output gate of the formula.

Then $Q(P_1, \ldots, P_m)$ is a $1/8$-error probabilistic polynomial for $f$ of degree

$$60c_1 \cdot \max_i (\log(s/s_i) + 1) \cdot (c_2(\frac{1}{\varepsilon} \log(s_i + 1)))^d.$$ So long as $c_2 \geq 20c_1$, it suffices to show that for all $i$,

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

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

$$(b+1) \left(\frac{a}{d} + 1\right)^d \leq (a+b) \left(\frac{a}{d+1} + 1\right)^{d+1}.$$ 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 $p_{a,d}(a/d) = 0.$

\hfill \Box
\textbf{Corollary 11.} Fix any constant \(d\) and let \(n \in \mathbb{N}\) be a growing parameter. Let \(f\) be any \((1/4,n)\)-Approximate Majority. Then any AC^0[⊕] formula of depth \(d\) computing \(f\) must have size \(\exp(\Omega(dn^{1/2(d-1)})\) for all \(d \leq O(\log n)\), where asymptotic notation \(O(\cdot)\) and \(\Omega(\cdot)\) hide absolute constants (independent of \(d\) and \(n\)).

\textbf{Proof.} Say that \(F\) is an AC^0[⊕] formula of depth \(d\) and size \(s\) computing \(f\). Then, by Lemma 9, we see that \(F\) has a 1/8-error probabilistic polynomial \(P\) of degree \(D \leq O(1/2 \log s + 1)^{d-1}\). 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 5 implies that the degree of \(P\) must be \(\Omega(\sqrt{n})\). Hence, we obtain \(O(1/2 \log s + 1)^{d-1} \geq \Omega(\sqrt{n})\). It follows that

\[
\frac{1}{s} \geq \exp(\Omega(dn^{1/2(d-1)}) - O(d)).
\]

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 \(d \leq \varepsilon \log n\).

\section{Upper Bound}

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

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

\[
\begin{align*}
\exp(-\beta) &\leq 1 - \beta \exp(-\beta), \\
1 - \beta &\geq \exp(-\beta - \beta^2) \geq \exp(-2\beta).
\end{align*}
\]

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

We need the following technical lemma.

\textbf{Lemma 12.} Let \(A, s, n, \gamma\) be positive reals, \(M, n \in \mathbb{N}\), and \(\gamma \in (\frac{1}{4}, \frac{1}{10})\) be such that \(e^A \geq n^3\), \(n \geq \frac{1}{M}\), and \(s \leq n\). Define \(I_0(\gamma) := \{y \in \{0,1\}^M \mid |y|_1 \leq Me^{-A}(1-\gamma)\}\) and \(I_1(\gamma) := \{y \in \{0,1\}^M \mid |y|_1 \geq Me^{-A}(1+\gamma)\}\). If we choose \(S \subseteq [M]\) of size \(t := |e^A \cdot s|\) by picking \(t\) random elements from \(M\) with replacement, then

\[
\begin{align*}
x \in I_0(\gamma) \Rightarrow \Pr_S[\bigvee_{j \in S} x_j = 0] &\geq \exp(-s) \cdot \exp(s\gamma/2), \\
x \in I_1(\gamma) \Rightarrow \Pr_S[\bigvee_{j \in S} x_j = 0] &\leq \exp(-s) \cdot \exp(-s\gamma).
\end{align*}
\]

Further, if \(s\gamma \leq \varepsilon_0\), then the above probabilities can be lower bounded and upper bounded 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 \(J_0(\gamma) := \{y \in \{0,1\}^M \mid |y|_0 \leq Me^{-A}(1-\gamma)\}\) and \(J_1(\gamma) := \{y \in \{0,1\}^M \mid |y|_0 \geq Me^{-A}(1+\gamma)\}\), with the event \(\bigvee_{j \in S} x_j = 0\) being replaced by the event \(\bigwedge_{j \in S} x_j = 1\).

\textbf{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.

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

\[
\leq \exp(-(1 + \gamma) \cdot s) \leq \exp(-s) \cdot \exp(-s\gamma). \tag{1}
\]
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 $x \in I_1$.

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

$$
\Pr_{S}[\bigvee_{j \in S} x_j = 0] \geq \left(1 - \frac{1 - \gamma}{e^A}\right)^{e^A, s+1}
$$

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

$$
= \exp\left(-s + s \gamma - \frac{s}{e^A} - \frac{1 - \gamma}{e^A} - \frac{1}{e^{2A}}\right)
$$

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

$$
= \exp(-s) \cdot \exp(s \gamma (1 - \frac{2e^{-A}}{\gamma}))
$$

(2)

where for the second inequality we have used the fact that $e^{-A} \leq \frac{1}{n^d} \leq \varepsilon_0$, we have $1 - \frac{1 - \gamma}{e^A} \geq \exp(-\frac{1 - \gamma}{e^A} - \frac{1}{e^{2A}})$. Since $e^{-A} \leq \frac{1}{n^d} \leq \frac{1}{4\varepsilon_0} \leq \gamma/4$, we can lower bound the right hand side of (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 RHS of (2) can also be lower bounded 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 13.** For any growing parameter $n \in \mathbb{N}$ and $2 \leq d \leq O(\log n)$ and $\varepsilon > 0$, there is an $(\varepsilon, n)$-Approximate Majority $f_n$ computable by a monotone $AC^0$ circuit with at most $\exp(O(n^{1/2(2d-1)} \log(1/\varepsilon)/\varepsilon))$ many gates, where both $O(\cdot)$’s hide absolute constants (independent of $d$, $\varepsilon$).

**Proof.** We assume throughout that $\varepsilon$ 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 $\exp(O(n^{1/2(2d-1)} \log(1/\varepsilon)/\varepsilon))$ computing a $(4\varepsilon, n)$-Approximate Majority, which also implies the theorem.

Fix parameters $A = n^{1/2(2d-1)}$ and $M = [e^{10A}]$. We assume that $A \geq 10 \log n$ (which holds as long as $d \leq \frac{e \log n}{\log \log n}$ for an absolute constant $c > 0$) and that $\varepsilon \leq \varepsilon_0$.

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 $\gamma_i \leq A^i \gamma_0$ for each $i \in [d-2]$. As a result we also obtain

$$
\gamma_i = A^i \gamma_0 \exp(-2A (\gamma_0 + \gamma_1 + \cdots + \gamma_{i-1}))
$$

$$
\geq A^i \gamma_0 \exp(-2 \gamma_0 A (1 + A + A^2 + \cdots + A^{i-1})) \geq A^i \gamma_0 \exp(-3A^i \gamma_0).
$$

(3)
Let
\[
Y_{\varepsilon} = \left\{ x \in \{0,1\}^n \mid |x|_1 \geq \left( \frac{1}{2} + \frac{\varepsilon}{\sqrt{n}} \right) n \right\},
\]
\[
N_{\varepsilon} = \left\{ x \in \{0,1\}^n \mid |x|_1 \leq \left( \frac{1}{2} - \frac{\varepsilon}{\sqrt{n}} \right) n \right\}.
\]

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 \(i\)
\[
x \in N_{\varepsilon} \Rightarrow C_i(x) \in I_0(\gamma_i)
\]
\[
x \in Y_{\varepsilon} \Rightarrow C_i(x) \in I_1(\gamma_i)
\]
and similarly for even \(i\)
\[
x \in N_{\varepsilon} \Rightarrow C_i(x) \in J_0(\gamma_i)
\]
\[
x \in Y_{\varepsilon} \Rightarrow C_i(x) \in J_1(\gamma_i).
\]

After this is done, we will add on 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 respectively - 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 \(A\) 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
\[
Pr[b_i^x = 1] \leq \left( \frac{1}{2} - \gamma_0 \right)^A \leq \frac{1}{2^A} (1 - 2\gamma_0)^A \leq \frac{1}{e^A} (1 - \gamma_0 A)
\]
where the last inequality follows from the fact that \((1 - z)^A \leq (1 - z + A^2 z^2)\).

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

By a Chernoff bound, the probability that \(\frac{1}{M} \sum_i b_i^x \geq \frac{1}{M} (1 - \gamma_0 A)(1 + \delta)\) is bounded by \(\exp(-\Omega(\delta^2 M/e^A)) \leq \exp(-\Omega(e^{3A}/n^6)) \leq \exp(-n)\), since \(e^A \geq n^1\). Thus, with probability at least \(1 - \exp(-n)\), we have
\[
\frac{1}{M} \sum_i b_i^x \leq \frac{1}{e^A} (1 - \gamma_0 A)(1 + \delta)
\]
\[
\leq \frac{1}{e^A} (1 - \gamma_0 A + \delta) = \frac{1}{e^A} (1 - \gamma_0 A(1 - \frac{\delta}{\gamma_0 A}))
\]
\[
\leq \frac{1}{e^A} (1 - \gamma_0 A \exp(-28/\gamma_0 A)) \leq \frac{1}{e^A} (1 - \gamma_0 A \exp(-\gamma_0 A))
\]
\[
\leq \frac{1}{e^A} (1 - \gamma_1).
\]

Above, we have used the fact that \((1 - \frac{\delta}{\gamma_0 A}) \geq \exp(-28/\gamma_0 A)\) since \(\delta/\gamma_0 A \leq \varepsilon_0\) for large enough \(n\), as noted above.
If $x \in Y_z$, then the probability that $b^x_i = 1$ is given by

$$\Pr[b^x_i = 1] \geq \left(\frac{1}{2} + \gamma_0\right)^A \geq \frac{1}{e^A}(1 + \gamma_0) \geq \frac{1}{A}(1 + \gamma_0).$$

As above, we can argue that the probability that $\frac{1}{M} \sum_i b^x_i \leq \frac{1}{e^A}(1 + \gamma_0)(1 - \delta)$ is at most $\exp(-n)$. Thus, with probability $1 - \exp(-n)$

$$\sum_i b^x_i \geq \frac{1}{e^A}(1 + \gamma_0)(1 - \delta) \geq \frac{1}{e^A}(1 + \gamma_0 - 2\delta) = \frac{1}{e^A}(1 + \gamma_0(1 - \frac{2\delta}{\gamma_0})).$$

Thus, by a union bound over $x$, we can fix a choice of $T_1, \ldots, T_M$ so that (6) holds for all $x \in N_z$ and (7) holds for all $x \in Y_z$. Hence, (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 (4) holds and hence that $C_i(x) \in I_0(\gamma_i)$ or $I_1(\gamma_i)$ depending on whether $x \in N_z$ or $Y_z$. 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 12 with $s = A$. Note that the chosen parameters satisfy all the hypotheses of Lemma 12. Further we also have $s\gamma \leq A \cdot A^d \gamma_0 \leq A^{d-1} \cdot \frac{\gamma_i}{\sqrt{s}} \leq \varepsilon_0$.

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

By Lemma 12, we have

$$x \in N_z \Rightarrow \Pr_S[b^x_j = 0] \geq \exp(-A) \cdot (1 + A\gamma \exp(-A\gamma))$$

$$x \in Y_z \Rightarrow \Pr_S[b^x_j = 0] \leq \exp(-A) \cdot (1 - A\gamma \exp(-A\gamma))$$

Let $\delta = \frac{1}{\sqrt{A}}$. Note that $A\gamma \in \left[\frac{1}{\sqrt{n}}, \frac{1}{n^{1/2}d^{2/3}}\right]$ and hence for large enough $n$, $\frac{2\delta}{A\gamma \exp(-A\gamma)} \leq \varepsilon_0$.

Assume $x \in N_z$. In this case, the Chernoff bound implies that the probability that $\sum_{j \in [M]} b^x_j \leq M \exp(-A) \cdot (1 + A\gamma \exp(-A\gamma))(1 - \delta)$ is at most $\exp(-\Omega(\delta^2 M/e^A)) \leq \exp(-n)$.
When this event does not occur, we have
\[
\sum_i \frac{b_i^2}{M} \geq \frac{1}{e^A} (1 + A \gamma \exp(-A \gamma))(1 - \delta)
\]
\[
\geq \frac{1}{e^A} (1 + A \gamma \exp(-A \gamma) - 2\delta) = \frac{1}{e^A} (1 + A \gamma \exp(-A \gamma)(1 - \frac{2\delta}{A \gamma \exp(-A \gamma)}))
\]
\[
\geq \frac{1}{e^A} (1 + A \gamma \exp(-A \gamma) \cdot \exp(-\frac{4\delta}{A \gamma \exp(-A \gamma)}))
\]
\[
\geq \frac{1}{e^A} (1 + A \gamma \exp(-A \gamma) \cdot \exp(-A \gamma))
\]
\[
\geq \frac{1}{e^A} (1 + A \gamma \exp(-2A \gamma)) \geq \frac{1}{e^A} (1 + \gamma_{i+1}). \tag{9}
\]

We have used above that for large enough \(n\), \(\sum_{j \in [M]} b_j^2 \geq M \exp(-A) \cdot (1 - A \gamma \exp(-A \gamma))(1 + \delta)\) is at most \(\exp(-n)\). In this case, we get
\[
\sum_i \frac{b_i^2}{M} \leq \frac{1}{e^A} (1 - A \gamma \exp(-A \gamma))(1 + \delta)
\]
\[
\leq \frac{1}{e^A} (1 - A \gamma \exp(-A \gamma) + \delta) = \frac{1}{e^A} (1 - A \gamma \exp(-A \gamma)(1 - \frac{\delta}{A \gamma \exp(-A \gamma)}))
\]
\[
\leq \frac{1}{e^A} (1 - A \gamma \exp(-A \gamma) \cdot \exp(-\frac{2\delta}{A \gamma \exp(-A \gamma)}))
\]
\[
\leq \frac{1}{e^A} (1 - A \gamma \exp(-A \gamma) \cdot \exp(-A \gamma))
\]
\[
= \frac{1}{e^A} (1 - A \gamma \exp(-2A \gamma)) \geq \frac{1}{e^A} (1 - \gamma_{i+1}). \tag{10}
\]

By a union bound, we can fix \(T_1, \ldots, T_M\) so that (9) and (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 (4) or (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 (3) that \(\gamma \geq A^{d-2-\gamma_0} \exp(-3A^{d-2-\gamma_0}) \geq A^{d-2-\gamma_0}/2\).

Let \(M' = \lceil \exp(\frac{10A\log(1/\varepsilon)}{\varepsilon} + 10A) \rceil\). We choose \(M'\) many subsets \(T_1, \ldots, T_{M'} \subseteq [M]\) i.i.d. so that each \(T_j\) is picked as in Lemma 12 with \(s = 10A \log(1/\varepsilon)/\varepsilon\). Note that
\[
s\gamma \geq \frac{s A^{d-2-\gamma_0}}{2} = \frac{10A \log(1/\varepsilon)}{\varepsilon} \cdot \frac{A^{d-2}}{2} \cdot \frac{\varepsilon}{\sqrt{M}} \geq 5 \log(1/\varepsilon).
\]

Say \(g_1, \ldots, g_M\) 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 \(j\)th AND gate computes \(\bigwedge_{k \in T_j} g_k\). Let \(b_j^2\) be the output of the \(j\)th AND gate on \(C_{d-2}(x)\).

By Lemma 12, we have
\[
x \in N_\varepsilon \Rightarrow Pr_s[b_j^2 = 1] \leq \exp(-s) \cdot \exp(-s\gamma) \leq \varepsilon^2 \cdot \exp(-s)
\]
\[
x \in Y_\varepsilon \Rightarrow Pr_s[b_j^2 = 1] \geq \exp(-s) \cdot \exp(s\gamma/2) \geq \frac{\exp(-s)}{\varepsilon^2}. \tag{11}
\]
Say $x \in N_\varepsilon$. By a Chernoff bound, the probability that $\sum_j b_j^x \geq 2\varepsilon^2 M' \exp(-s)$ is at most $\exp(-\Omega(\sqrt{dM'} \exp(-s))) \leq \exp(-\Omega(\varepsilon^2 e^{10A})) \leq \exp(-n)$. Similarly, when $x \in Y_\varepsilon$, the probability that $\sum_j b_j^x \leq \frac{M' \exp(-s)}{2}$ is also bounded by $\exp(-n)$. By a union bound, we can fix a $T_1, \ldots, T_{M'}$ to get a circuit $C_{d-1}$ such that

$$x \in N_\varepsilon \Rightarrow |C_{d-1}(x)| \leq 2\varepsilon^2 \exp(-s) M'$$

$$x \in Y_\varepsilon \Rightarrow |C_{d-1}(x)| \geq \frac{1}{2\varepsilon^2} \exp(-s) M'. \quad (12)$$

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 the output gates of $C_{d-1}$ indexed by the subset $T$.

From (12) it follows that

$$x \in N_\varepsilon \Rightarrow \Pr_T[C_d'(x) = 1] \leq |T| \cdot 2\varepsilon^2 \exp(-s) \leq 4\varepsilon^2 < \varepsilon$$

$$x \in Y_\varepsilon \Rightarrow \Pr_T[C_d'(x) = 0] \leq \left(1 - \frac{\exp(-s)}{2\varepsilon^2}\right)^{\exp(s)} \leq \exp(-1/2\varepsilon^2) < \varepsilon.$$

The final inequalities in each case above hold as long as $\varepsilon$ 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 $2\varepsilon$ fraction of inputs from $N_\varepsilon \cup Y_\varepsilon$. We fix such a choice for $T$ and the corresponding circuit $C$.

We have

$$\Pr_{x \in \{0,1\}^n} [C(x) \neq \text{Maj}_n(x)] \leq \Pr_{x \in Y_\varepsilon \cup N_\varepsilon} [C(x) \neq \text{Maj}_n(x)] + \Pr_{x \in \{0,1\}^n} [x \notin Y_\varepsilon \cup N_\varepsilon]$$

$$\leq 2\varepsilon + \Pr_{x \in \{0,1\}^n} [x \notin Y_\varepsilon \cup N_\varepsilon].$$

Finally by Stirling’s approximation we get

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

Hence we see that the circuit $C$ computes a $(4\varepsilon, n)$-Approximate Majority, which proves Theorem 13.

The circuit has depth $d$ and size $O(M') = \exp(O(n^{1/2(1-d)} \log(1/\varepsilon)/\varepsilon))$.

## 5 Conclusion

Our main results extend straightforwardly to $\text{AC}^0[\text{MOD}_p]$ for any fixed prime $p$. The proofs are exactly the same except for the fact that the approximating polynomials of degree $O(\frac{1}{d} \log s)^{d-1}$ from Section 3 are constructed over $\mathbb{F}_p$.

Using the fact [14] 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 $\text{AC}^0[\text{MOD}_p]$ formula computing the Parity function on $n$ variables must have depth $\Omega(\log n)$. This strengthens a result of Rossman [13] which gives this statement for $\text{AC}^0$ formulas.

## Acknowledgements

We thank Rahul Santhanam for valuable discussions. We also thank the organizers of the 2016 Complexity Semester at St. Petersburg, where this collaboration began.
References


