Lower bounds for circuits with MOD_m gates

Arkadev Chattopadhyay
McGill University, Montreal, Canada
achatt3@cs.mcgill.ca

Navin Goyal
McGill University, Montreal, Canada
navin@cs.mcgill.ca

Pavel Pudlák
Czech Academy of Sciences, Prague, Czech Republic
pudlak@math.cas.cz

Denis Thérien
McGill University, Montreal, Canada
denis@cs.mcgill.ca

Abstract

Let $\mathrm{CC}[m]$ be the class of circuits in which all gates are MOD_m gates. In this paper we prove lower bounds for circuits in $\mathrm{CC}[m]$ and related classes.

- Circuits in which all gates are MOD_m gates need $\Omega(n)$ gates to compute the MOD_q function, when m and q are co-prime. No non-trivial bounds were known before for computing MOD_q functions. Our argument is based on a new theorem about the boolean solutions of systems of linear equations over $\mathbb{Z}_m$, which may be of independent interest.
- When m is prime we get a similar theorem for systems of non-linear equations of small degree. As a consequence, we obtain linear lower bounds on the number of MOD_q gates in circuits of type $(\mathrm{MOD}_p \circ \mathrm{MOD}_q \circ \mathrm{AND}_{O(1)})$ computing MOD_r function where $(r, q) = (r, p) = 1$. The study of such circuits was intiated by Barrington et al. [3] as an important step towards understanding $\mathrm{CC}[m]$ circuits of constant depth.
- $\mathrm{CC}[m]$ circuits of constant depth need superlinear number of wires to compute both the AND and MOD_q functions. To prove this, we show that any circuit computing such functions has a certain connectivity property that is similar to that of superconcentration. We show a superlinear lower bound on the number of edges of such graphs extending results on superconcentrators.

1 Introduction

Proving lower bounds on the size of boolean circuits needed to compute explicit functions is of fundamental importance in theoretical computer science. Since the problem has proved to be very hard in general, various restricted models of circuits have been considered. One of the most fruitful directions has been the study of small depth circuits. The result (see [1, 10, 15, 30]) that circuits constructed using unrestricted fan-in OR, AND and NOT gates with constant depth (the class of circuits denoted by AC^0) need exponential size to compute the PARITY function, remains a jewel of this area. Smolensky [26], extending the work of [25], showed that sub-exponential size AC^0 circuits augmented with MOD_m gates (such circuits define the class $\mathrm{ACC}^0[m]$) cannot compute MOD_q if $(m, q) = 1$ and m is a prime power. However, the seemingly innocuous extension of these lower bounds to $\mathrm{ACC}^0[m]$ circuits for general m has remained open despite extensive efforts.

One of the main impediments seems to be understanding the power of MOD_m counting in this context. Define $\mathrm{CC}^0[m]$ to be the class of constant depth circuits composed only of MOD_m gates. Since it is difficult to compute the MOD_m function using AND and OR gates, it is a natural task to determine the smallest size $\mathrm{CC}^0[m]$ circuits computing AND and OR. It is known that both AND and MOD_q functions are impossible to compute by constant depth circuits composed entirely of MOD_m gates when m is a prime power. In contrast, it is also known that depth two MOD_6 circuits can compute every boolean function in exponential size [3]. A conjecture of [19] and a special case of a conjecture of Smolensky respectively imply that $\mathrm{CC}^0[m]$ circuits computing AND and MOD_q need exponential size whenever $(m, q) = 1$. Most known lower bounds, e.g., [3, 17, 13, 12] work only for special classes of $\mathrm{CC}^0[m]$ circuits. We do not even know if the satisfiability problem (SAT) can be solved by depth-2 linear size $\mathrm{CC}[6]$ circuits, when the gates used are *generalized* MOD_6 gates (see Section 2 for the definition of generalized MOD gate) [8].

The currently best known lower bound for AND for $\mathrm{CC}^0[m]$ is linear in the number of variables [28]. Previous to this work, no linear lower bounds were known for MOD_q. The difficulty in proving such lower bounds may be partly explained by the fact mentioned above that depth two $\mathrm{CC}[m]$ circuits can compute all boolean functions if m contains at least two different prime factors, but not if m is a prime power. The advantage of cmoposites over prime powers in computing the AND and MOD_q functions is also witnessed in the closely related setting of polynomials over $\mathbb{Z}_m$ where m is a composite which is not a prime power [2, 5, 14].

As a special case of $\mathrm{CC}^0[m]$ [3] considered $\mathrm{MOD}_p \circ \mathrm{MOD}_q$ circuits (those having depth two with a MOD_p gate at the output and a single layer of MOD_q gates at the input). A number of papers [3, 13, 27] show exponential lower bounds for such circuits computing AND and MOD_r, where $(r, p) = (r, q) = 1$. [3] formulate the Constant Degree Hypothesis (CDH) whose special case asserts that circuits of the type $\mathrm{MOD}_p \circ \mathrm{MOD}_q \circ \mathrm{AND}_{O(1)}$ (layered depth-3 circuits with AND gates of constant fan-in in the input layer, MOD_q gates in the middle layer, and a MOD_p gate at the output) require exponentially many MOD_q gates to compute AND. Some progress towards proving CDH is made by [29, 13, 12]. While obtaining the general CDH remains wide open, previous to our work even no linear lower bounds on the number of MOD_q gates were known without restricting the type of sub-circuits rooted at each MOD_q gate.

While the number of gates has been the more popular measure of circuit size, number of wires has also been studied fairly extensively, e.g., [9, 23, 24, 16]. The method in [16] is able to give a superlinear bound on the number of wires in ACC^0 circuits for only those functions, that have high communication complexity. Consequently, their method fails to give bounds on simple functions like AND and MOD_q.

Our results. Let $\mathrm{CC}[m]$ denote the class of circuits consisting of MOD_m gates *without* any depth restriction. In our discussion, unless otherwise specified, we always consider generalized MOD_m gates.

Let q be a positive integer and $b \in \{0, \ldots, q-1\}$. Define the bth MOD_q-residue class of $\{0,1\}^n$ by

$$M_{n,q}(b) = \{x = (x_1, \ldots, x_n) \in \{0,1\}^n \mid \sum_{i=1}^{n} x_i = b \mod q\}.$$

Lower bounds on the number of gates. One of the technical contributions of this paper is to prove the following *uniformity* property of boolean solutions of a system of linear equations over $\mathbb{Z}_m$ (see Lemma 4 and Theorem 5): If the number of equations in the system is at most dn for a small constant $d > 0$ then the boolean solutions to the system are essentially uniformly distributed among all the MOD_q-residue classes of $\{0,1\}^n$. The proof of this fact uses ideas from additive number theory, Fourier analysis and exponential sums. We apply the uniformity property to obtain:

Theorem 1 *For all positive integers q and m such that $(q, m) = 1$, $\mathrm{CC}[m]$ circuits computing $\mathrm{MOD}_q(x_1, \ldots, x_n)$ have size $\Omega(n)$.*

We say that a boolean function f is (c, m)-hard if the following holds: there does not exist a system L of cn homogeneous linear equations in n variables over Z_m such that f is constant over points in the boolean hypercube that satisfy L. We will show that for every such f and a $CC^0[m]$ circuit C having less than cn gates, there exists a boolean vector $b \in \{0,1\}^n$, such that $C(b) = C(0^n)$. Hence such a circuit cannot compute MOD_q. The main result in [28] essentially shows that AND is (c, m)-hard for all m. The uniformity property of the set of boolean solutions to a system of linear equations in $\mathbb{Z}_m$ implies that MOD_q is (c, m)-hard, whenever m and q are co-prime and $c = c(m, q)$ is some constant independent of n. Thus we get Theorem 1.

Lower bounds for circuits of type $\mathrm{MOD}_p \circ \mathrm{MOD}_q \circ \mathrm{AND}_{O(1)}$. For the case when $m = p$ is prime we can show a similar uniformity property of the set of boolean solutions to a system of small degree polynomial equations over $\mathbb{Z}_p$ (Lemma 10 and Theorem 11). This is done in Section 3 making use of the probabilistic method and a certain strong version of the Chevalley-Warning Theorem. This uniformity property yields the following :

Theorem 2 *For all primes p and q and integer r such that $(p, r) = (q, r) = 1$, circuits of type $(\mathrm{MOD}_p \circ \mathrm{MOD}_q \circ \mathrm{AND}_{O(1)})$ need $\Omega(n)$ MOD_q gates to compute both AND and MOD_r functions.*

Lower bound for number of wires. We give super-linear lower bounds on the number of wires in $\mathrm{CC}^0[m]$ circuits computing AND and MOD_q. To state our result more precisely, define for $d = 1, 2, \ldots,$,

$$\lambda_1(n) = \lceil \log_2 n \rceil,$$
$$\lambda_{d+1}(n) = \min\{i \in \mathbb{N};\ \lambda_d^{(i)}(n) \leq 1\},$$

where the superscript i denotes the i-times iterated function.

Theorem 3 *For every q and d there exist $\delta > 0, c > 0$ such that every circuit computing a (c,m)-hard boolean function $F(x_1, \ldots, x_n)$ that has depth $d+1$ and uses only MOD_m gates, has at least $\delta n \lambda_d(n)$ wires.*

We consider the bounded depth directed graph of a boolean circuit. The proof of the above theorem involves first showing that such graphs must satisfy a certain connectivity property similar to that of superconcentrators. We next prove a superlinear lower bound on the number of edges in such graphs. This theorem is stronger than lower bounds proved on bounded depth superconcentrators (when the depth of superconcentrator is even) and enables us to prove lower bounds on $\mathrm{CC}^0[m]$ circuits for which we cannot use superconcentrators. .

2 Bounds on the number of gates

For any vector $x \in \{0,1\}^n$, let x_i refer to its ith component, and $|x|$ denote its *weight* i.e. $\#\{i \mid x_i = 1\}$. For every positive integer m, we define the boolean function $\mathrm{MOD}_m : \{0,1\}^n \to \{0,1\}$ in the following way: $\mathrm{MOD}_m(x) = 1$ iff $\sum_{i=1}^n x_i \neq 0$ (mod m). For each $A \subseteq Z_m$, the *generalized* MOD_m^A boolean gate computes the following function : $\mathrm{MOD}_m^A(x) = 1$ iff $\sum_{i=1}^n x_i \ \in A$. The set A is called the accepting set of the MOD gate. We remark that the standard gate used in the literature is the one that has the accepting set $\{1, \ldots, m-1\}$. To avoid notational clutter, we shall denote by MOD_m^* a generalized gate without explicitly referring to its accepting set. However, in circuits that we consider, each gate would have its own accepting set that may or may not be the same as that of others.

Let θ be a set of r linear homogeneous forms $\theta_1, \ldots, \theta_r$, each of which is in n variables $x_1, \ldots, x_n$ over $\mathbb{Z}_m$, where m is a positive integer. Every such θ defines a linear map from $\mathbb{Z}_m^n$ into $\mathbb{Z}_m^r$ in a natural way. For any vector $v \in \mathbb{Z}_m^r$, let $K^\theta(v)$ denote the set of boolean points that are mapped to v by θ i.e. the set $\{x \in \{0,1\}^n \,|\, 1 \leq i \leq r,\ \theta_i(x) = v_i\}$.

We shall show the following lemma that essentially says that the elements of $K^\theta(v)$ are more or less uniformly distributed among the q mod classes, whenever q and m are relatively prime to each other:

Lemma 4 (Linear Uniformity Lemma) *For all positive integers q, m with $(q,m) = 1$, there exists a constant $\gamma = \gamma(m,q) < 1$, such that for all positive integers n, b, vector $v \in \mathbb{Z}_m^r$ and linear mapping $\theta : \mathbb{Z}_m^n \to \mathbb{Z}_m^r$, if $K^\theta(v)$ is non-empty, then*

$$\big||K^\theta(v) \cap M_{n,q}(b)| - |K^\theta(v)|/q\big| \leq (2\gamma)^n \qquad (1)$$

The Uniformity Lemma above becomes meaningful when the size of $K^\theta(v)$ is large enough so that the term $(2\gamma)^n$ in (1) behaves as an error-term. In this case, the points in $K^\theta(v)$ are almost *uniformly* distributed among the $M_{n,q}(b)$ classes for various values of b. We note that results in [28, 3] imply a lower bound of $(\frac{\alpha}{\alpha-1})^n \cdot \frac{1}{\alpha^r}$ for $|K^\theta(v)|$ when it is non-zero, where $\alpha = \alpha(m)$ is a constant. This is still not large to offset $(2\gamma)^n$. We obtain a sufficiently large bound on size of $K^\theta(v)$ in the Theorem below:

Theorem 5 *For any $v \in \mathbb{Z}_m^r$, if $K^\theta(v)$ is non-empty, then*

$$|K^\theta(v)| \geq \frac{2^n}{c^r}. \tag{2}$$

The proof of the Uniformity Lemma uses an exponential sum argument. Exponential sums have been previously used in similar contexts [7, 11]. As is standard, we use the notation $e_m(x)$ to denote $e^{2\pi ix/m}$, where i is the complex square root of -1.

Proof: [of Uniformity Lemma] Suppose $K^\theta(v)$ is non-empty. Then, $\theta(a) = v$ for some boolean vector a. Substituting $x_i = x_i - a_i$ and $b = b - \sum_{i=1}^n a_i$, for $1 \leq i \leq n$, we reduce to the case of v being the all-zero vector. For removing clutter, we denote $K^\theta(0^r)$ by K^θ. We first write $|K^\theta \cap M_{n,q}(b)|$ as an exponential sum and then estimate this exponential sum by grouping the terms appropriately.

$$|K^\theta \cap M_{n,q}(b)| = \sum_{x\in\{0,1\}^n} \Big[\prod_{i=1}^r \big(\frac{1}{m}\sum_{j=0}^{m-1} \mathrm{e}_m(j\theta_i(x))\big)\big(\frac{1}{q}\sum_{j=0}^{q-1} \mathrm{e}_q(j(\sum_{k=1}^n x_k - b))\big)\Big]. \tag{3}$$

The above identity is immediate from the well-known and simple fact that $\frac{1}{m}\sum_{j=0}^{m-1} \mathrm{e}_m(ja)$ is 1 if $a = 0$ and is 0 otherwise, for every positive integer m. We now rewrite the right hand side (RHS) in (3) as

$$(3) = \sum_{x\in\{0,1\}^n} \frac{1}{q}\prod_{i=1}^r \big(\frac{1}{m}\sum_{j=0}^{m-1} \mathrm{e}_m(j\theta_i(x)\big) + \sum_{x\in\{0,1\}^n} \Big[\prod_{i=1}^r \big(\frac{1}{m}\sum_{j=0}^{m-1} \mathrm{e}_m(j\theta_i(x))\big)\big(\frac{1}{q}\sum_{j=1}^{q-1} \mathrm{e}_q(j(\sum_{k=1}^n x_k - b))\big)\Big]. \tag{4}$$

The first term on the RHS is easily seen to be $|K^\theta|/q$. Hence, we get the following:

$$\big||K^\theta \cap M_{n,q}(b)| - |K^\theta|/q\big| = \Big|\sum_{x\in\{0,1\}^n} \Big[\prod_{i=1}^r \big(\frac{1}{m}\sum_{j=0}^{m-1} \mathrm{e}_m(j\theta_i(x))\big)\big(\frac{1}{q}\sum_{j=1}^{q-1} \mathrm{e}_q(j(\sum_{k=1}^n x_k - b))\big)\Big]\Big| \tag{5}$$

We now estimate the RHS of 5. To do this, let us multiply out the terms in the summand inside the abolute value and then sum the resulting terms. We obtain $m^r(q-1)$ terms after multiplying out the terms in the summand, each of which gives rise to a sum of the form:

$$\frac{\mathrm{e}_q(-jb)}{m^s q} \sum_{x\in\{0,1\}^n} \mathrm{e}_m(j_1\theta_1(x)+\ldots+j_r\theta_r(x))\mathrm{e}_q(j\sum_{k=1}^{n} x_k). \tag{6}$$

where $j \neq 0$. Writing $a_1x_1+\ldots+a_nx_n := j_1\theta_1(x)+\ldots+j_r\theta_r(x)$, using the trigonometric identity $1+e^{i2\rho} = 2e^{i\rho}\cos(\rho)$, and taking absolute values, we have

$$|(6)| = \left|\frac{1}{m^r q}\prod_{i=1}^{n}(1+\mathrm{e}_m(a_i)\mathrm{e}_q(j))\right| = \left|\frac{2^n}{m^r q}\prod_{i=1}^{n}\cos\left(\pi(\frac{a_i}{m}+\frac{j}{q})\right)\right|. \tag{7}$$

Let $\gamma = \max_{a_i\in Z_q;\, j\in Z_m}|\cos\left(\pi(\frac{a_i}{m}+\frac{j}{q}\right)|$. Since, m and q are co-prime and $j \neq 0$, it can be verified that $\gamma < 1$. Hence,

$$|(7)| \leq \frac{2^n\gamma^n}{m^r q}. \tag{8}$$

Using the triangle inequality on the RHS of (5) and plugging in the bound of (8), we get

$$\big||K^\theta \cap M_{n,q}(b)| - |K^\theta|/q\big| \leq m^r(q-1)\frac{(2\gamma)^n}{m^r q}. \tag{9}$$

This gives us the Uniformity Lemma. ∎

We now want to prove Theorem 5. To do so, we will have to introduce a notion from additive combinatorics: for any abelian group G, the *Davenport constant* of G (denoted by $s(G)$) is the smallest integer k such that every sequence of elements of G having length at least k, has a non-empty subsequence that sums to zero. Olson[21] showed that there exists a connection between $s(G)$ and the set of boolean solutions to the equation $g_1x_1+\ldots+g_nx_n = 0$ (denoted by $K(G,n)$), where each $g_i \in G$.

Theorem 6 (Olson's Theorem) $|K(G,n)| \geq \mathit{max}\{1, 2^{n+1-s(G)}\}$.

Note that the group we are interested in, is $\mathbb{Z}_m^r$ i.e. an equation in n variables over Z_m^r is equivalent to r equations over $\mathbb{Z}_m$ in the same set of variables. Recalling the argument as used at the beginning of the proof of the Uniformity Lemma, we get the following corollary:

Corollary 7 *For every θ and $v \in \mathbb{Z}_m^r$ such that $K^\theta(v)$ is non-empty, we have $|K^\theta(v)| \geq 2^{n+1-s(\mathbb{Z}_m^r)}$.*

To the best of our knowledge, determining $s(Z_m^r)$ for $r \geq 3$ and arbitrary m, is an open question. However, the independent works of [20, 28] based on Fourier analysis, imply the following upper bound:

Theorem 8 $s(Z_m^r) \leq (m \log\ m)r$.

Theorem 5 follows by combining Corollary 7 and bound on $s(Z_m^r)$ given by Theorem 8. The Uniformity Lemma and Theorem 5 immediately imply that

Corollary 9 *There is a constant $d' \in (0,1)$ depending on m and q such that if $r \leq d'n$ then $K^\theta(v) \cap M_{n,q}(b)$ is nonempty, for every $b \in \{0, \ldots, q-1\}$, whenever $K^\theta(v)$ is non-empty.*

We now show the lower bound on the number of gates needed by $\mathrm{CC}^0[m]$ circuits to compute the MOD_q function:

Proof:[of Theorem 1] Let the gates in the circuit be $G_1, \ldots, G_r$, where $r = o(n) < d'n$ and d' is given by Corollary 9. Let i_G be the set of all indices k such that G_k feeds into G_i. Consider the all-zero assignment $a = 0^n$ to the input variables. Let $\overline{G_i}(a) \in \mathbb{Z}_m$ and $G_i(a) \in \{0,1\}$ be respectively the value to which the ith gate evaluates on a *internally* and the boolean value it outputs in the circuit. For each gate i, we form the following affine equation : $\sum_{j=1}^n c_j^i x_j + \sum_{k \in i_G} G_k(a) = \overline{G_i}(a)$, where c_j^i is the number (modulo m) of copies of input bit x_i fed into G_i. By Corollary 9 if $r \leq d'n$ then there is a $b \in \{0,1\}^n$ such that all r affine equations are satisfied and $\mathrm{MOD}_q(b) \neq 0$. Hence for assignment b, each gate in the circuit evaluates (internally, and hence for the boolean outputs) to the same value as it evaluated to for assignment a. Thus, such a circuit cannot be computing the MOD_q function. ∎

3 Nonlinear Uniformity

In this section, we show that the linear uniformity theorem can be strengthened when m is a prime (we denote this prime by p). This will immediately yield Theorem 2. Let $S = \{\phi_1, \ldots, \phi_r\}$ be a set of r polynomials over $\mathbb{Z}_p$, where ϕ_i has degree d_i. Let $D = D(S) = d_1 + \ldots + d_r$ be the total degree of the system, and $\Delta = \Delta(S) = \max_{1 \leq i \leq r} d_i$ be the maximum degree among all polynomials in S. For $v \in \mathbb{Z}_p^r$, let K_n^S represent the set of points in $\{0,1\}^n$, that satisfy $\phi_i = v_i$ for all $1 \leq i \leq r$. We have

Lemma 10 (Nonlinear Uniformity Lemma) *Using the notation above, for all positive integers b, p, q, vector $v \in \mathbb{Z}_p^r$ with $(p,q) = 1$ and p prime, there exist constants α, β such that for all n and polynomial mapping $S : \mathbb{Z}^n \to \mathbb{Z}^r$, if $K_n^S(v)$ is non-empty, then*

$$\left||K_n^S(v) \cap M_{n,q}(b)| - |K_n^S|/q\right| \leq \left(\frac{2}{e^{\alpha/\beta^\Delta}}\right)^n. \tag{10}$$

The proof of this lemma, which appears in Appendix A, has a similar overall structure as the linear uniformity theorem, but now requires the use of some estimates on exponential sums due to [7, 11]. We want to use the nonlinear uniformity lemma to show that K_n^S intersects all residue classes mod q if the sum of the degrees of the polynomials in S is not too large. This will follow if we can show that $|K_n^S|$ is much larger than the right hand side in (10). The next theorem achieves this:

Theorem 11 *Using the notation above, we have* $|K_n^S(v)| \geq 2^n/p^{(p-1)D}$.

Before embarking on the proof we recall a strong form of the Chevalley-Warning theorem, whose elementary proof can be found in the book of Lidl and Niederreiter [18].

Theorem 12 (Chevalley-Warning) *Let $\phi_1, \ldots, \phi_s$ be s polynomials in $\mathbb{F}_a[x_1, \ldots, x_n]$, where $\mathbb{F}_a$ is a field of cardinality a. Let $D = \sum_{i=1}^{s} \deg(\phi_i) < n$, be the total degree of the system. Then, if the system of equations, $\phi_i(x_1, \ldots, x_n) = 0$, where $1 \leq i \leq s$, has a solution then it has at least a^{n-D} solutions in $\mathbb{F}_a^n$.*

Proof:[of Theorem 11] We will assume that K_n^S is nonempty, else there is nothing to prove. Recall that Fermat's little theorem says that for $y \in \mathbb{Z}_p$ we have $y^{p-1} = 1$ iff $y \neq 0$. To study the boolean solutions of S, we use the technique of replacing each variable x_i by y_i^{p-1} in every equation. Call the new system of equations S'.

Here we pause to give some intuition for the proof. We can lower bound the number of $\mathbb{Z}_p$-solutions of the system S' using the Chevalley-Warning theorem. However we want a lower bound on the number of boolean solutions of S. An immediate approach is to estimate how many $\mathbb{Z}_p$-solutions of S' can lead to the same boolean solution of S. This gives the following:

Note that the total degree of the system of new equations is $(p-1)D$. Theorem 12 can be applied to this new system of equations to conclude that the solution space in $\mathbb{Z}_p^n$ (denoted by K') has size at least $p^n/p^{(p-1)D}$. For any vector v in $\{0,1\}^n$, let $|v|$ denote the number of 1's in v. On the other hand, using Fermat's little theorem we get the following relation:

$$|K'| = \sum_{v \in K_n^S} (p-1)^{|v|} \leq |K_n^S| \cdot (p-1)^n \tag{11}$$

Combining these two observations we get $|K_n^S| \geq (\frac{p}{p-1})^n \cdot \frac{1}{p^{(p-1)D}}$. This however falls much short of what we need for Lemma 11. The way we resolve this difficulty is to consider maps from $\mathbb{Z}_p$-solutions to the boolean solutions more carefully. In fact, we consider a family of maps and then use a probabilistic argument to show that there is a choice of a map from this family that allows us to transfer the lower bound on the number of $\mathbb{Z}_p$-solutions to a good lower bound on the number of boolean solutions. We now continue with the proof.

Consider the equation $x^{p-1} - 1 = (x^{(p-1)/2} - 1)(x^{(p-1)/2} + 1) = 0$ in $\mathbb{Z}_p$. The solution set of this equation is $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$. Let S_p be the set of elements in $\mathbb{Z}_p$ that satisfy $f(x) = x^{(p-1)/2} - 1 = 0$. Clearly, $|S_p| = (p-1)/2$. Further, f evaluates to $p-2$ for every element of $\mathbb{Z}_p^*$ not in S_p. It can be verified that we can choose constants $a, b, c \in \mathbb{Z}_p$ such that the function $g(x) = a\cdot(f(x))^2 + b\cdot f(x) + c$ will evaluate to 0 for every element in $S_p \cup \{0\}$ and to 1 for all other elements. The degree of g is $p-1$.

Now consider the following random process: let f be a random function that is g with probability $1/2$ and $1-g$ with probability $1/2$. Let $f_1, \ldots f_n$ be n independent random functions each of which is identically distributed as f. Let $F : \mathbb{Z}_p^n \to \{0,1\}^n$ be the function defined by $F = f_1 \times \ldots \times f_n$. In each of the given equations, we replace each variable x_i by $f_i(x_i)$. Let N' be the random variable representing the number of solutions in $\mathbb{Z}_p^n$ for the system of equations obtained by the above process. Our bound will be obtained by estimating $\mathbf{E}[N']$ in two ways. The random system of equations that we get has total degree $(p-1)D$. Applying Chevalley-Warning, one thus gets $\mathbf{E}[N'] \geq p^{n-(p-1)D}$.

We count $\mathbf{E}[N']$ in another way. For any boolean vector u, let $F^{-1}(u)$ represent the set of vectors in $\mathbb{Z}_p^n$ that get mapped to u by F. Using linearity of expectation, one gets the following:

$$\mathbf{E}[N'] = \sum_{u \in K_n^S} \mathbf{E}[|F^{-1}(u)|] \tag{12}$$

Since each f_i is independent, for any $u \in K_n^S$, we get

$$\mathbf{E}[|F^{-1}(u)|] = \prod_{i=1}^{n} \mathbf{E}[|f_i^{-1}(u_i)|] \tag{13}$$

It is easily verified that $\mathbf{E}[|f_i^{-1}(u_i)|] = p/2$ for every i. Combining these observations we get $\mathbf{E}[N'] = (p/2)^n \cdot |K_n^S| \geq p^{n-(p-1)D}$. This immediately yields the bound we are looking for. ∎

Proof sketch of Theorem 2: The proof follows along the same lines as the proof of Theorem 1, only more simply. Briefly, suppose that the number of input MOD_q gates is $o(n)$. Then, using the nonlinear uniformity theorem we can fool the layer of MOD_q gates in the sense that there are two settings of the inputs such that the output of the MOD_q gates is the same on both the inputs but the MOD_k function takes different values, and thus the circuit is not computing MOD_k. It should be noted that this argument actually shows a stronger result, namely the lower bound holds irrespective of what is the output gate.

4 Lower bound on the number of wires

In this section we prove superlinear lower bound on the number of wires needed in a CC^0 circuit to compute (c, m)-hard functions, namely Theorem 3 .

This section is organized as follows. After setting up some notation we prove a superlinear lower bound on the number of edges in bounded depth graphs with a certain connectivity property. The proof is then completed by showing that the circuits in Theorem 3 satisfy this property and hence have superlinear number of edges.

Notation. Let G be a finite directed acyclic graph with a distinguished set of indegree zero vertices V_0, which will be called *input vertices*. Let X be a subset of input vertices. We shall say that a subset of vertices S *separates* X, if for every two different input vertices $x, y \in X$, every vertex v and every pair of directed paths p, q starting in x and y respectively and ending in v, at least one of the paths must contain a vertex from S. S may contain input vertices.

We shall say that X is ε*-separable*, if there exists an S such that S separates X and $|S| \leq \varepsilon|X|$.

We shall say that G is ε*-inseparable*, if for every subset of input vertices X, if $|X| \geq 2$, then X is not ε-separable. ($\varepsilon < 1$, as X separates itself.)

Define, for $d = 1, 2, \ldots,$

$$\lambda_1(n) = \lceil \log_2 n \rceil,$$

$$\lambda_{d+1}(n) = \min\{i \in \mathbb{N};\ \lambda_d^{(i)}(n) \leq 1\},$$

where the superscript i denotes the i-times iterated function.[1]

We can now state the theorem about graphs that we will use for our lower bound on the number of wires.

Theorem 13 *For every $\varepsilon > 0$ and every integer $d \geq 1$, there exists $\delta > 0$ such that for all n, if G has depth d, n inputs and it is ε-inseparable, then it has at least $\delta n \lambda_d(n)$ edges.*

We shall prove a stronger version of this theorem. For a set of inputs X of G, define

$$s(X) = \min\{|S|;\ S \text{ separates } X\}.$$

Let n be the number of input vertices, let $2 \leq t \leq n$, and $\varepsilon > 0$. We shall say that G is *weakly* t, ε*-inseparable*, if for all k, $t \leq k \leq n$,

$$\mathop{\mathbf{E}}_{|X|=k}(s(X)) > \varepsilon k.$$

The greater generality (in particular, the bound on the expectation, instead of an absolute bound) is needed for the proof.

Theorem 14 *For every $\varepsilon > 0$ and every integer $d \geq 1$, there exists $\delta > 0$ such that for every $2 \leq t \leq n$, every weakly t, ε-inseparable G of depth d with n input vertices has at least $\delta n \lambda_d(\frac{n}{t})$ edges.*

This theorem is proved by induction on the depth d. We shall assume w.l.o.g. that G is stratified into levels $V_0, V_1, \cdots, V_d$ and edges are only between consecutive levels. The following two lemmas formalize the induction base and the induction step.

[1] Note that the functions λ_i defined in [24] are different.

Lemma 15 *For every $\varepsilon > 0$, there exists $\delta > 0$ such that if G has depth 1, has n input vertices and it is weakly t, ε-inseparable, where $2 \leq t \leq n$, then it has more than $\delta n \log \frac{n}{t}$ edges.*

The proof appears in Appendix 4.

Lemma 16 *For every integer $d \geq 1$, reals $\varepsilon > 0$, and $\gamma > 0$, there exists $\delta > 0$ such that for every n, if*

(i) for every $2 \leq t \leq n$, every weakly $t, \frac{\varepsilon}{2}$-inseparable G of depth d with n input vertices has at least $\gamma n \lambda_d(\frac{n}{t})$ edges,

then

(ii) for every $2 \leq t \leq n$, every weakly t, ε-inseparable G of depth $d+1$ with n input vertices has at least $\delta n \lambda_{d+1}(\frac{n}{t})$ edges.

The proof appears in Appendix 4.

Proof:[Proof of Theorem 3] Let $0 < \varepsilon < \gamma$, let $\delta > 0$ be given by Theorem 13 for these ε and d. Suppose that the circuit has $< \delta n \lambda_d(n)$ edges. Then, by Theorem 13, there exists a set of inputs X which is ε-separated in the depth d graph obtained by removing the output gate from the circuit. Let S be the separating set augmented with the output gate. Then S is a separating set in the whole circuit and $|S| \leq \varepsilon|X| + 1$. We may moreover require that $|X| \geq \log n$, thus if n is sufficiently large, $|S| \leq \gamma|X|$.

Furthermore, for every $v \in S$, disconnect v from its inputs and set it to be the constant equal to the boolean value computed at v when all inputs are 0. Let C' be the resulting circuit. Let $v \in S$ and let w be an input gate of v in C. Then in C', the gate w only depends on at most one input from X, because S is a separating set. Thus if we put back the original MOD_m gate on v, the boolean function computed at v will be some MOD_m function G_v.

Thus in order to get a contradiction with the assumption that C computes $F(x_1, \ldots, x_n)$, we need only to find a boolean assignment $a \neq 0^n$ of $x_1, \ldots, x_n$ such that the variables outside X are set to 0 and the following holds: For every $v \in S$,

$$G_v(a) = G_v(0^n), \tag{14}$$

but $F(a) \neq F(0^n)$.

On the left hand side of (14) we replace each boolean function $G_v(\cdot)$ by its underlying linear form that takes values in $\mathbb{Z}_m$.

Then if the resulting linear system over $\mathbb{Z}_m$ is satisfied then so is (14). The assumption that F is (c, m)-hard guarantees the existence of a boolean solution $a \neq 0^n$ to this system such that $F(a) \neq F(0^n)$. Thus C cannot compute F. ∎

References

[1] M. Ajtai Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic* 24 (1983), 1–48.

[2] D. A. M. Barrington, R. Beigel, S. Rudich Representing Boolean Functions as Polynomials Modulo Composite Numbers. *Computational Complexity* 4 (1994), 367–382.

[3] D. Barrington, H. Straubing and D. Thérien, Non-uniform automata over groups. *Information and Computation* 89(2) (1990) 109–132.

[4] N. Bhatnagar, P. Gopalan, R. J. Lipton, The Degree of Threshold Mod 6 and Diophantine Equations Electronic Colloquium on Computational Complexity *ECCC* (2004) TR 022.

[5] N. Bhatnagar, P. Gopalan, R. J. Lipton, Symmetric Polynomials over Zm and Simultaneous Communication Protocol. FOCS (2003).

[6] P. van Emde Boas and D. Kruyswijk, A combinatorial problem on finite abelian groups III, Z.W. *(Math. Centrum, Amsterdam)* 1969–008.

[7] J. Bourgain, Estimation of certain exponential sums arising in complexity theory. *C. R. Acad. Sci. Paris*, Ser I 340 (2005), no. 9, 627–631.

[8] H. Caussinus, A Note on a Theorem of Barrington, Straubing and Thrien. *Inf. Process. Lett.* 58(1) (1996) 31–33.

[9] A. K. Chandra, S. Fortune, R. J. Lipton, Lower Bounds for Constant Depth Circuits for Prefix Problems. *ICALP* 1983 109–117.

[10] M. Furst, J.B. Saxe and M. Sipser Parity, circuits and the polynomial-time hierarchy. *Math. Systems Theory* 18 (1984), 13–27.

[11] F. Green, A. Roy and H. Straubing, Bounds on an exponential sum arising in boolean circuit complexity. *C. R. Acad. Sci. Paris*, Ser I 341 (2005), 279–282.

[12] V. Grolmusz, A Degree-Decreasing Lemma for (MOD-q - MOD-p) Circuits. *Discrete Mathematics and Theoretical Computer Science*, 4(2) (2001) 247–254.

[13] V. Grolmusz, G. Tardos, Lower Bounds for (MOD-p - MOD-m) Circuits. *SIAM J. Comput.* 29(4) (2000) 1209–1222.

[14] K. A. Hansen, On Modular Counting with Polynomials. To appear in Computational Complexity Conference 2006.

[15] J. Håstad, Computational limitations on small depth circuits. Ph.D Thesis, MIT, (1986).

[16] M. Koucký, P. Pudlák and D. Thérien, Boundeddepth circuits: Separating wires from gates. *37th ACM STOC*, 2005, pp.257-265.

[17] M. Krause, S. Waack, Variation Ranks of Communication Matrices and Lower Bounds for Depth Two Circuits Having Symmetric Gates with Unbounded Fan-In. *FOCS* 777–782 (1991).

[18] R. Lidl, H. Niederreiter, Finite Fields. Cambridge University Press, 1997.

[19] P. McKenzie, P. Pladeau, D. Thrien, NC1: The Automata-Theoretic Viewpoint. *Computational Complexity* 1: 330-359 (1991).

[20] R. Meshulam, An uncertainty inequality and zero subsums. *Discrete Mathematics* (84) 1990, 197–200.

[21] J.E. Olson, A combinatorial problem on finite abelian groups, II. *J. Number Theory* 1 (1969), 195–199.

[22] P. Pudlák, Communication in bounded depth circuits. *Combinatorica* 14(2), pp.203–216, 1994.

[23] P. Ragde and A. Wigderson Linear-size constant-depth polylog-threshold circuits. *Information Processing Letters*, 39 (1991), 143–146

[24] R. Raz and A. Shpilka, Lower bounds for matrix product in bounded depth circuits with arbitrary gates. *SIAM J. on Computing* 32(2), 2003, pp.488-513.

[25] A. Razborov, Lower bounds on the size of bounded-depth networks over a complete basis with logical addition. *Mathematical Notes of the Academy of Sci. of the USSR*, 41(4): 333–338, 1987.

[26] R. Smolensky, Algebraic methods in the theory of lower bounds for boolean circuit complexity. *19th STOC* (1987), 77–82.

[27] H. Straubing, D. Thérien, A Note on MODp - MODm Circuits. accepted in Theory of Computing Systems.

[28] D. Thérien, Circuits constructed with MOD_q gates cannot compute AND in sublinear size. Comput. Complexity 4 (1994), no. 4, 383–388.

[29] P. Y. Yan, I. Parberry, Exponential Size Lower Bounds for Some Depth Three Circuits Inf. Comput. 112(1): 117–130 (1994)

[30] A. Yao, Separating the polynomial-time hierarchy. *26th FOCS*, (1985), 1–10.

A Proof of Theorem 10

We now state an upper-bound for an exponential sum that appeared in [7, 11]:

Fact 17 *Let q, m be any relatively prime numbers. Further, let*

$$S = \sum_{x\in\{0,1\}^n} e_m(\phi(x))e_q(a\sum_{i=1}^{n} x_i) \tag{15}$$

where $\phi(x) = \phi(x_1, \ldots, x_n)$ is a polynomial of degree d with coefficients in Z_q. Then, there exists $0 < \alpha < 1$ such that $|S| \leq (2\mu)^n$, where $\mu < 1 - \frac{\alpha}{(m2^m)^d}$ and α depends only on m and q.

Now we can complete the proof of Theorem 10.

Proof: [of Theorem 10] For simplicity we will work with the case where $v = 0$ is the all-zero vector; other cases are handled similarly. We write K_n^S for $K_n^S(0)$. As in the proof of Theorem 4, we get the following:

$$|K_n^S \cap M_{n,q}(b)| = \sum_{x\in\{0,1\}^n} \Big[\prod_{i=1}^{r} \big(\frac{1}{p}\sum_{j=0}^{p-1} \mathrm{e}_p(j\phi_i(x))\big)\big(\frac{1}{q}\sum_{j=0}^{q-1} \mathrm{e}_q(j(\sum_{k=1}^{n} x_k - b))\big)\Big]. \tag{16}$$

As before, this can be re-written as :

$$(16) = |K_n^S|/q + R \tag{17}$$

where R is a sum of $p^s(q-1)$ terms, each of which is of the form

$$\frac{\mathrm{e}_q(-jb)}{p^s q} \sum_{x\in\{0,1\}^n} \mathrm{e}_p(j_1\phi_1(x) + \ldots + j_r\phi_r(x))\mathrm{e}_q(j\sum_{k=1}^{n} x_k). \tag{18}$$

Note that the degree of the form $j_1\phi_1(x)+\ldots+j_r\phi_r(x)$ is at most $\Delta(S)$ for every $(j_1, \ldots, j_r) \in [m]^r$. Using the bound on (15) in Fact 17 and the fact that $1 - x \leq e^{-x}$, one can write

$$|R| \leq \frac{q-1}{q}\big(\frac{2}{e^{\alpha/\beta^{\Delta}}}\big)^n \tag{19}$$

where $\beta = p2^p$. Applying the bound in (19) to (17), we get (10) proving Theorem 10. ∎

We can easily combine Theorem 10 and Theorem 11 to get the following:

Corollary 18 *There exist constants α, β that depend only on m and q such that if*

$$D(S) \cdot \beta^{\Delta(S)} < \frac{\alpha}{\log p} n \tag{20}$$

then $K_n^S \cap M_{n,q}(b)$ is nonempty, for every $b \in \{0, \ldots, q-1\}$.

B Proofs from Section 4

Proof:[of Lemma 15] Suppose G is weakly t, ε-inseparable. Let $v_1, v_2, \dots$ be all vertices on the level 1 (the level 0 being the input vertices) ordered by the decreasing indegrees $d_1 \geq d_2 \geq \dots$. For $t \leq q \leq \frac{\varepsilon n}{2}$ consider the undirected graph H_q with the set of vertices being the input vertices of G and edges (x, y) such that $x \to v_i, y \to v_i$ in G for some $i > q$. Thus H_q has $m \leq \sum_{i>q} \binom{d_i}{2}$ edges. Let X be a random subset of inputs of cardinality $k = \lceil \frac{2q}{\varepsilon} \rceil$ (thus $t \leq k \leq n$). The expected number of edges on X is $\frac{m}{\binom{n}{2}} \binom{k}{2}$.

Observe that if there are ℓ edges of H_q on X, then $s(X) \leq \ell + q$ (take the vertices $v_1, \cdots, v_q$ and one vertex from each edge). Thus we have

$$\frac{m}{\binom{n}{2}} \binom{k}{2} + q \geq \mathbf{E}(s(x)) > \varepsilon k.$$

Since $q \leq \varepsilon k/2$, we have

$$\frac{m}{\binom{n}{2}} \binom{k}{2} > \frac{\varepsilon k}{2}.$$

Substituting for m and simplifying we get

$$\sum_{i>q} \frac{\binom{d_i}{2}}{\binom{n}{2}} > \frac{\varepsilon}{k-1}.$$

Since $d_i \leq n$, we can estimate $\frac{\binom{d_i}{2}}{\binom{n}{2}} \leq \frac{d_i^2}{n^2}$. Thus we get

$$\sum_{i>q} \frac{d_i^2}{n^2} > \frac{\varepsilon}{k-1} = \frac{\varepsilon}{\lceil \frac{2q}{\varepsilon} \rceil - 1} \geq \frac{\varepsilon^2}{2q}.$$

By Lemma 4 of [22], this implies

$$\sum_i \frac{d_i}{n} \geq \delta_1 \log \frac{\lfloor \frac{\varepsilon n}{2} \rfloor}{t},$$

for some $\delta_1 > 0$ depending only on ε. Hence if $t = o(n)$, we get

$$\sum_i d_i \geq \delta n \log \frac{n}{t}.$$

Otherwise use the trivial lower bound εt on the number of edges. ∎

Proof:[of Lemma 16] Suppose (i) holds true. Let G be weakly t, ε-inseparable directed graph with depth $d+1$ and n input vertices.

Let us briefly sketch the idea of the proof before doing detailed computations. We would like to distinguish two cases: either there are a lot of vertices of high degree on the first level, or not. In the first case there are, clearly, many edges. In the second case we can delete the vertices on the first level that have large degrees, connect inputs directly to the second level and then we can apply (i) to the resulting depth d graph. However, this does not quite work, as after deleting the vertices with high degree, the degrees of the remaining vertices on level 1 are still too large. Therefore we have to consider also vertices with intermediate degrees. If the number of those vertices would be small, then a random set of inputs would meet only a few edges connected to them.

Let $deg(v)$ denote the indegree of a vertex v. Let t be given, $2 \le t \le n$. Put $r = \frac{n}{t}$,

$$A_0 = \{v \in V_1;\ deg(v) > \lambda_d(r)\},$$

$$A_i = \{v \in V_1;\ \lambda_d^{(i+1)}(r) < deg(v) \le \lambda_d^{(i)}(r)\},\ \text{for } i \ge 1.$$

Let E denote the set of edges of G.

Claim. For every i, $1 \le i \le \lambda_{d+1}(r)/2 - 3$, at least one of the following three inequalities is satisfied:

1. $|A_0 \cup \cdots \cup A_{i-1}| \ge \frac{\varepsilon}{4} \frac{n}{\lambda_d^{(i+1)}(r)}$;
2. $|\{(u,v) \in E;\ u \in V_0, v \in A_i \cup A_{i+1} \cup A_{i+2}\}| \ge \frac{\varepsilon}{4} n$;
3. $|\{(u,v) \in E;\ u,v \notin A_0 \cup \cdots \cup A_{i+2}\}| \ge \gamma n \frac{\lambda_d^{(i+2)}(r)}{\lambda_d^{(i+3)}(r)}$.

Proof of Claim. Let i be given and suppose that conditions (1) and (2) are false. Let $n/\lambda_d^{(i+1)}(r) \le k \le n$. Observe that $n/\lambda_d^{(i+1)}(r) = n/\lambda_d^{(i+1)}(n/t) \ge t$, since $\lambda_d(x) \le x$ for all x. Let $X \subseteq V_1$ be a random subset of size k. We shall show that if we remove from G all edges incident with $A_0 \cup \cdots \cup A_{i+2}$, then

$$\mathbf{E}(s'(X)) > \frac{\varepsilon}{2} k,$$

where $s'(X)$ denotes $s(X)$ in the modified graph, which we shall denote by G'.

Indeed, let $a = |A_0 \cup \cdots \cup A_{i-1}|$, $b(X) = |\{(u,v) \in E;\ u \in X, v \in A_i \cup A_{i+1} \cup A_{i+2}\}|$. Then

$$s(X) \le a + b(X) + s'(X).$$

Hence

$$\mathbf{E}(s'(X)) \ge \mathbf{E}(s(X) - b(X) - a) = \mathbf{E}(s(X)) - \mathbf{E}(b(X)) - a.$$

By non-1, $a < \frac{\varepsilon}{4} \frac{n}{\lambda_d^{(i+1)}(r)} \le \frac{\varepsilon}{4} k$. By non-2, we have $\mathbf{E}(b(X)) < \frac{\varepsilon}{4} k$, (each edge from $\{(u,v) \in E;\ u \in V_0, v \in A_i \cup A_{i+1} \cup A_{i+2}\}$ is chosen with probability k/n; use the linearity of expectation).

Thus G' is weakly $n/\lambda_d^{(i+1)}(r), \frac{\varepsilon}{2}$-inseparable.

We shall further modify G' by removing all edges between V_1 and V_2 and adding, for every path (u, v, w) in G' with $u \in V_0, v \in V_1, w \in V_2$, the edge (u, w). The resulting graph will be denoted by G''. It has depth d (the first level being $V_1 \cup V_2$, the second level being V_3 etc.) and at most $\lambda_d^{(i+3)}(r)$-times more edges.

Furthermore, G'' is also weakly $n/\lambda_d^{(i+1)}(r), \frac{\varepsilon}{2}$-inseparable. To see that, observe that if X is a set of inputs (in G' and G'') and S is a separating set for X in G'', then S is a separating set for X also in G'. Indeed, let S be a separating set for X in G'' and let $(v_0, \cdots, v_j)$ and $(u_0, \cdots, u_j)$ be two paths in G', $v_0, u_0 \in X$, $v_0 \neq u_0$ and $v_j = u_j$. Then if $j = 1$, these paths are also paths in G'', and if $j > 1$, $(v_0, v_2, \cdots, v_j)$ and $(u_0, u_2, \cdots, u_j)$ are paths in G''. In both cases they contain an element from S, whence the original pair of paths also contains an element from S. Thus separating sets are at least as large in G'' as in G'.

By the assumption (i), G'' must have at least $\gamma n \lambda_d(\lambda_d^{(i+1)}(r)) = \gamma n \lambda_d^{(i+2)}(r)$ edges. Hence G' has at least $\gamma n \lambda_d^{(i+2)}(r)/\lambda_d^{(i+3)}(r)$ edges, which proves 3. This finishes the proof of the Claim.

To finish the proof of Lemma 16, we shall use the inequality

$$\frac{\lambda_d^{(i)}(r)}{\lambda_d^{(i+1)}(r)} \geq \tfrac{1}{2}\lambda_{d+1}(r),$$

for every $i \leq \lambda_{d+1}(r)/2 - 1$, which was proved in [22] as Lemma 5. By the Claim it suffices to consider the following three cases.

1. Suppose for some $i \leq \lambda_{d+1}(r)/2 - 3$ the condition (i) of Claim is satisfied. Then, since every $v \in A_0 \cup \cdots \cup A_{i-1}$ has degree $> \lambda_d^{(i)}(r)$, the number of edges in G is at least

$$\frac{\varepsilon}{4}\frac{n}{\lambda_d^{(i+1)}(r)}\lambda_d^{(i)}(r) \geq \frac{\varepsilon}{8} n \lambda_{d+1}(r).$$

2. Suppose for *all* $i \leq \lambda_{d+1}(r)/2 - 3$ the condition (ii) of Claim is satisfied. Then the number of edges of G is at least

$$\tfrac{1}{3}(\lambda_{d+1}(r)/2 - 3)\frac{\varepsilon}{4} n = \Omega(n\lambda_{d+1}(r)).$$

3. Suppose for some $i \leq \lambda_{d+1}(r)/2 - 3$ the condition (iii) of Claim is satisfied. Then the number of edges of G is at least

$$\gamma n \frac{\lambda_d^{(i+2)}(r)}{\lambda_d^{(i+3)}(r)} \geq \tfrac{1}{2}\gamma n \lambda_{d+1}(r).$$

∎