Lower Bounds for Circuits with Few Modular and Symmetric Gates

Arkadev Chattopadhyay[1] and Kristoffer Arnsfelt Hansen[2]

[1] School of Computer Science, McGill University
achatt3@cs.mcgill.ca
[2] Department of Computer Science, University of Aarhus, Denmark
arnsfelt@daimi.au.dk

Abstract. We consider constant depth circuits augmented with few modular, or more generally, arbitrary symmetric gates. We prove that circuits augmented with $o(\log^2 n)$ symmetric gates must have size $n^{\Omega(\log n)}$ to compute a certain (complicated) function in $\mathbf{ACC^0}$.
This function is also hard on the average for circuits of size $n^{\epsilon \log n}$ augmented with $o(\log n)$ symmetric gates, and as a consequence we can get a pseudorandom generator for circuits of size m containing $o(\sqrt{\log m})$ symmetric gates.
For a composite integer m having r distinct prime factors, we prove that circuits augmented with s MOD_m gates must have size $n^{\Omega(\frac{1}{s}\log^{\frac{1}{r-1}} n)}$ to compute MAJORITY or MOD_l, if l has a prime factor not dividing m. For proving the latter result we introduce a new notion of representation of boolean function by polynomials, for which we obtain degree lower bounds that are of independent interest.

1 Introduction

Strong size lower bounds have been obtained for several classes of circuits. In particular *constant depth* circuits ($\mathbf{AC^0}$ circuits) require exponential size to compute even simple functions such as PARITY [1–4]. More generally, if we also allow gates computing MOD_q for a prime power q we have exponential lower bounds for computing MAJORITY, and if l has a prime divisor not dividing q also for computing MOD_l [5, 6]. If we however allow gates computing MOD_m for an arbitrary integer m, we have no nontrivial lower bounds. The corresponding circuit class $\mathbf{ACC^0}$ is the smallest natural circuit class, for which we have no nontrivial lower bounds. Another very interesting class of constant depth circuits (containing $\mathbf{ACC^0}$) is the class of constant depth threshold circuits ($\mathbf{TC^0}$ circuits), i.e constant depth circuits built entirely from MAJORITY gates.

Viewing $\mathbf{TC^0}$ circuits as $\mathbf{AC^0}$ circuits augmented with MAJORITY gates instead, raises a natural question: Can lower bounds be proved if we limit the amount of MAJORITY gates used. This question was answered affirmatively in a series of papers. Aspnes et al [7] proved that $\mathbf{AC^0}$ circuits with a MAJORITY gate on top, i.e $\mathbf{MAJ} \circ \mathbf{AC^0}$ circuits, and in fact even $\mathbf{AC^0}$ circuits with a single MAJORITY gate anywhere in the circuit, require exponential size to compute the MOD_2 function. Beigel et al [8] improved this lower bound to allow for $o(\log n)$ MAJORITY gates, and Beigel [9] showed that the lower bound also holds for $\mathbf{AC^0}$ circuits augmented with $n^{o(1)}$ MAJORITY gates. Finally Barrington and Straubing [10] generalized this lower bound, showing that it also holds for circuits computing the MOD_m function for any constant m. In fact, the proofs of these lower bounds show that $\mathbf{AC^0}$ circuits of size $2^{n^{o(1)}}$ augmented with $n^{o(1)}$ MAJORITY gates must differ from the MOD_m function on a constant fraction all inputs of length n.

The analogous question for $\mathbf{ACC^0}$ is: Can lower bounds be proved if we limit the amount of MOD_m gates used. Note that although these functions *can* be computed by small $\mathbf{TC^0}$ circuits, the above lower bounds for $\mathbf{AC^0}$ circuits augmented with few MAJORITY gates are not strong enough to be applied. Indeed, these lower bounds are proved for circuit *computing* the MOD_m functions.

Another way of viewing $\mathbf{TC^0}$, is as $\mathbf{AC^0}$ augmented with gates computing arbitrary symmetric functions, since every such function is computable by a depth 2 threshold circuit and MAJORITY is obviously symmetric. Thus, regarding $\mathbf{TC^0}$, we could also ask, if lower bounds can be proved if we limit the amount of symmetric functions used. Also note that since the MOD_m functions are symmetric, an affirmative answer would answer both questions.

Our most powerful lower bound is a superpolynomial lower bound for circuits containing $o(\log^2 n)$ arbitrary symmetric gates. Combining a technique of Beigel [9] for reducing the number of symmetric gates with a lower bound for $\mathbf{MAJ} \circ \mathbf{SYM} \circ \mathbf{AC^0}$ circuits by Hansen and Miltersen [11], one can immediately obtain superpolynomial lower bounds for $\mathbf{AC^0}$ circuits augmented with up to $o(\log\log n)$ MOD_m gates. Our significantly stronger result, is obtained by instead generalizing the results of Håstad and Goldmann [12], Razborov and Wigderson [13] and Hansen and Miltersen [11]. We now give the definition of the functions we prove the lower bounds for. The so-called Generalized Inner Product [14] is defined as $\mathrm{GIP}_{n,k} = (\mathrm{MOD}_2)_n \circ \mathrm{AND}_k$, i.e. the MOD_2 function of n conjunctions each consisting of k variables. Let further $f_{n,k} = (\mathrm{MOD}_2)_n \circ \mathrm{AND}_k \circ (\mathrm{MOD}_2)_n$, i.e. $\mathrm{GIP}_{n,k}$ with each input replaced by the MOD_2 function of n variables.

Theorem 1. *Any $\mathbf{AC^0}$ circuit augmented with $o(\log^2 n)$ arbitrary symmetric gates computing $f_{n,\log n}$ must have size $n^{\Omega(\log n)}$.*

It follows as a corollary of the following theorem.

Theorem 2. *Any $\mathbf{MAJ} \circ \mathbf{ANY}_s \circ \mathbf{SYM} \circ \mathbf{ANY}_t$ circuit computing $GIP_{n,t+1}$ must have size $2^{\Omega\left(\frac{n}{st2^t}\right)}$. For constants $\delta > 0$ and $\gamma > 0$ such that $\delta + \gamma < 1$, if $t = \delta \log n$ and $s = n^\gamma$, any $\mathbf{MAJ} \circ \mathbf{ANY}_s \circ \mathbf{SYM} \circ \mathbf{AC^0}$ circuit computing $f_{n,t+1}$ must have size $n^{\Omega(\log n)}$.*

For circuits augmented with MOD_m gates we are able to give lower bounds for computing much simpler functions than that used for proving Theorem 1. In particular are all these functions symmetric, and the results are thus incomparable.

Theorem 3. *Let m be a positive integer with $r \geq 2$ distinct prime factors. Any* $\mathbf{AC^0}$ *circuit augmented with s MOD_m gates require size $n^{\Omega(\frac{1}{s} log^{\frac{1}{r-1}} n)}$ to compute* MAJORITY *or* MOD_l*, if l has a prime factor not dividing m.*

This complements the mentioned results about circuits with few MAJORITY gates: We cannot compute MOD_m with few MAJORITY gates and we cannot compute MAJORITY with few MOD_m gates. Note also that the lower bounds are for the same functions and under the same conditions as the mentioned lower bounds for circuits with MOD_q gates for prime powers q.

Previously, Hansen and Miltersen [11] used results on *weak representation* of boolean function by polynomials over $\mathbf{Z}_m$ as introduced and proved by Green [15] to obtain exponential lower bounds for circuits augmented with just one MOD_m gate. Theorem 3 is proved, by greatly generalizing this technique. We introduce a new notion of representation of boolean functions be polynomials over $\mathbf{Z}_m$ for which we prove lower bounds and finally apply these to obtain circuit lower bounds.

Beigel and Maciel [16] showed how to convert a $\mathbf{MAJ} \circ \mathbf{OR} \circ \mathbf{MOD}_m$ circuit into a $\mathbf{MAJ} \circ \mathbf{OR}_{O(1)} \circ \mathbf{MOD}_m$ circuit with only a polynomial increase in size, which they could then apply to results by Krause and Pudlák [17] to obtain lower bounds for $\mathbf{MAJ} \circ \mathbf{OR} \circ \mathbf{MOD}_m$ circuits. Applying their technique to our results we immediately obtain the following as a corollary to Theorem 2.

Theorem 4. *Any* $\mathbf{MAJ} \circ \mathbf{OR} \circ \mathbf{MOD}_m \circ \mathbf{ANY}_t$ *circuit computing $GIP_{n,t+1}$ must have size $2^{\Omega\left(\frac{n}{t2^t}\right)}$. Any* $\mathbf{MAJ} \circ \mathbf{OR} \circ \mathbf{MOD}_m \circ \mathbf{AC^0}$ *circuit computing $f_{n,\log n}$ must have size $n^{\Omega(\log n)}$.*

Viola [18] showed that by combining and modifying the proofs of [13] and [11] one can obtain a function that is hard on the average for $\mathbf{SYM} \circ \mathbf{AC^0}$ circuits of size $n^{\epsilon \log n}$. Viola used his result for constructing a pseudorandom generator for these circuits, using the construction of Nisan and Wigderson [19], and generalized it to circuits with a *constant* number of symmetric gates using the technique of Beigel [9]. Adding our constructions, one can prove that the same function is hard on the average for $\mathbf{ANY}_{n^\gamma} \circ \mathbf{SYM} \circ \mathbf{AC^0}$ circuits of size $n^{\epsilon \log n}$ for $\gamma < 1$, and thus also for $\mathbf{AC^0}$ circuits of size $n^{\epsilon \log n}$ with $\gamma \log n$ symmetric gates. This gives a pseudorandom generator for circuits of size m with $o(\sqrt{\log m})$ symmetric gates. Subsequently to this result, Viola showed how to improve this to $o(\log m)$ symmetric gates [18].

1.1 Organisation of Paper

In Sect. 2 we introduce the notation and previous results we will use in our proofs. In Sect. 3 we prove Theorem 1 and 2. In Sect. 4 we prove Theorem 3 as well as our degree lower bounds used for this.

2 Preliminaries

2.1 Constant Depth Circuits

We consider circuits built from families of unbounded fanin gates. Inputs are allowed to be boolean variables and their negations as well as the constants 0 and 1. Let $x_1, \ldots, x_n$ be boolean inputs. For a positive integer m, the MOD_m function is 1 if and only if $\sum_{i=1}^n x_i \not\equiv 0 \pmod m$. Generally, for $A \subset \mathbf{Z}_m$ the MOD_m^A function is 1 is and only $\sum_{i=1}^n x_i \in A \pmod m$. The MAJORITY function is 1 if and only if $\sum_{i=1}^n x_i \geq \frac{n}{2}$. A function is symmetric if and only if it only depends on the sum of its inputs.

Let **AND** and **OR** denote the families of unbounded fanin AND and OR gates. Let $\mathbf{MOD}_m$, **MAJ**, **SYM** and **ANY** denote the families of MOD_m, MAJORITY, all symmetric gates and *any* kind of gate, respectively. If G is a family of boolean gates and $\mathcal{C}$ is a family of circuits we let $G \circ \mathcal{C}$ denote the class of circuits consisting of a gate from G taking circuits from $\mathcal{C}$ as inputs. If we need to specify a specific bound on the fanin of some of the gates, this will be specified by a subscript.

$\mathbf{AC^0}$ is the class of functions computed by constant depth circuits built from AND and OR gates. $\mathbf{ACC^0}$ is the the analogous class of functions computed when we also allow unbounded fanin MOD_m gates for constants m, and similarly is $\mathbf{TC^0}$ the class of functions computed when we instead allow unbounded fanin MAJORITY gates.

2.2 Multiparty Communication Complexity

We consider the "Number on the Forehead" model of multiparty communication as introduced by Chandra, Furst and Lipton [20]. Here k players wish to evaluate a boolean function $f(x_1, \ldots, x_k)$ where $x_i \in \{0,1\}^n$ and player i knows all input except x_i. They exchange messages according to a fixed protocol by broadcasting, and we are interested in the number of bits that must be exchanged in order to evaluate f. We let $C_\epsilon(f)$ denote the minimal number of bits that must be exchanged in the worst case, for a deterministic protocol that computes f correctly with probability at least $1-\epsilon$ when x is chosen uniformly at random.

Babai, Nisan and Szegedy [14] studied the k-party communication complexity of the Generalized Inner Product, and proved $C_{\frac{1}{2}-\epsilon}(\mathrm{GIP}_{n,k}) = \Omega(\frac{n}{4^k} + \log \epsilon)$. We state an improvement due to Chung and Tetali [21].

Theorem 5. $C_{\frac{1}{2}-\epsilon}(GIP_{n,k}) = \Omega(\frac{n}{2^k} + \log \epsilon)$.

Håstad and Goldmann [12] observed that depth 2 threshold circuit can be evaluated efficiently by a multiparty protocol if the fanin of the bottom gates is restricted. We observe that essentially the same protocol works for depth 3 circuits if we restrict the fanin of both the top and bottom gates.

Proposition 6. *Let f be a boolean function computed by a $\mathbf{ANY}_s \circ \mathbf{SYM} \circ \mathbf{ANY}_t$ circuit of size S. Then there is a $t+1$ player protocol computing f with $1 + st \log S$ bits exchanged.*

Proof. Since the bottom fanin of the circuit is less than the number of players, every $\mathbf{ANY}_t$ gate can be evaluated by at least one player. In the protocol we fix a partition of these gates to the players, such that every player can evaluate all assigned gates. Now for each of the symmetric gates, the first t players compute the sum of the assigned inputs and send the results separately to player $t+1$. Player $t+1$ can then compute the output of the circuit and communicate the result with 1 extra bit of communication. $\square$

2.3 The Discriminator Lemma and The Switching Lemma

Let C be a circuit taking n inputs and f a boolean function on n variables. We say that C is an ϵ-discriminator for f if $\Pr[C(x) = 1 | f(x) = 1] - \Pr[C(x) = 1 | f(x) = 0] \geq \epsilon$. The so-called Discriminator Lemma by Hajnal et al [22], states that if a circuit with a MAJORITY gate at the output computes a boolean function f, then one of the inputs to the output gate is an ϵ-discriminator for f.

Lemma 7. *Let f be a boolean function computed by a circuit C with a* MAJORITY *gate as the output gate, and let $C_1, \ldots, C_s$ be the subcircuits of C whose output gates are the inputs to the output of C. Then for some i, C_i is an $\frac{1}{s}$-discriminator for f.*

A *restriction* on a set V of boolean variables is a map $\rho : V \to \{0, 1, \star\}$. It acts on a boolean function $f : V \to \{0, 1\}$, creating a new boolean function f_ρ on the set of variables for which $\rho(x) = \star$, obtained by substituting $\rho(x)$ for $x \in V$ whenever $\rho(x) \neq \star$. The variables x for which $\rho(x) = \star$ are called *free*; the other variables *set*. Let R_n^l denote the set of all restriction ρ leaving l of n variables free.

A decision tree is a binary tree, where the internal nodes are labeled by variables and leafs are labeled by either 0 or 1. On a given input x, its value is the value of the leaf reached by starting at the root, and at any internal node labeled by x_i proceeding to the left child if $x_i = 0$ and to the right child otherwise. We will use the following version of Håstads Switching Lemma due to Beame [23].

Lemma 8. *Let f be a DNF formula in n variables with terms of length at most r. Let $l = pn$ and pick ρ uniformly at random from R_n^l. Then the probability that f_ρ does not have a decision tree of depth at most d is less than $(7pr)^d$.*

The advantage of using Beame's switching lemma is that it directly gives us a decision tree. If we convert a decision tree into a DNF, we have that all terms are *mutually contradictory*, i.e. we can view it as a sum of terms , instead as an OR of AND's. This will allow us to absorb the sum into a symmetric gate.

2.4 Representation by Polynomials

For composite m there are several different ways of defining representation by polynomials. To obtain our lower bounds we will define new a notion which we will call *weak generalized* representation. Let f be a boolean function in n variables, and let P be a polynomial in n variables over $\mathbf{Z}_m$.

– P is a *strong* representation of f if $f(x) = P(x)$ for all $x \in \{0,1\}^n$.
– P is a *one-sided* representation of f if $f(x) = 0 \Leftrightarrow\ P(x) \equiv 0 \pmod m$ for all $x \in \{0,1\}^n$.
– P is a *weak* representation of f if $P \not\equiv 0$ and $P(x) \not\equiv 0 \pmod m \Rightarrow f(x) = 1$ for all $x \in \{0,1\}^n$.
– P is a *generalized*[3] representation of f if there is a set $S \subset \mathbf{Z}_m$ such that $f(x) = 1 \Leftrightarrow\ P(x) \in S$.
– P is a *weak generalized* representation of f if there is a set $S \subset \mathbf{Z}_m$ and an $\bar{x} \in \{0,1\}^n$ such that $P(\bar{x}) \in S$ and that for all $x \in \{0,1\}^n$ we have $P(x) \in S \Rightarrow f(x) = 1$.

The minimal degree of a polynomial satisfying the above properties will be called the strong, one-sided, weak, generalized and weak generalized MOD_m degree, respectively. Note that a strong representation is also a one-sided representation. A one-sided representation is also a weak representation as well as a generalized representation, and these are both weak generalized representations. For a weak generalized representation we can assume that $|S| = 1$, in fact, if P is a weak generalized representation there exist $a \in \mathbf{Z}_m$ such that $P - a$ is a weak generalized representation with respect to $\{0\}$ of the same boolean function.

For convenience we will also consider a representation of a boolean function by more than one polynomial. Let f be as before and let $P_1, \ldots, P_s$ be polynomials in n variables over $\mathbf{Z}_m$. We say $P_1, \ldots, P_s$ is a *simultaneous weak* representation of f if there is an $\bar{x} \in \{0,1\}^n$ such that $P_i(\bar{x}) \not\equiv 0 \pmod m$ for all i, and if it holds that whenever $P_i(x) \not\equiv 0 \pmod m$ for all i, we have that $f(x) = 1$. The s-simultaneous weak MOD_m degree of f is the minimum over all choices of polynomials $P_1, \ldots, P_s$ over $\mathbf{Z}_m$ representing f of the maximal degree of $P_1, \ldots, P_s$.

We will several times use the following well known lemma. (cf. [24]).

Lemma 9. *Let $q = p^k$ for a prime p, let P be a polynomial of degree d in n variables over $\mathbf{Z}_q$ and let $S \subseteq Z_q$. Then there exists another polynomial P' of degree at most $(q-1)d$ in n variables over $\mathbf{Z}_p$ such that $P(x) \in S \Rightarrow P'(x) = 1$ and $P(x) \notin S \Rightarrow P'(x) = 0$ for all $x \in \{0,1\}$.*

Thus if $q = p^k$ for a prime p, the strong MOD_p degree of f is at most $(q-1)$ times the generalized MOD_q degree of f, and similarly the weak MOD_p degree of f is at most $(q-1)$ times the weak generalized MOD_q degree of f.

The following lemma shows, that s-simultaneous weak degree and weak generalized degree are essentially the same, when s is a constant.

Lemma 10. *Let m be a positive integer and let $m = q_1 \cdots q_t$ be the factorization in prime powers $q_i = p_i^{k_i}$, let $m' = p_1 \cdots p_t$ and let f be a boolean function. The weak generalized $\mathrm{MOD}_{m'}$ degree of f is at most $s(q-1)$ times the s-simultaneous weak MOD_m degree of f, where q is the largest prime power factor of m.*

[3] This notion was actually called *weak* representation in [24], but we prefer to reserve this name for the representation introduced by Green [15], which is analogous to the weak degree of a voting polynomial defined by Aspnes et al [7].

On the other hand, the $(m-1)$-simultaneous weak MOD_m *degree of f is at most as large as the weak generalized* MOD_m *degree of f.*

3 Circuits with few Symmetric Gates

Proof (of Theorem 2). Let C be a $\mathbf{MAJ} \circ \mathbf{ANY}_s \circ \mathbf{SYM} \circ \mathbf{ANY}_t$ circuit of size S which computes $f = \mathrm{GIP}_{n,t+1}$. From Lemma 7 we have a $\mathbf{ANY}_s \circ \mathbf{SYM} \circ \mathbf{ANY}_t$ subcircuit C_i that is an $\frac{1}{S}$-discriminator for f, i.e. $\Pr[C_i(x) = 1 \mid f(x) = 1] - \Pr[C_i(x) = 1 \mid f(x) = 0] \geq \frac{1}{S}$. From the Binomial Theorem we have $\Pr[f(x) = 0] - Pr[f(x) = 1] = (1 - 2^{-t})^n$ and thus $\Pr[f(x) = 0] = \frac{1}{2} + \frac{(1-2^{-t})^n}{2}$. For $a, b \in \{0, 1\}$ let $P_{ab} = Pr[C_i(x) = a \mid f(x) = b]$. We thus get $\Pr[C_i(x) = f(x)] = P_{11} \Pr[f(x) = 1] + P_{00} \Pr[f(x) = 0] = \frac{1}{2} + \frac{1}{2}(P_{11} - P_{10}) + \frac{(1-2^{-t})^n}{2}(1 - P_{11} - P_{10}) \geq \frac{1}{2} + \frac{1}{2S} - \frac{(1-2^{-t})^n}{2} \geq \frac{1}{2} + \frac{1}{4S}$, where we assume without loss of generality that $S \leq \frac{1}{2} e^{\frac{n}{2^t}}$. Combining this with Theorem 5 and Proposition 6 we have $1 + st \log S = \Omega(\frac{n}{2^t} + \log \frac{1}{4S})$ and we can conclude that $S = 2^{\Omega\left(\frac{n}{st2^t}\right)}$.

For the second part, let C be a depth $d+3$ $\mathbf{MAJ} \circ \mathbf{ANY}_s \circ \mathbf{SYM} \circ \mathbf{AC^0}$ circuit of size $n^{\epsilon \log n}$ computing $f_{n,t+1}$. Let $m = n^2(t+1)$. We choose a random restriction $\rho \in R_m^{m^{\frac{2}{3}}}$, and argue that for sufficiently small ϵ, with positive probability, $f_{n,t+1}$ contains $\mathrm{GIP}_{n,t+1}$ as a subfunction and the function computed by C_ρ is also computed by a $\mathbf{MAJ} \circ \mathbf{ANY}_s \circ \mathbf{SYM} \circ \mathbf{AND}_t$ circuit of size $n^{\epsilon \log n} 2^t = n^{\delta + \epsilon \log n}$. The statement then follows from the first part.

The probability that ρ fails the first requirement is at most $n(t+1)$ times the probability that a random subset of size $m^{\frac{2}{3}}$ does not intersect a fixed subset of size n (corresponding to the inputs of the MOD_2 functions substituted in $\mathrm{GIP}_{n,t+1}$). This happens with probability at most $2^{-n^{\Omega(1)}}$.

For the other part, view ρ as a composition of several restrictions $\rho_1, \ldots, \rho_d$ where $\rho_i \in R_{m_{i-1}}^{m_i}$ and $m_i = m\left(m^{\frac{1}{3d}}\right)^{-i}$. Assume that after having applied the first $i-1$ restrictions, that all functions computed at level $i-1$ of C are computed by decision trees of depth at most t, and hence DNFs with terms of size at most t. Assume without loss of generality that the gates at level i are OR gates, and thus also computable by decision trees of depth at most t. By Lemma 8, the probability that the function computed by such an OR gate can not be computed by a decision tree of depth t after applying ρ_i is at most $\left(7\frac{m_i}{m_{i-1}} t\right)^t = \left(7m^{-\frac{1}{3d}} \delta \log n\right)^{\delta \log n} = n^{-\Omega(\log n)}$. Thus for ϵ sufficiently small, ρ will convert all gates at level d into decision trees of depth at most t. By rewriting these into DNFs with *mutually contradictory* terms, allows us to simply directly feed these terms into the symmetric gate above, resulting in a $\mathbf{MAJ} \circ \mathbf{ANY}_s \circ \mathbf{SYM} \circ \mathbf{AND}_t$ circuit computing the same function. □

Proof (of Theorem 1). Let C be a $\mathbf{AC^0}$ circuit of size S augmented with s symmetric gates. Let $g_1, \ldots, g_s$ denote the symmetric gates, such that there is no path from the output of g_j to an input of g_i if $i < j$. For $\alpha \in \{0,1\}^s$, let C_i^α

be the $\mathbf{SYM} \circ \mathbf{AC^0}$ subcircuit of C with g_i as output, where each g_j for $j < i$ is replaced by the constant α_j. Similarly, let C^α be the $\mathbf{AC^0}$ circuit obtained from C where every g_i is replaced by α_i. Note that since the subcircuit of C with output g_i contain at most $i-1$ other symmetric gates than g_i, we have at most $\sum_{i=1}^{s} 2^{i-1} = 2^s - 1$ different $\mathbf{SYM} \circ \mathbf{AC^0}$ subcircuits of C. We can now compute the same function as C by a $\mathbf{OR}_{2^s} \circ \mathbf{AND}_{s+1} \circ \mathbf{SYM} \circ \mathbf{AC^0}$ circuit of size $O(2^s S)$ constructed as follows: The output OR gate is fed by ANDs corresponding to all $\alpha \in \{0,1\}^s$. The AND gates takes C^α as input, as well as C_i^α if $\alpha_i = 1$ and $\neg C_i^\alpha$ (which is also a $\mathbf{SYM} \circ \mathbf{AC^0}$ circuit) otherwise. If $s = o(\log^2 n)$ Theorem 2 gives that $S = n^{\Omega(\log n)}$. $\square$

4 Circuits with few Modular Gates

4.1 Degree Lower Bounds

Barrington and Tardos obtained the following lower bound on the generalized degree of the OR function.

Theorem 11 ([24]). *Let m be a positive integer with $r \geq 2$ distinct prime factors, and let q be the smallest maximal prime power of m. The generalized degree of the* OR *function on n variables is at least $\left(\left(\frac{1}{q-1} - o(1)\right) \log n\right)^{\frac{1}{r-1}}$.*

As a corollary to this we obtain the same lower bounds for the weak generalized degree of the MAJORITY and ¬MAJORITY functions.

Theorem 12. *Let m be a positive integer with $r \geq 2$ distinct prime factors, and let q be the smallest maximal prime power of m. The weak generalized* MOD_m *degree of the* MAJORITY *and* ¬MAJORITY *functions on n variables is at least $\left(\left(\frac{1}{q-1} - o(1)\right) \log n\right)^{\frac{1}{r-1}}$.*

We will use the following lower bound on the degree of weak representation due to Green [15] in a crucial way for proving Lemma 15.

Theorem 13. *Let m and l be positive relative prime integers. The weak* MOD_m *degree of the* MOD_l *and* $\neg\mathrm{MOD}_l$ *functions on n variables is at least $\lfloor \frac{n}{2(l-1)} \rfloor$.*

Corollary 14. *Let l be a positive integer and let $q = p^a$ for a prime p not dividing l. The weak generalized* MOD_q *degree of the* MOD_l *and* $\neg\mathrm{MOD}_l$ *functions on n variables is at least $\frac{1}{(q-1)} \lfloor \frac{n}{2(l-1)} \rfloor$.*

For a subset $S \subseteq \{1, \ldots, n\}$, let $\chi(S) \in \{0,1\}^n$ denote its characteristic vector. Conversely for $x \in \{0,1\}^n$, let $\sigma(x) \subseteq \{1, \ldots, n\}$ be the set of indices where $x_i = 1$.

Lemma 15. *Let P be a polynomial of degree d in n variables over $\mathbf{Z}_q$ where $q = p^a$ for a prime p, and let l be a positive integer not divisible by p. If $k \geq 1$*

satisfies $n \geq 2(l-1)\left(k+(q-1)\sum_{i=1}^{d}(d+1-i)\binom{k}{i}\right)$, *then there exists pairwise disjoint nonempty sets* $S^1,\ldots,S^k \subseteq \{1,\ldots,n\}$ *such that for every* $y \in \{0,1\}^k$ *we have* $P(\sum_{i=1}^{k} y_i\chi(S^i)) \equiv P(0) \pmod{q}$ *and furthermore we have* $|S^i| \not\equiv 0 \pmod{l}$ *for all* i.

Proof. Assume without loss of generality that $P(0)=0$. We will find the sets S^i inductively with $|S^i| \leq s_i$, where $s_j = 2(l-1)\left(1+(q-1)\sum_{i=1}^{d}\binom{j-1}{i-1}(d+1-i)\right)$. First pick a set S of $s_1 = 2(l-1)(d(q-1)+1)$ variables. If we consider the polynomial obtained from P by substituting 0 for all variables not in S, we obtain a polynomial which can not be a weak generalized representation of $\neg\mathrm{MOD}_l$ with respect to the set $\{0\}$, by Corollary 14. Thus there is a subset $S^1 \subseteq S$ such that $P(\chi(S^1)) = 0 = P(0)$ and $\neg\mathrm{MOD}_l(\chi(S^1)) = 0$, that is $|S^1| \not\equiv 0 \pmod{l}$.

In the general case, assume for $j < k$, that we have already found sets $S^1,\ldots,S^j$, where $|S^i| \leq s_i$ and $|S^i| \not\equiv 0 \pmod{l}$ for all $i \leq j$ and we have $P(\sum_{i=1}^{t} y_i\chi(S^i)) = 0 = P(0)$ for all $y \in \{0,1\}^j$. Pick a set S of size s_{j+1} of the remaining variables.

For any $y \in \{0,1\}^j$, let P_y be the polynomial obtained from P be substituting y_i for all variables in S^i for all i, and further substituting 0 for all remaining variables not in S. Let P'_y be the polynomial over $\mathbf{Z}_p$, obtained using Lemma 9, that is a strong representation of the boolean function of which P_y is a generalized representation with respect to $\{0\}$. That is $P'_y(x) \equiv 0 \pmod{p} \Leftrightarrow P_y(x) \not\equiv 0 \pmod{q}$ and $P'_y(x) \equiv 1 \pmod{p} \Leftrightarrow P_y(x) \equiv 0 \pmod{q}$.

Let $R = \prod_{y\in\{0,1\}^j} P'_y$. Observe that R only take the values 0 and 1 modulo p, and that $R(x) \equiv 1 \pmod{p}$ if and only if $P'_y(x) \equiv 1 \pmod{p}$ for all y, that is, if and only if $P_y(x) \equiv 0 \pmod{q}$ for all y. If R is not a weak representation of $\neg\mathrm{MOD}_l$, we could thus find a new set S^{j+1} as before. But since the degree of R is $2^j(q-1)d$, this does not give our desired bound on the number of variables s_{j+1} required for this contradiction. However, exactly as in [24] we can use inclusion-exclusion to construct an equivalent polynomial R', i.e $R'(x) \not\equiv 0$ if and only if $P_y(x) \equiv 0 \pmod{q}$ for all y, of degree $(q-1)\sum_{i=0}^{d-1}\binom{j}{i}(d-i)$. From Theorem 13 and the choice of s_{j+1} we have that R' is not a weak representation of $\neg\mathrm{MOD}_l$. We can thus find $S^{j+1} \subseteq S$ such that $R'(\chi(S^{j+1})) \not\equiv 0 \pmod{p}$ and $\neg\mathrm{MOD}_l(\chi(S^{j+1})) = 0$. It follows that $P_y(\chi(S^{j+1})) = 0$ for all y and $|S^{j+1}| \not\equiv 0 \pmod{l}$. To allow the induction to go through, we need that $n \geq \sum_{j=1}^{k} s_j$, which is exactly the requirement stated. $\square$

Theorem 16. *Let* m *be a positive integer with* $r \geq 2$ *distinct prime factors, let* q *be the smallest maximal prime power factor of* m *and let* p *be a prime not dividing* m. *For all* $a \in \mathbf{Z}_p$ *the weak generalized* MOD_m *degree of the* $\mathrm{MOD}_p^{\{a\}}$ *and* $\neg\mathrm{MOD}_p^{\{a\}}$ *functions on* n *variables is at least* $\left(\left(\frac{1}{2(p-1)^2(q-1)} - o(1)\right)\log n\right)^{\frac{1}{r-1}}$.

Proof. The idea of the proof is to succesively use Lemma 15 to convert a given representation into another representation (on fewer variables) that depend on less prime factors, and finally use Corollary 14.

Let $n = n(m,d)$ denote the maximal number of variables, for which there is a weak generalized representation over $\mathbf{Z}_m$ of degree d, for any of the $\mathrm{MOD}_p^{\{a\}}$ and $\neg\mathrm{MOD}_p^{\{a\}}$ functions. We need to prove that $\log n(m,d) \leq (2(p-1)^2(q-1)+o(1))d^{r-1}$. Let $m = q_1 m_1$ where q_1 is a maximal prime power divisor of m different from q.

Assume that P is a polynomial in n variables of degree d over $\mathbf{Z}_m$ which is a weak generalized representation of f with respect to $\{0\}$, where f is either $\mathrm{MOD}_p^{\{a\}}$ or $\neg\mathrm{MOD}_p^{\{a\}}$ for some $a \in \mathbf{Z}_p$. By assumption there exists $\bar{x} \in \{0,1\}^n$ such that $P(\bar{x}) \equiv 0 \pmod m$ and $f(\bar{x}) = 1$. If $|\sigma(\bar{x})| < \frac{n}{2}$ let P' be the polynomial obtained from P by setting the variables indexed by $\sigma(\bar{x})$ to 1. Otherwise, if $|\sigma(\bar{x})| \geq \frac{n}{2}$ we can let P' be the polynomial where variables x_i are substituted with $1 - x_i$ if $i \in \sigma(\bar{x})$ and otherwise set to 0, and modify the following proof accordingly. In either case the number n' of variables in P' is at least $\frac{n}{2}$.

For a given integer k, let $k' = (p-1)k$ and assume that n' is at least $2(p-1)\left(k' + (q_1-1)\sum_{i=1}^{d}(d+1-i)\binom{k'}{i}\right)$. Then using Lemma 15 we can find pairwise disjoint nonempty sets $S'^1, \ldots, S'^{k'} \subseteq \{1,\ldots,n'\}$ such that for every $y \in \{0,1\}^{k'}$ we have $P'(\Sigma_{i=1}^{k'} y_i \chi(S'^i)) \equiv P'(0) \equiv 0 \pmod{q_1}$ and furthermore we have $|S'^i| \not\equiv 0 \pmod p$ for all i. Choosing the most occuring residue $b \in \mathbf{Z}_p \setminus \{0\}$ among $|S'^i|$ modulo p and extending the sets to $\{1,\ldots,n\}$, we have pairwise disjoint nonempty sets $S^1,\ldots,S^k \subseteq \{1,\ldots,n\}$ such that $P(\bar{x} + \Sigma_{i=1}^{k} y_i\chi(S^i)) \equiv P(\bar{x}) \equiv 0 \pmod{q_1}$ for every $y \in \{0,1\}^k$, and $|S^i| \equiv b$ for all i.

In case f is $\mathrm{MOD}_p^{\{a\}}$, we have $P(\bar{x} + \Sigma_{i=1}^{k} y_i\chi(S^i)) \equiv 0 \pmod m \Rightarrow |\sigma(\bar{x})| + \sum_{i=1}^{k} y_i|S^i| \not\equiv a \pmod p \Rightarrow \sum_{i=1}^{k} y_i \not\equiv b^{-1}(a - |\sigma(\bar{x})|) \pmod p$. In case f is $\neg\mathrm{MOD}_p^{\{a\}}$ we have $P(\bar{x} + \Sigma_{i=1}^{k} y_i\chi(S^i)) \equiv 0 \pmod m \Rightarrow |\sigma(\bar{x})| + \sum_{i=1}^{k} y_i|S^i| \equiv a \pmod p \Rightarrow \sum_{i=1}^{k} y_i \equiv 0 \pmod p$. Since $P(\bar{x} + \Sigma_{i=1}^{k} y_i\chi(S^i)) \equiv 0 \pmod{m_1} \Rightarrow P(\bar{x} + \Sigma_{i=1}^{k} y_i\chi(S^i)) \equiv 0 \pmod m$, we thus have that $P(\bar{x} + \Sigma_{i=1}^{k} y_i\chi(S^i))$ is a weak generalized MOD_{m_1} representation of either $\mathrm{MOD}_p^{\{b^{-1}(a-|\sigma(\bar{x})|)\}}$ or $\neg\mathrm{MOD}_p^{\{0\}}$. Thus for $k = n(m_1,d)+1$ we must have $n' < 2(p-1)\left(k' + (q_1-1)\sum_{i=1}^{d}(d+1-i)\binom{k'}{i}\right)$.

If $r = 2$ then $m_1 = q$ and from Corollary 14 we have that $k \leq 2(p-1)\left((q-1)d+1\right)$. Using $n(m,d) \leq O\left(d2^{k'}\right)$ gives $\log n(m,d) \leq O(\log d) + (p-1)k \leq \left(2(p-1)^2(q-1)+o(1)\right)d$.

If $r > 2$, we have by induction that $\log k \leq (2(p-1)^2(q-1)+o(1))d^{r-2}$. Using $n(m,d) \leq O(k'^d)$ gives $\log n(m,d) \leq O(1) + d(\log(p-1) + \log k) \leq (2(p-1)^2(q-1)+o(1))d^{r-1}$. □

Corollary 17. *Let m be a positive integer with $r \geq 2$ distinct prime factors, let q be the smallest maximal prime power factor of m and let l be a positive integer having a prime factor p not dividing m.*

Then the weak generalized MOD_m degree of the $\mathrm{MOD}_l^{\{a\}}$ and $\neg\mathrm{MOD}_l^{\{a\}}$ functions on n variables is at least $\left(\left(\frac{1}{2(p-1)^2(q-1)} - o(1)\right)\log n\right)^{\frac{1}{r-1}}$.

4.2 Circuit Lower Bounds

Proof (of Theorem 3). Let C be a depth d $\mathbf{AC^0}$ circuit of size $n^{\frac{\epsilon}{s}\log^{\frac{1}{r-1}} n}$ containing s MOD_m gates $g_1, \ldots, g_s$ computing a function f. Assume there is no path from the output of g_j to an input of g_i if $i < j$. For $\alpha \in \{0,1\}^s$ let C_i^α be the $\mathbf{MOD}_m \circ \mathbf{AC^0}$ subcircuit of C with g_i as output, where every g_j for $j < i$ is replaced by the constant α_j. Similarly, let C^α be the $\mathbf{AC^0}$ circuit obtained from C where every g_i is replaced by α_i. We will choose a random restriction $\rho \in R_n^{\sqrt{n}}$ and show that for every $\delta > 0$ we can choose ϵ sufficiently small such that with high probability we can obtain polynomials p_i^α and q^α, of degree $\frac{\delta}{s}\log^{\frac{1}{r-1}} n$, such that $C_{i,\rho}^\alpha(x) = 1 \Leftrightarrow p_i^\alpha(x) \not\equiv 0 \pmod{m}$ and $C_\rho^\alpha(x) = q^\alpha(x)$ for all x.

If we obtain this we can construct a simultaneous weak representation, using $s+1$ of the polynomials, of either f_ρ or $\neg f_\rho$ as follows: Pick a *maximal* set G of the MOD_m gates that are 1 at the same time for some input x in C_ρ and define α such that $\alpha_i = 1 \Leftrightarrow g_i \in G$. If there exist x such that all the gates in G evaluate to 1 on x and at the same time $C_\rho(x) = 1$, then $\{p_i^\alpha \mid g_i \in G\} \cup \{q^\alpha\}$ is a simultaneous weak representation of f_ρ. Otherwise $\{p_i^\alpha \mid g_i \in G\}$ is a simultaneous weak representation of $\neg f_\rho$.

Note that if f is MOD_l then f_ρ is $\mathrm{MOD}_l^{\{a\}}$ for some a. If f is MAJORITY and the number of 0 and 1 assigned by ρ differ by at most 1 (which happens with probability $\Omega(n^{-\frac{1}{2}})$), we can fix at most 1 extra variable such that f_ρ will compute MAJORITY. In any case we can pick δ sufficiently small and obtain a contradiction to the degree lower bounds in Theorem 12 and Corollary 17, using Lemma 10.

The polynomials are obtained similarly as in the proof of Theorem 2 by applying a series of restrictions $\rho_1, \ldots, \rho_d$, where ρ_i leaves $n\left(n^{\frac{1}{2d}}\right)^{-i}$ variables free, simultaneously on all of the at most $2^{s+1}-1$ different circuits defined above. Here again the crucial step is to feed the terms of a DNF directly to the MOD_m gates, thus obatining a polynomial over $\mathbf{Z}_m$. $\square$

Acknowledgement We thank Emanuele Viola for sending his paper [18].

The first author is supported by a postgraduate scholarship from the Natural Sciences and Engineering Research Council (NSERC) of Canada and research grants of Prof. Denis Thérien. The second author is supported by BRICS, Basic Research in Computer Science, a Centre of the Danish National Research Foundation.

References

1. Furst, M., Saxe, J.B., Sipser, M.: Parity, circuits, and the polynomial-time hierarchy. Mathematical Systems Theory **17** (1984) 13–27
2. Ajtai, M.: Σ_1^1-formulae on finite structures. Annals of Pure and Applied Logic **24** (1983) 1–48
3. Håstad, J.: Computational limitations of small-depth circuits. MIT Press (1987)

4. Yao, A.C.C.: Separating the polynomial–time hierarchy by oracles. In: 26th Annual Symposium on Foundations of Computer Science, IEEE Computer Society Press (1985) 1–10
5. Razborov, A.A.: Lower bounds for the size of circuits of bounded depth with basis $(\wedge, \oplus)$. Mathematical Notes of the Academy of Science of the USSR **41** (1987) 333–338
6. Smolensky, R.: Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In: 19th Annual ACM Symposium on Theory of Computing. (1987) 77–82
7. Aspnes, J., Beigel, R., Furst, M.L., Rudich, S.: The expressive power of voting polynomials. Combinatorica **14** (1994) 135–148
8. Beigel, R., Reingold, N., Spielman, D.A.: PP is closed under intersection. Journal of Computer and System Sciences **50** (1995) 191–202
9. Beigel, R.: When do extra majority gates help? Polylog(n) majority gates are equivalent to one. Computational Complexity **4** (1994) 314–324
10. Barrington, D.A.M., Straubing, H.: Complex polynomials and circuit lower bounds for modular counting. Computational Complexity **4** (1994) 325–338
11. Hansen, K.A., Miltersen, P.B.: Some meet-in-the-middle circuit lower bounds. In: 29th International Symposium on Mathematical Foundations of Computer Science. Volume 3153 of Lecture Notes in Computer Science., Springer (2004) 334–345
12. Håstad, J., Goldmann, M.: On the power of small-depth threshold circuits. Computational Complexity **1** (1991) 113–129
13. Razborov, A., Wigderson, A.: $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. Information Processing Letters **45** (1993) 303–307
14. Babai, L., Nisan, N., Szegedy, S.: Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. Journal of Computer and System Sciences **45** (1992) 204–232
15. Green, F.: A complex-number fourier technique for lower bounds on the mod-m degree. Computational Complexity **9** (2000) 16–38
16. Beigel, R., Maciel, A.: Upper and lower bounds for some depth-3 circuit classes. Computational Complexity **6** (1997) 235–255
17. Krause, M., Pudlák, P.: On the computational power of depth-2 circuits with threshold and modulo gates. Theoretical Computer Science **174** (1997) 137–156
18. Viola, E.: Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. In: 20th Annual IEEE Conference on Computational Complexity, IEEE Computer Society Press (2005) (to appear).
19. Nisan, N., Wigderson, A.: Hardness vs randomness. Journal of Computer and System Sciences **49** (1994) 149–167
20. Chandra, A.K., Furst, M.L., Lipton, L.: Multi-party protocols. In: 15th Annual ACM Symposium on Theory of Computing, ACM Press (1983) 94–99
21. Chung, F.R.K., Tetali, P.: Communication complexity and quasi randomness. SIAM Journal on Discrete Mathematics **6** (1993) 110–123
22. Hajnal, A., Maass, W., Pudlák, P., Szegedy, M., Turán, G.: Threshold circuits of bounded depth. Journal of Computer and System Sciences **46** (1993) 129–154
23. Beame, P.: A switching lemma primer. Technical Report UW-CSE-95-07-01, Department of Computer Science and Engineering, University of Washington (1994) Availible online at www.cs.washington.edu/homes/beame.
24. Tardos, G., Barrington, D.A.M.: A lower bound on the mod 6 degree of the or function. Computational Complexity **7** (1998) 99–108