Linear systems over composite moduli

Arkadev Chattopadhyay*
School of Mathematics
Institute for Advanced Study
Princeton, NJ, USA
achatt3@ias.edu

Avi Wigderson†
School of Mathematics
Institute for Advanced Study
Princeton, NJ, USA
avi@ias.edu

Abstract— We study solution sets to systems of *generalized* linear equations of the form

$$\ell_i(x_1, x_2, \cdots, x_n) \in A_i \pmod m$$

where $\ell_1, \ldots, \ell_t$ are linear forms in n Boolean variables, each A_i is an arbitrary subset of $\mathbb{Z}_m$, and m is a *composite* integer that is a product of two distinct primes, like 6. Our main technical result is that such solution sets have exponentially small correlation, i.e. $\exp\big(-\Omega(n)\big)$, with the boolean function MOD_q, when m and q are relatively prime. This bound is independent of the number t of equations.

This yields progress on limiting the power of constant-depth circuits with modular gates. We derive the first exponential lower bound on the size of depth-three circuits of type MAJ $\circ$ AND $\circ$ MOD_m^A (i.e. having a MAJORITY gate at the top, AND/OR gates at the middle layer and *generalized* MOD_m gates at the base) computing the function MOD_q. This settles an open problem of Beigel and Maciel [5], for the case of such modulus m.

Our technique makes use of the work of Bourgain [6] on estimating exponential sums involving a *low-degree polynomial* and ideas involving matrix rigidity from the work of Grigoriev and Razborov [15] on *arithmetic circuits* over finite fields.

Keywords-boolean circuit complexity; constant-depth circuits; modular gates; exponential sums; matrix rigidity

1. Introduction

1.1. Background

Proving strong lower bounds on the size of constant-depth boolean circuits comprising MOD_m gates for computing an explicit function is a fundamental open problem in theoretical computer science. Despite the fact that Razborov [23] and Smolensky [24] obtained strong lower bounds, more than twenty years ago, on the size of circuits of constant depth having AND, OR and MOD_p gates, it has proved surprisingly difficult to extend that result to composite modular counting (see for example [2], [25], [22], [16], [27], [4], [18], [26]). The class of boolean functions that can be computed by circuits of constant depth and polynomial size, having AND, OR and MOD_m gates, where m is any fixed positive integer, is called ACC^0. It is the smallest naturally arising circuit complexity class that currently cannot be separated from NP.

Part of the difficulty of this problem was explained by surprising upper bounds, where a composite modulus, even a MOD_6 gate, allows more efficient algorithms than a prime modulus MOD_p. The canonical example of this power is that *every* Boolean function can be computed by a depth-2 circuit of MOD_6, whereas for any prime p, a depth-2 circuit (indeed, any constant depth circuit) of MOD_p gates can only compute Boolean functions which are constant degree polynomials over $\mathbb{Z}_p$, an exponentially small fraction of all Boolean functions. Yet another example of that power was demonstrated by Barrington, Beigel and Rudich [3]. They showed that while polynomials representing the AND function on n variables require degree $\Omega(n)$ over the field F_p for any fixed prime p, this function has degree $O(\sqrt{n})$ over the ring $\mathbb{Z}_6$. Moreover, if m has t distinct prime factors the degree upper bound drops further to $n^{1/t}$. This advantage of a composite modulus is not restricted to just computing the AND function, but also comes into play for computing MOD_q as exhibited by Hansen [20].

Another distinction surfaces when defining MOD_m as a Boolean function. A flexibility, used in many of these upper bounds, is to pick a subset $A \subset \mathbb{Z}_m$, and let $\mathrm{MOD}_m^A(z_1, \cdots, z_k)$ output 1 if $z_1 + \cdots + z_k (\mathrm{mod}\ m) \in A$ and 0 otherwise. It is easy to see that if $m = p$ is prime, than the choice of A is immaterial, in the sense that constant-depth circuits of such gates (with varying A's) can be simulated with similar size and depth circuits in which A is fixed for all MOD_p gates, say $A = \{0\}$. This reduction uses the identity $x^p \equiv x$ over the field F_p, which fails for rings $\mathbb{Z}_m$ for composite m.

Indeed, it is known even in contexts outside of circuit complexity that the flexibility of choosing an arbitrary accepting set A affords non-trivial advantage over choosing a singleton accepting set. A striking example of this is the recent design of 3-query locally decodable codes of sub-exponential length by Efremenko [10], using the earlier intriguing construction of set systems by Grolmusz [17]. Finally, and this point will be crucial for this work, linear systems of equations modulo m are completely understood when $m = p$ is a prime, due to the availability of division

*Research supported partially by a NSERC postdoctoral fellowship and NSF Grant CCF-0832797.
†Research partially supported by NSF grants CCF-0832797 and DMS-0835373.

and Gaussian elimination. This breaks down when m is composite, and some of the upper bounds use the strange structure of Boolean solutions to linear equations over $\mathbb{Z}_m$.

Can this extra power and complexity of composite moduli help significantly in computing functions using modular gates? It remains consistent with our knowledge that circuits, comprising only MOD_6 gates, of depth-three and linear size can compute an NP-complete function like SAT. On the other hand, Smolensky [24] conjectured that circuits having AND, OR and MOD_m gates, cannot even compute the MOD_q function in sub-exponential size and constant depth, when m, q are co-prime. This remains an outstanding conjecture and is one of the driving themes of past work and our work here.

1.2. Past lower bounds

To attack Smolensky's conjecture researchers have considered a variety of restricted models, and have tried to prove weaker lower bounds in attempt to develop proof techniques dealing with modular counting.

Chattopadhyay and Hansen [9] have proved super-polynomial lower bounds on the size of AC^0 circuits augmented with a *few* MOD_m gates for computing MOD_q. Chattopadhyay, Goyal, Pudlák and Thérien [8] proved linear lower bounds on the number of gates and super-linear lower bounds on the number of wires, for circuits with only MOD_m gates computing MOD_q.

Some exponential lower bounds were obtained for more restricted models, in which there is only a *single* layer of modular gates in the circuit. Both were achieved for depth-three circuits only.

One such result, following a sequence of earlier results [11], [22], [12], [13], is Bourgain's exponential lower bound [6], for $\mathrm{MAJ} \circ \mathrm{MOD}_m^A \circ \mathrm{AND}_{o(\log n)}$ circuits, in which the modular gates are in the middle layer, and the bottom layer has AND gates of small fan-in (up to $o(\log n)$ for fixed m, q). The intense interest in this kind of circuits followed from the surprising observation by Allender [1] that showed that these circuits can simulate in quasi-polynomial size and poly-logarithmic bottom fan-in, circuits of arbitrary but constant depth and quasi-polynomial size comprising AND, OR and MOD_{p^k} gates, where p is any prime dividing m and k is a constant integer. No non-trivial lower bounds are currently known for such circuits, even when the bottom fan-in is $\log n + 1$.

The other result is by Beigel and Maciel [5], who proved exponential lower bounds for $\mathrm{MAJ} \circ \mathrm{AND} \circ \mathrm{MOD}_m^{\{0\}}$ circuits for computing MOD_q, in which the modular gates are at the bottom layer and have a singleton accepting set. To prove that, they use an argument similar to the one used by Razborov and Smolensky in the case of MOD_p gates, to reduce the fan-in of the AND gates to a constant. They, then use arguments from the earlier work of Krause and Pudlák [22] who proved exponential lower bounds for $\mathrm{MAJ} \circ \mathrm{AND}_{O(1)} \circ \mathrm{MOD}_m$ circuits, i.e. AND gates in the middle layer are restricted to have constant fan-in. Unfortunately, the Beigel-Maciel technique breaks down for general MOD_m^A gates. In particular, there is no known way of reducing the fan-in of AND gates when they receive their inputs from generalize MOD_m gates. In fact, as a MOD_m gates with a singleton accepting set is not closed under complementation, no non-trivial lower bound was known even for circuits of type $\mathrm{AND} \circ \mathrm{OR} \circ \mathrm{MOD}_m^{\{0\}}$.

1.3. New results and techniques

In this paper, we improve upon the result of Beigel and Maciel, obtaining exponential lower bounds for $\mathrm{MAJ} \circ \mathrm{AND} \circ \mathrm{MOD}_m^A$ circuits computing MOD_q. Specifically, we can handle general modular gates MOD_m^A in the bottom layer. Let $A \subseteq \mathbb{Z}_m$ be an arbitrary set. Then, the boolean function MOD_m^A is defined as follows: $\mathrm{MOD}_m^A(x_1, \ldots, x_n) = 1$ iff $x_1 + \cdots + x_n \in A$ modulo m. We show:

Theorem 1 (Main Theorem) *Let m, q be co-prime integers such that m is square-free and has at most two prime factors. Let C be any circuit of type $\mathrm{MAJ} \circ G \circ \mathrm{MOD}_m^A$ where G is either AND or OR gate and the MOD_m gates at the base have arbitrary accepting sets. If C computes MOD_q then the top fan-in, and hence the circuit size, must be $2^{\Omega(n)}$.*

It is worth noting that for the special case of a singleton accepting set A, our technique yields a much stronger consequence as stated in Theorem 6 in Section 3.1. In particular, it proves lower bounds for depth-four circuits without making any assumptions on m, save that it is a constant. This subsumes the result of Beigel and Maciel [5] for constant modulus m.

Like other results for circuits with a top level Majority gate (see [19], [11], [12]), the key technical part is obtaining an exponentially small correlation bound of the target MOD_q function with any depth-2 sub-circuit of our circuit. This we obtain by an exponential sum bound which is the main technical contribution of this paper. We note that the depth-2 circuits we consider are of the form $\mathrm{AND} \circ \mathrm{MOD}_m^A$, which accept solutions to a system of linear equations MOD_m (more precisely, equations of the form $\ell_i(x) \in A_i$ where each ℓ_i is a linear form and A_i is a subset of $\mathbb{Z}_m$). We show that such solution sets have only exponentially small correlation with MOD_q.

Define the correlation of a function f with MOD_q, denoted by $\mathrm{Corr}(f, \mathrm{MOD}_q)$, as follows:

$$\max_{a \in \mathbb{Z}_q} \left\{ \left| \Pr_x \left[f(x) = 1 \,\middle|\, \sum_i x_i = a \pmod q \right] - \Pr_x \left[f(x) = 1 \,\middle|\, \sum_i x_i = 0 \pmod q \right] \right| \right\}$$

Lemma 2 (Main Technical Result) *Let C be any circuit of type $G \circ \mathrm{MOD}_m^A$, where m is a fixed square-free integer that has at most two distinct prime factors and G is either an AND or an OR gate. Then, for any fixed q that is co-prime with m,*

$$Corr\big(C, \mathrm{MOD}_q\big) \leq exp\big(-\Omega(n)\big) \tag{1}$$

Note the "duality" with Bourgain's similar exponential correlation bound for $\mathrm{MOD}_m^A \circ \mathrm{AND}$, in which the order of the conjunction and modular counting are reversed.

In order to prove our result, we are naturally lead to study the set of boolean solutions to linear systems of t equations of the form

$$\ell_i(x_1, x_2, \cdots, x_n) \in A_i \pmod m \tag{2}$$

(where $\ell_1, \cdots \ell_t$ are linear forms). We first show that when each A_i is a singleton, then using simple exponential sums, one can prove exponentially small upper bounds on the correlation between the solution set of such equations and MOD_q. This provides a very short and simple alternative proof to the Biegel-Maciel result. We also show how to extend this to solutions of *polynomial* equations of low degree, as long as the modular gates are *singleton*. Using this, we get exponential lower bounds on depth-4 circuits of the form $\mathrm{MAJ} \circ \mathrm{AND} \circ \mathrm{MOD}_m \circ \mathrm{AND}$, when the bottom AND gates have sub-logarithmic fan-in.

It is easy to see that, running over all possible choices of elements $a_i \in A_i$, the solution set to the above system is the union over $\exp(t)$ "normal" linear systems of the form $\ell_i(x_1, x_2, \cdots, x_n) = a_i \pmod m$. But as t may be arbitrarily large, one cannot simply use a union bound.

The main idea that we use to overcome this difficulty originates in the world of *arithmetic circuits*. To see the connection, observe that when working in e.g. F_3, addition in the field is a (non Boolean) MOD_3 gate, while multiplication (when restricted to the nonzero elements 1, -1) is equivalent to a MOD_2 computation. Thus MOD_6 gates can easily perform both field operations. This explains their power, mentioned before, to compute every Boolean function in depth two. A natural idea, which has been used on arithmetic circuits, is to focus on the linear forms ℓ_i, and treat differently the cases when they are "low rank" and "high rank". Intuitively, thinking that equations in (2) were over a field, for high rank $\geq r$ there will only be $\exp(-r)$ solutions to such systems (and so low correlation with any nontrivial function), and for high rank $\leq r$ the union bound above will only be over $\exp(r)$, as opposed to $\exp(t)$ cases of singleton equations, which can be hopefully handled by simpler methods.

And this idea can be made to work! However, its implementation is quite complex. The main problem of course is that we are not over a field. Thus, even standard notions of rank are problematic, and linear algebraic methods as above cannot be used directly. To resolve this, we borrow and generalize the ingenious definitions of "rigidity-rank" and "communication-rank", introduced by Grigoriev and Razborov [15] to handle a related problem, in the context of depth-3 arithmetic circuits over finite fields. On the high-rank part these generalizations are straightforward. In the low-rank part they raise complications special to the fact that we use composite moduli. In particular, we need to handle the special case of a combination of *sparse* linear systems (where each equation has few nonzero coefficients) with low-rank systems. We do so using estimates of exponential sums by Bourgain involving *low-degree* polynomials over $\mathbb{Z}_m$. This is the part which restricts our result to handle only moduli m with just two distinct prime factors.

Our analysis further reveals that in the low-rank case we can prove exponential correlation bounds not only for $\mathrm{AND} \circ \mathrm{MOD}_m^A$ circuits, but also for restricted $\mathrm{MOD}_m^A \circ \mathrm{MOD}_m^A$ circuits. While still far from general lower bounds for such depth-2 circuits of composite modular gates, we hope that our partial result here may be useful in attacking this important challenge. Currently, no super-linear lower bounds are known in general for such depth-two circuits.

Finally, it is worth noting that our work is interesting from another point of view. Recently, Hansen and Koucký [21] have observed that polynomial size ACC^0 circuits can be simulated by poly-size $\mathrm{OR} \circ \mathrm{AND} \circ \mathrm{CC}^0$ circuits, where CC^0 denotes constant-depth circuits comprising only modular gates. Our result is a natural first step towards obtaining lower bounds for circuits of type $\mathrm{OR} \circ \mathrm{AND} \circ \mathrm{CC}^0$.

1.4. Organization

After some preliminaries in Section 2, we present the lower bound in the "low-rigid" case in Section 3, and in the complementary "high-rigid" case in section 4, motivating in both cases the exact definitions of these terms. Section 5 combines them.

2. Preliminaries

The main tool that we use for lower bounding the size of our circuits for computing MOD_q is the so-called ϵ-Discriminator Lemma, introduced by Hajnal et.al.[19]. We state here a specialized version of it that is particularly convenient for our work, and has been also used in earlier works (see for example [6], [8]).

Lemma 3 (Discriminator Lemma) *Let C be a circuit that has a MAJORITY gate at its output that is being fed by t sub-circuits $C_1, \ldots, C_t$. If C computes MOD_q, then there exists a sub-circuit C_i, such that $Corr\big(C_i, \mathrm{MOD}_q\big) \geq 1/t$.*

The way it is useful for us is the following: let $\mathrm{e}_q(y)$ represent the function that is obtained by raising the q-th primitive root of unity to its yth power, i.e. $\mathrm{e}_q(y) = \exp\big(2\pi i y/q\big)$, where i is the non-trivial square-root of unity. Recall the

following elementary fact: for any integer y, the expression $(1/q)\sum_{a=0}^{q-1} \mathrm{e}_q(ay)$ evaluates to 1 if $y \equiv 0 \pmod q$, and otherwise evaluates to 0. This gives rise to the following useful fact: $\mathrm{MOD}_q^{\{b\}} = (1/q)\sum_{a=0}^{q-1} \mathrm{e}_q\big(a(\sum_i x_i - b)\big)$. For any $f : \{0,1\}^n \to \{0,1\}$ and $a \in \mathbb{Z}_q$, let $S(f,q,a) = \mathbb{E}_x\big[f(x)\mathrm{e}_q(a\sum_i x_i)\big]$. Then, the above identities can be easily made to yield (see for example [6], [8])

$$\mathrm{Corr}\big(f, \mathrm{MOD}_q\big) \leq$$
$$2q \cdot \max\big\{S(f,q,a) \,\big|\, a \in \mathbb{Z}_q, a \neq 0\big\} + 2^{-\Omega(n)} \qquad (3)$$

The Discriminator Lemma along with (3), immediately sets us a target of obtaining exponentially small upper bounds for the quantity $S(f,q,a)$, where f is the output of a circuit like $\mathrm{AND} \circ \mathrm{MOD}_m^A$. Such a bound would yield desired exponential lower bounds on the top fan-in of $\mathrm{MAJ} \circ \mathrm{AND} \circ \mathrm{MOD}_m^A$, for computing MOD_q. The key to obtaining such upper bounds on $S(f,q,a)$ will be the usage of exponential sums.

We will need estimates of exponential sums that were first obtained in the breakthrough work of Bourgain [6] and refined progressively in further works [14], [30], [7]. We state the most refined estimate below:

Theorem 4 ([7]) *Let m, q be two fixed positive co-prime integers and let P be any n-variate multilinear polynomial of degree d with coefficients in $\mathbb{Z}_m$. Then, there exists a constant $\beta = \beta(m,q)$ such that the following holds:*

$$\left|\mathbb{E}_{x\in\{0,1\}^n} \mathrm{e}_m\big(P(x)\big)\mathrm{e}_q\Big(\sum_i x_i\Big)\right| \leq \exp(-\beta^d n). \qquad (4)$$

We point out that the above estimate fails to give anything non-trivial when the degree d of the polynomial P is more than $\log n$. Finding exponentially small upper bounds for the exponential sum in (4) for $d > \log n$, even when m is prime, remains a very interesting open problem.

3. Low rigid-rank systems of equations

This section will deal with systems of equations which have low rigid-rank, a notion we will define below. We start this section with four subsections dealing with special cases, allowing us to introduce technical background, develop some necessary machinery, and motivate the definition and use of rigid-rank in the final subsection.

3.1. MOD_m *gates with a singleton accepting set*

In this subsection, we prove a simple exponential sum, for systems of equations in which the modular gates have an accepting set of size 1. Without loss of generality, such gates have accepting set $\{0\}$. This will yield a correlation bound that yields an alternative proof to the main result of Beigel-Maciel [5].

Lemma 5 *Let m be any positive integer, and $C = \mathrm{AND} \circ \mathrm{MOD}_m^{\{0\}}$ be any depth-two circuit. Then the correlation of C with MOD_q is $\exp\big(-\alpha n\big)$ for some constant $0 < \alpha(m,q) < 1$, when m, q are co-prime.*

Proof: Let ℓ_i be the linear form associated with the ith MOD_m gate at the base of C and the fan-in of the output AND gate be t. Then,

$$S(C,q,b) =$$
$$\mathbb{E}_{x\in\{0,1\}^n}\left[\left\{\prod_{j=1}^{t}\left(\frac{1}{m}\sum_{a=0}^{m-1}\mathrm{e}_m(a\ell_i(x))\right)\right\}\mathrm{e}_q\left(b\sum_{i=1}^{n}x_i\right)\right]$$

Expanding the product of sums into a sum of products along with the linearity of expectation yields

$$\frac{1}{m^t}\sum_{j=1}^{m^t}\mathbb{E}_{x\in\{0,1\}^n}\Big[\mathrm{e}_m(r_j(x))\mathrm{e}_q\big(b(x_1+\cdots+x_n)\big)\Big] \qquad (5)$$

where, each r_j is a linear polynomial obtained by a $\mathbb{Z}_m$-linear combination of ℓ_i's. Writing $r_j(x) = a_{j,1}x_1 + \cdots + a_{j,n}x_n$, we can separate variables and obtain

$$\left|\mathbb{E}_{x\in\{0,1\}^n}\Big[\mathrm{e}_m\big(r_j(x)\big)\mathrm{e}_q\big(b(x_1+\cdots+x_n)\big)\Big]\right|$$
$$= \prod_{i=1}^{n}\left|\mathbb{E}_{x_i\in\{0,1\}}\Big[\mathrm{e}_m\big(a_{j,i}x_i\big)\mathrm{e}_q\big(bx_i\big)\Big]\right| \leq \exp\big(-\alpha n\big)$$

for some $0 < \alpha < 1$, where the last inequality follows from the simple fact that every term in the product is bounded away from 1 in absolute value . Thus, using triangle inequality, we get $S(f,q,b) \leq 2^{-\alpha n}$ for all $b \in \mathbb{Z}_q$. Applying (3) with the Discriminator Lemma proves our lemma. ∎

First observe that the proof works with any singleton accepting set, not just the set $A = \{0\}$, simply by adding the affine shift in the exponential sum. Further, note that the above bound already yields exponential lower bounds for depth-three circuits of type $\mathrm{MAJ} \circ \mathrm{AND} \circ \mathrm{MOD}_m^{\{0\}}$. As mentioned above, such a bound was obtained by Beigel and Maciel [5], through different techniques. The advantage of using our technique is that in tandem with powerful estimates by Bourgain [6] of exponential sums involving low-degree polynomials, our argument yields the following significantly stronger result for depth-four circuits.

Theorem 6 *Let C be a depth-four circuit of type $\mathrm{MAJ} \circ \mathrm{AND} \circ \mathrm{MOD}_m^{\{0\}} \circ \mathrm{AND}_{o(\log n)}$, where the bottom AND gates have fan-in of $o(\log n)$. If C computes MOD_q, then the top fan-in of C must be $2^{\Omega(n)}$, when m, q are co-prime fixed integers.*

We do not give a formal proof of this theorem here, but point out that it follows in a very similar way as the proof of Lemma 5, where instead of exponentiating linear polynomials we exponentiate degree d polynomials over $\mathbb{Z}_m$

if the fan-in of the bottom AND gates are at most d. In the step that is analogous to (5) in the proof above, we get exponential sums of the type in (4). Plugging their bounds from Theorem 4, yields Theorem 6.

As suggested by Beigel and Maciel, a natural next step is to tackle the problem of obtaining strong lower bounds for depth-three circuits with generalized modular gates at the bottom, i.e. circuits of type $\mathrm{MAJ} \circ \mathrm{AND} \circ \mathrm{MOD}_m^A$, where A is an arbitrary subset of $\mathbb{Z}_m$. Let us see what happens when we try to apply the same method for such gates.

3.2. Generalized MOD_m^A gates and systems with few equations

The upshot of this subsection is that the above argument can be extended to general modular gates of the form $\mathrm{MOD}_m^A(z_1, \cdots, z_k))$ output 1 iff $z_1 + \cdots + z_k \in A (\mathrm{mod}\ m)$, as long as the number of such gates is small. The lemma below essentially appears in [8].

Lemma 7 *Let C be a circuit of type $\mathrm{AND} \circ$ MOD$_m^A$, with top fan-in t. Then, Corr$(C, \mathrm{MOD}_q) \leq (m-1)^t 2^{-\alpha n}$ for some constant $0 < \alpha = \alpha(m, q) < 1$.*

Proof: The proof simply reduces to Lemma 5. Note that the output of every MOD_m^A gate can be expressed as a sum of at most $m-1$ simple MOD_m gates with singleton accepting sets, one for each element of A, since at most one of them can be 1 for any fixed input x. Expressing every modular gate in C this way, we get that $C = \sum_{j=1}^{m^t} C_j$, where each $C_j(x)$ tests if x satisfies the linear system $\ell_i(x) = a_{j,i}$ for all t, where each of the constants $a_{j,i}$ is an element of A_i. Note that for every input x at most one of the $C_j(x)$ can output 1. This, linearity of expectation and triangle inequality allow us to derive directly $\mathrm{Corr}(C, \mathrm{MOD}_q) \leq \sum_{j=1}^{m^t} \mathrm{Corr}(C_j, \mathrm{MOD}_q) \leq (m-1)^t 2^{-\alpha n}$. ∎

This bound is useful only for fan-in $t \leq \delta n$, for some constant δ. Hence, it cannot provide a super-linear lower bound. We next show that this is possible if the system of equations has low rank.

3.3. Low-rank systems

We note that henceforth, unless otherwise stated, we consider generalized modular gates with arbitrary accepting sets. Thus, we may assume, w.l.o.g, that the linear forms associated with all MOD_m gates are homogeneous. The *rank* of C, denoted by $\mathrm{rank}(C)$, is defined as the size of a minimal subset S of the set of its underlying linear forms, such that every linear form of C is generated by a $\mathbb{Z}_m$-linear combination of the forms in S.

Lemma 8 *Let C be a circuit of type AND $\circ$ MOD$_m^A$. Then, Corr$(C, \mathrm{MOD}_q) \leq (m-1)^{rank(C)} 2^{-\alpha n}$ for some constant $0 < \alpha = \alpha(m, q) < 1$.*

Proof: Let $\ell_1, \ldots, \ell_t$ be the linear forms in our circuit. Let $r = \mathrm{rank}(C)$, and assume w.l.o.g., that $\ell_1, \ldots, \ell_r$ span the remaining $t-r$ forms. Now we can write $C = \sum_{j \in J} C_j$ going over all possible r-tuples of values of the singletons composing A_i for $i \leq r$, and keeping only those tuples for which satisfying these r equations implies satisfying the remaining $t-r$ equations determined by them. Thus $|J| \leq (m-1)^r$ and we conclude as in the proof of Lemma 7. ∎

We will now see that, using another idea, we can handle more general situations than just low-rank systems. For this we take a detour to a different restriction on our gates.

3.4. Sparse MOD_m gates

Here we handle generalized modular gates with few inputs. Let us call a linear form k-sparse if the number of non-zero coefficients appearing in it is at most k. A mod gate is called k-sparse if the associated linear form is k-sparse. We show that AND of sparse gates has small correlation with MOD_q.

Lemma 9 *Let C be a $G \circ \mathrm{MOD}_m^A$ circuit in which each bottom gate is k-sparse. Then, Corr$\big(C, \mathrm{MOD}_q\big) \leq exp(-\beta^k n)$.*

Proof: Consider any MOD_m^A gate at the base. As it is computing a boolean function of at most k-variables, there is a polynomial of degree at most k over $\mathbb{Z}_m$ that *exactly* represents the output of the gate. Let $P_1, \ldots, P_t$ be these polynomials for the t gates at the bottom. Then, one can write the following:

$$S(C, q, b) = \left| \mathbb{E}_{x \in \{0,1\}^n} \left[\left(\prod_{j=1}^{t} \left(\frac{1}{m} \sum_{a=0}^{m-1} \mathrm{e}_m(aP_i(x)) \right) \right) \mathrm{e}_q\left(b \sum_{i=1}^{n} x_i \right) \right] \right| \tag{6}$$

Mimicking the argument as in the proof of Lemma 5, one obtains the following upper bound on $S(C, q, b)$:

$$\frac{1}{m^t} \sum_{j=1}^{m^t} \left| \mathbb{E}_{x \in \{0,1\}^n} \left[\mathrm{e}_m(s_j(x)) \mathrm{e}_q\big(b(x_1 + \cdots + x_n)\big) \right] \right| \tag{7}$$

where, each s_j is a polynomial of degree at most k obtained by a $\mathbb{Z}_m$-linear combination of P_i's. Applying in the estimate given by (4) to the RHS of (7), proves our result. ∎

3.5. Low rigid-rank

We now combine Lemma 9 and Lemma 8 in the following way, to show that we can handle systems of equations which can be made low rank after a sparse change to each equation. This is inspired by Valiant's famous notion of rigidity [28], [29], used to attack (so far unsuccessfully) size-depth tradeoffs for computing linear systems over fields. We use the following definition:

A depth-two circuit of type $\mathrm{AND} \circ \mathrm{MOD}_m^A$ is called (k, r)-sparse if its associated linear forms $\ell_1, \ldots, \ell_t$ satisfy the

following property: each ℓ_i can be written as $\ell'_i + L_i$ such that the set $\{L_i | 1 \leq i \leq t\}$ has rank r and every ℓ'_i is k-sparse.

Lemma 10 *Let C be a depth-two circuit of type $\mathrm{AND} \circ \mathrm{MOD}_m^A$ that is (k,r)-sparse. Then, $Corr\big(C, \mathrm{MOD}_q\big) \leq m^r exp(-\beta^k n)$*

Proof: As before, we look at the possible evaluations of the various linear forms. Let t be the top fan-in, and let $\ell_i = \ell'_i + L_i$. Wlog, assume that $L_1, \ldots, L_r$ are the linearly independent forms that span every other L_i. Then our idea is to split the sum into at most m^r different ones, corresponding to the possible evaluations of $L_1, \ldots, L_r$. Let u be any such evaluation, where u_i represents the evaluation of L_i. Given u, we know what each L_i evaluates to in $\mathbb{Z}_m$, for all $i \leq t$. Hence, we know the set of values in $\mathbb{Z}_m$, denoted by A_i^u, that ℓ'_i could evaluate to so that ℓ_i evaluates to some element in A_i. In other words, $A_i^u = \{a \in \mathbb{Z}_m | a + u_i \in A_i\}$. Since, ℓ'_i depends on at most k variables, there exists a multilinear polynomial P_i^u over $\mathbb{Z}_m$ of degree at most k such that $P_i^u(x) = 0 \pmod m$ iff $\ell'_i(x) \in A_i^u$. These observations allow us to write $S\big(C, q, b\big)$ as the following:

$$\Big| \sum_{u \in [m]^r} \mathbb{E}_x \Big[\Big(\prod_{j=1}^{r} \frac{1}{m} \sum_{a=0}^{m-1} \mathrm{e}_m\big(a(L_j(x) - u_1)\big) \Big) \times \\ \times \Big(\prod_{i=1}^{t} \frac{1}{m} \sum_{a=0}^{m-1} \mathrm{e}_m\big(aP_i^u(x)\big) \Big) \mathrm{e}_q\big(b \sum_{i=1}^{n} x_i\big) \Big] \Big|$$

Expanding out the product of sums into sum of products yields the following upper bound on $S(C, q, b)$,

$$\sum_{u \in [m]^r} \frac{1}{m^{r+t}} \sum_{i=1}^{m^r} \sum_{j=1}^{m^t} \Big| \mathbb{E}_x \Big[\mathrm{e}_m\big(R_i^u(x) + Q_j^u(x)\big) \times \\ \times \mathrm{e}_q\big(b \sum_{i=1}^{n} x_i\big) \Big] \Big|,$$

where each $Q_j^u(x)$ is a polynomial of degree at most k obtained by a $\mathbb{Z}_m$-linear combination of the t polynomials $P_1^u, \ldots, P_t^u$, and each R_i^u is a linear polynomial obtained by the ith $\mathbb{Z}_m$-linear combination of the L_i's. Thus, applying the bounds from (4), we are done. ∎

As it happens, even low rigid-rank will not suffice, and we now generalize this result further in the next section.

3.6. Low rigid-rank in one prime factor of m

Let $m = p_1 p_2$ with both p_i prime. We now show how to bound the exponential sum even if the given linear system is (k,r)-sparse only modulo one of them.

Let C be a depth-two circuit of type $G \circ \mathrm{MOD}_m^A$, with top fan-in t. Then, let $\mathcal{L} = \{\ell_1, \ldots, \ell_t\}$ be the set of the associated linear forms over $\mathbb{Z}_m$. Using Chinese remaindering, let $\mathcal{L}^{p_1} = \{\ell_1^{p_1}, \ldots, \ell_t^{p_1}\}$ and $\mathcal{L}^{p_2} = \{\ell_1^{p_2}, \ldots, \ell_t^{p_2}\}$ be respectively the corresponding set of linear forms over $\mathbb{Z}_{p_1}$ and $\mathbb{Z}_{p_2}$.

Lemma 11 *If $\mathcal{L}^{p_1}$ (or $\mathcal{L}^{p_2}$) is (k,r)-sparse, then, $Corr\big(C, \mathrm{MOD}_q\big) \leq m^r exp(-\beta^{k+m} n)$, when m, q are co-prime and $\beta = \beta(m,q) > 0$.*

Proof: Assume $\mathcal{L}^{P_1}$ is (k,r)-sparse. For each $1 \leq i \leq t$, let $\ell_i^{p_1} = \ell'_i + L_i$, be the decomposition, over $\mathbb{Z}_{p_1}$, of linear forms such that each ℓ'_i has k variables with non-zero coefficients. Wlog, let $L_1, \ldots, L_r$ span all the other L_i's. The idea is we split the correlation into $(p_1)^r$ different sums corresponding to the $(p_1)^r$ possible evaluations of $L_1, \ldots, L_r$. Let u be any such evaluation. Then for every i, $1 \leq i \leq t$ we know what L_i evaluates to. There are thus at most $|A_i|$ possible evaluations of ℓ'_i over $\mathbb{Z}_{p_1}$ that will keep $\ell_i^{p_1}$ evaluate to something that is admissible in the accepting set. For each such evaluation of ℓ'_i, we know the set of admissible evaluations of $\ell_i^{p_2}$ over $\mathbb{Z}_{p_2}$. It is simple to verify that the characteristic function of the set of points of the cube which result in admissible evaluations for ℓ'_i and $\ell_i^{p_2}$ can be exactly represented over $\mathbb{Z}_{p_2}$ by a polynomial of degree at most $k + p_2 - 1$. We call this polynomial, as before, P_i^u. Finding this polynomial allows us to carry on our calculations in an identical way as in the proof of Lemma 10. Thus, $S(C, q, b)$ can be written as follows:

$$\Big| \sum_{u \in [p_1]^r} \mathbb{E}_x \Big[\Big(\prod_{j=1}^{r} \frac{1}{p_1} \sum_{a=0}^{p_1-1} \mathrm{e}_{p_1}\big(a(L_j(x) - u_1)\big) \Big) \times \\ \times \Big(\prod_{i=1}^{t} \frac{1}{p_2} \sum_{a=0}^{p_2-1} \mathrm{e}_{p_2}\big(aP_i^u(x)\big) \Big) \mathrm{e}_q\big(b \sum_{i=1}^{n} x_i\big) \Big] \Big|$$

Expanding out the product of sums into sum of products yields the following bound:

$$\sum_{u \in [p_1]^r} \frac{1}{p_1^r p_2^t} \sum_{i=1}^{p_1^r} \sum_{j=1}^{p_2^t} \Big| \mathbb{E}_x \Big[\mathrm{e}_{p_1}\big(R_i^u(x)\big) \mathrm{e}_{p_2}\big(Q_j^u(x)\big) \times \\ \times \mathrm{e}_q\big(b \sum_{i=1}^{n} x_i\big) \Big] \Big|,$$

where each $Q_j^u(x)$ is a polynomial of degree at most $k + p_2 - 1$ obtained by a $\mathbb{Z}_{p_2}$-linear combination of the t polynomials $P_1^u, \ldots, P_t^u$, and each R_i^u is a linear polynomial obtained by the ith $\mathbb{Z}_{p_1}$-linear combination of the L_i's. Note that using Chinese remaindering, one can combine $Q_j^u(x)$ and R_i^u to a polynomial over $\mathbb{Z}_m$ of degree at most $k + p_2 - 1$. Thus, applying the bounds from (4), we are done. ∎

Finally, we observe that the lemma above can be generalized to systems which can be decomposed into a few subsystems, each sparse modulo one of the prime factors of m. It is this generalization that will be needed for proving our Main Lemma. Let S_{p_1}, S_{p_2} be a partition of $[t]$, such that the set of linear forms, over Z_{p_i}, indexed by elements

of S_i is (k_i, r_i)-sparse (over Z_{p^i}), where each p^i is either p_1 or p_2. Further, assume $|S_i| = s_i$. Then,

Lemma 12 *Let C be a circuit whose underlying linear forms admit a partition into sets S_{p_1}, S_{p_2} as described above. Then, if m, q are co-prime and $m = p_1 p_2$,*

$$Corr\big(C, \mathrm{MOD}_q\big) \leq m^r exp\Big(- \frac{\beta n}{\big(m2^{m-1}\big)^{k+m-1}}\Big) = m^r exp\big(- \beta_0(m,q,k)n\big),$$

where $r = r_1 + r_2$ and $k = max\{k_1, k_2\}$.

The proof of the lemma above is a natural adaptation of the argument used for proving Lemma 11. We leave it for the full version of our paper.

4. A SET OF GATES HAVING HIGH RIGID RANK

Next, we extend a result from the work of Grigoriev and Razborov [15] about a set of linear forms. The overall plan is to show that in this case, the probability that a random Boolean input will satisfy any such system is exponentially small in n, independent of the number of equations. Naturally, this is what one expects over a field, and when the inputs are chosen randomly from that field. We work over a ring, and the inputs are only Boolean. Nevertheless, notions of rank introduced in [15] naturally extend to yield the result, as well as complement the low rigid rank case we used in the previous section.

Given a set $\mathcal{L}$ of t linear forms in n variables over $\mathbb{Z}_m$, we identify them in the natural way with a $t \times n$ matrix denoted by $A(\mathcal{L})$. When clear from the context, we simply denote the matrix by A. Define the k-rigid rank of this matrix over $\mathbb{Z}_p$, denoted by $\mathrm{rrank}_k^p\big(A\big)$, as the rank (over $\mathbb{Z}_p$) of a minimal rank matrix that differs from A in at most k entries per row. Let $m = p_1 p_2 \cdots p_s$ be a product of s distinct prime numbers. The k-communication rank of A, over $\mathbb{Z}_m$, denoted by $\mathrm{ccrank}_k^m\big(A\big)$ is the maximum number r such that there exists a subset I of the rows of A satisfying the following: a) $|I| = r$, and b) For each $1 \leq i \leq s$, there exists k pairwise disjoint subsets $J_1^i, \ldots, J_k^i$ of the columns of A, each of size r, such that every sub-matrix $A_{\{I, J_j^i\}}$ has rank r, i.e. has full rank, over $\mathbb{Z}_{p_i}$. The notions of rigid rank and communication rank are related. If a matrix A has high rigid rank over every $\mathbb{Z}_{p_i}$, then we expect that the rank is well distributed over the columns in the sense that several disjoint sub-matrices of A should have high rank. This intuition is captured by the following lemma.

Lemma 13 (extension of Lemma 3.3 of [15]) *Let $m = p_1 p_2 \cdots p_s$ be any number, where each p_i is a distinct prime. Let A be any $t \times n$ matrix with entries in $\mathbb{Z}_m$ such that the $ccrank_k^m(A) = r$. Then, the rows of A can be partitioned into s sets $I_1, \ldots, I_s$, such that the sk-rigid rank of $A_{\{I_i, [n]\}}$ over $\mathbb{Z}_{p_i}$ is at most $(sk+1)r$.*

Proof: From the assumption on A, there exists s pairwise disjoint families of subsets of columns of A, denoted by $\mathcal{J}^1, \ldots, \mathcal{J}^s$, where each $\mathcal{J}^i = \{J_1^i, \ldots, J_k^i\}$ contains k pairwise disjoint sets of columns, each of size r. Further, there exists a set of rows I, such that each sub-matrix $A_{\{I, J_j^i\}}$ has full rank over $\mathbb{Z}_{p_i}$, i.e. has rank r. Trying to enlarge r (which we can't, since r is maximal) we derive some structure leading to bound on rigid-rank.

Consider any row ρ that is not in I. For each J_j^i, notice that the vector $A_{\{\rho, J_j^i\}}$ modulo p_i is obtained by a unique linear combination of the rows of $A_{\{I, J_j^i\}}$. Label this linear combination θ_j^i. Define a set of columns $J_j^i[\rho]$ in the following way: a column c *outside* of $\cup_{i=1}\mathcal{J}^i$ is in $J_j^i[\rho]$ precisely if the linear combination θ_j^i when applied to the elements of the vector $A_{\{I,c\}}$ *fails* to produce the element $A_{\{\rho,c\}}$ modulo p_i.

The first thing to observe is that for each such row ρ, there exists i, j such that $|J_j^i[\rho]| \leq sk$. Otherwise, for all i, j we have $|J_j^i[\rho]| \geq sk+1$. Then, it is easy to verify, that we can add one distinct element from each $J_j^i[\rho]$ to J_j^i and add ρ to I, certifying that $\mathrm{ccrank}_k^m(A)$ is at least $r+1$. This yields a contradiction. Hence, for each row ρ there exists an i_ρ, and a j_ρ, such that $|J_{j_\rho}^{i_\rho}| \leq sk$.

Let $I_a = \{\rho \notin I | i_\rho = a\}$, for all $1 \leq a \leq s$. We now show that the sk-rigid rank of any such I_a is at most $(sk+1)r$. Consider any $\rho \in I_a$. We change the at most sk elements of $J_{j_\rho}^{i_\rho}$ to agree with each coordinate of the vector generated by the linear combination $\theta_{j_\rho}^{i_\rho}$ of the rows of the sub-matrix $A_{\{I, J_{j_\rho}^{i_\rho}\}}$ modulo p_j. Let the modified row be ρ'. We finish the argument by showing that the rank of the set of slightly perturbed rows $\mathcal{L}_a = \{\rho' | \rho \in I_a\}$ is at most $(sk+1)r$ over $\mathbb{Z}_{p_j}$. Define $\delta_{j,t}^i$ to be the unit n dimensional vector that has a zero in every co-ordinate, except the co-ordinate that corresponds to the tth column of the set J_j^i where it has a 1. Then, it is simple to verify that the set of vectors $I \cup \{\delta_{j,t}^i | 1 \leq i \leq s; 1 \leq j \leq k; 1 \leq t \leq r\}$ generates every vector in $\mathcal{L}_a$, proving that the sk-rigid rank of I_a is at most $skr + r$.

Finally, we note that we could add the rows in I to any one of the set I_a, without increasing the sk-rigid rank of the resulting system beyond $(sk+1)r$. ∎

The lemma above yields the following convenient dichotomy: when m is a product of two distinct primes, either the given set of $\mathrm{AND} \circ \mathrm{MOD}_m^A$ sub-circuits can be partitioned into two sets of sub-circuits, each of which gives rise to a linear system with low rigid-rank w.r.t. one prime, OR, the m-communication rank of the entire system is large. The former case was handled in the last section. The latter is handled below by extending a result of Grigoriev and Razborov [15]. They worked with linear forms over finite fields and we, extend it to forms over finite rings of the form $\mathbb{Z}_m$. Before we proceed with this extension, we need a detour into arithmetic combinatorics of sumsets over these

rings.

4.1. Sumsets over $\mathbb{Z}_m$

Let A, B be subsets of any group. Then, the *sumset* $A+B$ is defined as the set $\{c = a_i + b_i \,|\, a_i \in A,\, b_i \in B\}$.

When the underlying group is $\mathbb{Z}_p$, p prime, then the famous Cauchy-Davenport lemma states that the sumset always grows (if it has room to grow), more specifically that $|A+B| \geq \min\{|A|+|B|-1, p\}$. Thus adding enough subsets would cover the whole group $\mathbb{Z}_p$. This fails, of course, over rings $\mathbb{Z}_m$ when m is composite, due to the existence of subrings. However, in this subsection we show that a weaker statement still holds for these rings. This can be roughly stated ad follows: adding sufficiently many 2-sets, each pair differing modulo one of the divisors of m, will eventually generate $\mathbb{Z}_m$.

Lemma 14 *Let $m = p_1 \cdots p_s$ be a product of s distinct primes. For each $1 \leq i \leq s$, consider m subsets $A^i_1, \ldots, A^i_m \subset \mathbb{Z}_m$, each $|A^i_j| = 2$, such that the mod p_i component of the two elements of A^i_j are different. Then, $\sum_{i=1}^{s} \sum_{j=1}^{m} A^i_j = \mathbb{Z}_m$.*

For any set A and element $a \in \mathbb{Z}_m$, let A_a represent the translate of A by a, i.e. $A_a = A+\{a\}$. The following simple observation is useful for estimating the size of sumsets:

Observation 15 *Let $A, B \subseteq \mathbb{Z}_m$. Then, for any $a, b \in \mathbb{Z}_m$, $|A_a + B_b| = |A + B|$.*

Proof: (*of Lemma 14*) We prove by induction of s that $|\sum_{i=1}^{s} \sum_{j=1}^{m} A^i_j| = p_1 p_2 \cdots p_s$. This is clearly equivalent to Fact 14. For $s = 0$, this is vacuously true. Assume, by Inductive hypothesis, it is true for $s = t$. Let $m' = p_1 \cdots p_t$. View $B, A^{t+1}_1, \ldots, A^{t+1}_m$ as subsets of $\mathbb{Z}_{m'} \times \mathbb{Z}_{p_{t+1}}$. Then, the hypothesis implies, via Chinese remaindering, that the sumset $B = \sum_{i=1}^{t} \sum_{j=1}^{m} A^i_j$ has the following property: for each element $a \in \mathbb{Z}_{m'}$, there is an element (a, b) in B, with $b \in \mathbb{Z}_{p_{t+1}}$. We show that for every set B satisfying this property, the following holds: $|B+A^{t+1}_j| = \min\{|B|+1, m\}$. To show that, using Observation 15 about translates, assume w.l.o.g. that $A^{t+1}_j = \{(0,0), (a,b)\}$, for some $a \in \mathbb{Z}_{m'}$ and non-zero $b \in \mathbb{Z}_{p_{t+1}}$.

If $B = \mathbb{Z}_{m'} \times \mathbb{Z}_{p_{t+1}}$, then we have nothing to prove. Otherwise, there exists some $c \in Z_{p_{t+1}}$ and $a' \in \mathbb{Z}_{m'}$ such that (a', c) does not occur in B. By inductive property of B, there exists $c' \in \mathbb{Z}_{p_{t+1}}$ such that (a', c') does occur in B. Let $c = c' + d$ for $d \in Z_{p_{t+1}}$. As m', p_{t+1} are co-prime, there exists an integer k such that $k \equiv 0 \bmod(m')$ and $k \equiv b^{-1}d \bmod(p_{t+1})$. Now if $B + (a,b) \neq B$, we are done. Otherwise, $B = B + (a,b) = B + k(a,b)$. But, (a', c) occurs in $B + k(a,b)$, yielding a contradiction. Thus, in this case, $B + (a,b) \neq B$. Hence, $|B + A^{t+1}_j| \geq |B| + 1$ and we complete the induction. ∎

Lemma 14 is used next to prove the main result of this section.

4.2. The correlation bound

We are now ready to state the main result of this section.

Lemma 16 (extension of Lemma 4.1 of [15]) *Let $\mathcal{L} = \{\ell_1, \ldots, \ell_t\}$ be a system of t linear forms, in n variables, over $\mathbb{Z}_m$, where m is a fixed positive integer. Let $A(\mathcal{L})$ be the associated $t \times n$ matrix, with entries from $\mathbb{Z}_m$. If $r = ccrank^m_m\big(A(\mathcal{L})\big)$, then*

$$\Pr_{x \in_R \{0,1\}^n} \left[\bigwedge_{i=1}^{t} \ell_i(x) \in A_i \right] \leq exp\big(-\Omega(r)\big),$$

where each $A_i \subsetneq \mathbb{Z}_m$ is an arbitrary set.

Proof: The argument follows closely the one given in Grigoriev and Razborov [15]. From the definition of communication rank, we get a set of rows I, with $|I| = r$ such that there are m pairwise disjoint sets of columns, each of size r and denoted by $J^i_1, \ldots, J^i_m$, $1 \leq i \leq s$ and $A(\mathcal{L}^{p_i})_{\{I, J^i_j\}}$ has full rank, i.e. has rank r modulo p_i for every i. Put $d = ms$. It is convenient for us, here, to consider an enumeration of the above d disjoint sets of columns. With a slight abuse of notation, we denote this enumeration as $J_1, \ldots, J_d$. The relevant prime that comes into play, when considering the rank of the sub-matrix $A(L)_{\{I, J_j\}}$, is denoted by $p[j]$.

Let $x^1, \ldots, x^d$ be random variables representing boolean assignments to variables indexed by these sets of columns. Each x^i is thus a boolean vector of length r. Let us focus our attention only to linear forms corresponding to rows indexed by elements of I. We show below the following stronger bound that clearly implies the result we want to prove: let $\mathcal{J} = \cup_{i=1}^{d} J_i$. Let $\ell_i(x^1, \ldots, x^d)$ denote the linear form that is obtained by retaining in ℓ_i just the terms that correspond to variables indexed by sets in $\mathcal{J}$. Then, the following holds:

$$\Pr_{x^1, \ldots, x^d} \left[\bigwedge_{i \in I} \ell_i(x^1, \ldots, x^d) \in A_i \right] \leq \exp\big(-\Omega(r)\big),$$

where each $A_i \subsetneq \mathbb{Z}_m$ is an arbitrary subset.

Instead of showing this directly, we do the following: let $h(x^1, \ldots, x^d)$ be a random indicator variable that outputs 1 if $\ell_i(x^1, \ldots, x^d) \in A_i$, for all $i \in I$, and otherwise outputs 0. Then, (by now) a routine use of d-repeated applications of Cauchy-Schwartz inequality yields

$$\begin{aligned}
&\left(Pr_{x^1, \ldots, x^d} \left[\bigwedge_{i \in I} \ell_i(x^1, \ldots, x^d) \in A_i \right] \right)^{2^d} \\
&= \left(\mathbb{E}_{x^1, \ldots, x^d} \big[h(x^1, \ldots, x^d) \big] \right)^{2^d} \\
&\leq \mathbb{E}_{x^j_0, x^j_1;\, 1 \leq j \leq d} \left[\prod_{u \in \{0,1\}^d} h\big(x^1_{u_1} \ldots, x^d_{u_d}\big) \right] \qquad (8)
\end{aligned}$$

$$= \Pr_{x_0^1, x_1^1, \ldots, x_0^d, x_1^d} \left[\forall i \in I;\, \forall u \in \{0,1\}^d;\, \ell_i\left(x^1_{u_1}, \ldots, x^d_{u_d}\right) \in A_i \right] \tag{9}$$

The rest of the argument provides an exponentially small upper bound for the probability of the event (9), thereby proving our Lemma.

This probability is estimated in two steps. In the first step, we prove the following claim, which roughly asserts that every full rank block J^i_j is likely to provide us (when multiplied by a random Boolean vector) with a pair of distinct elements modulo p_i (to be later used in the sumset argument).

Claim: There exists constants γ and ν, such that with probability at least $1 - \exp(-\nu r)$, there exists a set of rows I' of size r/γ^d , such that for each $i \in I'$, we have $\ell_i(x_0^j) \neq \ell_i(x_1^j) \bmod(p[j])$, for all $1 \leq j \leq d$.

Before we prove this, let us see how, in the second step, it yields our desired result. We want to show that some equation fails to hold. Now if r is sufficiently large so that $r/\gamma^d \geq 1$, then Claim provides us at least one row k such that $\ell_k(x_0^{j'}) \neq \ell_k(x_1^{j'}) \bmod(p[j'])$, for $1 \leq j' \leq d$. Let J^i_j be the set that appears as $J_{j'}$ in our enumeration. Set $A^i_j = \{a_0, a_1\}$, where $a_0 = \ell_k(x_0^{j'})$ and $a_1 = \ell_k(x_1^{j'})$. Applying Lemma 14, we have that $\sum_{i=1}^{s} \sum_{j=1}^{m} A^i_j = \mathbb{Z}_m$. Hence, there exists a $u \in \{0,1\}^d$ such that $\ell_k\left(x^1_{u_1}, \ldots, x^d_{u_d}\right) \notin A_k$. The Claim shows that such a k exists with probability at least $(1 - \exp(-\nu r))$. The desired bound on the quantity in (9) follows.

Hence, all that remains is to prove the Claim. We do so by induction on d. Our inductive hypothesis is that the Claim holds for $d = u$. We show that it should then hold for $d = u + 1$. Consider I_u to be the set of all rows $k \in I$ for which every $\ell_k(x_0^j) \neq \ell_k(x_1^j) \pmod{p[j]}$, for all $j \leq u$. The inductive hypothesis is that $\Pr\left[|I_u| \geq r/\gamma^u\right] \geq \left(1 - \exp\left(\nu_u r\right)\right)$, for some constant ν_u. Let $p = p[u+1]$. Consider any fixed x_0^{u+1}. This fixes $y_0 = A(\mathcal{L}^p)_{\{I_u, J_{u+1}\}} \cdot x_0^{u+1}$, where $y_0 \in \mathbb{Z}_p^{|I_u|}$. We consider the set $\mathcal{V}$ of vectors in $\mathbb{Z}_p^{|I_u|}$ that differ from y_0 in at most $|I_u|/\gamma$ coordinates. Then,

$$|\mathcal{V}| = \sum_{j=0}^{|I_u|/\gamma} \binom{|I_u|}{j} (p-1)^j$$

We show

$$\forall v \in \mathcal{V};\ \Pr_{x_1^{u+1}} \left[A(\mathcal{L}^p)_{\{I_u, J_{u+1}\}} \cdot x_1^{u+1} = v \right] = 2^{-|I_u|}. \tag{10}$$

This is sufficient to complete the induction because, if γ is large enough, there exists a constant η such that $\Pr_{x_1^{u+1}} \left[A(\mathcal{L}^p)_{\{I_u, J_{u+1}\}} \cdot x_1^{u+1} \in \mathcal{V} \right] \leq \exp\left(-\eta r\right)$. In other words,

$$\Pr_{x_0^1, x_1^1, \ldots, x_0^{u+1}, x_1^{u+1}} \left[|I_{u+1}| \geq r/\gamma^{u+1} \right] \geq \left[1 - \exp\left(\nu_u r\right)\right]\left[1 - \exp\left(\eta r\right)\right] \geq 1 - \exp\left(-\nu_{u+1} r\right)$$

for an appropriately chosen constant ν_{u+1}. That proves the Claim, assuming (10).

We finish our argument by establishing (10). Since matrix $A(L^p)_{\{I_u, J_{u+1}\}}$ has rank $|I_u|$, we choose a subset of columns, denoted by $K \subseteq J_{u+1}$ so that $|K| = |I_u|$, such that the square matrix $A(L^p)_{\{I_u, K\}}$ has full rank over $\mathbb{Z}_p$. Let $x_1^{u+1}|_K$ be the projection of the vector x_1^{u+1} to coordinates indexed by K. Consider an arbitrary assignment to co-ordinates of x_1^{u+1} that do no not correspond to variables indexed by elements of K. For every such assignment and any $v \in \mathcal{V}$, there is exactly one vector in $\mathbb{Z}_p^{|I_u|}$ that $A(L^p)_{\{I_u, K\}} \cdot x_1^{u+1}|_K$ must evaluate to, so that $A(L^p)_{\{I_u, J_{u+1}\}} \cdot x_1^{u+1}$ evaluates to v. The full rank of $A(L^p)_{\{I_u, K\}}$, thus ensures that the probability of this happening is at most $2^{-|I_u|}$ and we are done. ■

5. Putting things together

In this section, we present the proof of our main lemma (Lemma 2) that depth-two circuits of type $G \circ \mathrm{MOD}_m^A$ have exponentially small correlation with MOD_q, when G is an AND or an OR gate and m contains exactly two distinct prime factors and is square-free.

Proof: (of Lemma 2) We assume that G is an AND gate in the argument below. The case when G is an OR gate is handled by using De-Morgan's law as follows: if f is the function computed by C, then $\neg f$ is computed by a depth-two circuit where the output gate is an AND gate and the base layer is the same as that of C with the accepting set of each MOD_m gate being the complement of what it was before. As the correlation of f with MOD_q is small iff the correlation of $\neg f$ and MOD_q is small, we are done by handling just the case when the output gate G is an AND gate.

Let $\mathcal{L}$ be the set of underlying linear forms, and $A(\mathcal{L})$ the associated matrix with entries from $\mathbb{Z}_m$. There are two cases to consider. First, assume that the m-communication rank of $A(\mathcal{L})$ is large, i.e. larger than αn for some constant $\alpha < 1$ to be set later. Then, Lemma 16 directly implies that the correlation of C and MOD_q is at most $\exp\left(-\delta(m,\alpha)n\right)$. In the other case, applying Lemma 13, we can partition $\mathcal{L}$ into two parts $\mathcal{L}^1$ and $\mathcal{L}^2$, such that each $\mathcal{L}^i$ has $2m$-rigid rank over $\mathbb{Z}_{p_i}$ at most $(2m+1)\alpha n$. Then, setting $k = 2m$, upper bounding both r_1 and r_2 by $(2m+1)\alpha n$, we apply Lemma 12 to obtain an upper bound of $\exp\left(-\{\beta_0(m,q,2m) - 2(2m+1)\alpha \log m\}n\right)$ on the correlation between C and MOD_q.

Thus, setting

$$\gamma(m,q) = \max_\alpha \left[\min\{\beta_0(m,q,2m) - 2(2m+1)\alpha \log m;\ \delta(m,\alpha)\}\right],$$

we see that the correlation is always at most $\exp(-\gamma n)$. ■

Using the Discriminator Lemma of Hajnal et.al [19], we get an exponential lower bound on depth-three circuits with generalized MOD_m gates at the base, that proves Theorem 1.

ACKNOWLEDGEMENT

We thank the anonymous referees for their helpful comments.

REFERENCES

[1] E. Allender, "A note on the power of threshold circuits," in *30th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, 1989, pp. 580–584.

[2] D. A. M. Barrington, "Some problems involving Razborov-Smolensky polynomials," in *Boolean function complexity*, ser. London Math.Soc.Lec.Note. Durham, 1990: Cambridge University Press, 1992, vol. 169, pp. 109–128.

[3] D. A. M. Barrington, R. Beigel, and S. Rudich, "Representing boolean functions as polynomials modulo composite numbers," *Computational Complexity*, vol. 4, pp. 367–382, 1994.

[4] D. A. M. Barrington and H. Straubing, "Lower bounds for modular counting by circuits with modular gates," *Computational Complexity*, vol. 8, no. 3, pp. 258–272, 1999.

[5] R. Beigel and A. Maciel, "Upper and lower bounds for some depth-3 circuit classes," *Computational Complexity*, vol. 6, no. 3, pp. 235–255, 1997.

[6] J. Bourgain, "Estimates of certain exponential sums arising in complexity theory," *C.R.Acad.Sci.Paris*, vol. Ser I 340, no. 9, pp. 627–631, 2005.

[7] A. Chattopadhyay, "Discrepancy and the power of bottom fan-in in depth-three circuits," in *48th Annual Symposium on Foundations of Computer Science (FOCS)*, 2007, pp. 449–458.

[8] A. Chattopadhyay, N. Goyal, P. Pudlák, and D. Thérien, "Lower bounds for circuits with MOD_m gates," in *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2006, pp. 709–718.

[9] A. Chattopadhyay and K. A. Hansen, "Lower bounds for circuits with few modular and symmetric gates," in *32nd International Colloquium on Automata, Languages and Programming (ICALP)*, 2005, pp. 994–1005.

[10] K. Efremenko, "3-query locally decodable codes of subexponential length," in *41st Annual Symposium on Theory of Computing (STOC)*, 2009, to appear.

[11] M. Goldmann, "A note on the power of Majority gates and modular gates," *Inf.Process.Lett.*, vol. 53, no. 6, pp. 321–327, 1995.

[12] F. Green, "Exponential sums and circuits with a single threshold gate and mod-gates," *Theory of Computing Systems*, vol. 32, pp. 453–466, 1999.

[13] ——, "The correlation between parity and quadratic polynomials mod 3," *J.Computer. Systems. Sciences*, vol. 69, no. 1, pp. 28–44, 2004.

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

[15] D. Grigoriev and A. A. Razborov, "Exponential lower bounds for depth-3 arithmetic circuits in algebras of functions over finite fields," *Applicable Algebra in Engineering, Communication and Computing*, vol. 10, no. 6, pp. 465–487, 2000.

[16] V. Grolmusz, "A weight-size trade-off for circuits and MOD m gates," in *26th Annual Symposium on Theory of Computing (STOC)*. ACM, 1994, pp. 68–74.

[17] ——, "Superpolynomial size set-systems with restricted intersections mod 6 and explicit ramsey graphs," *Combinatorica*, vol. 20, no. 1, pp. 71–86, 2000.

[18] V. Grolmusz and G. Tardos, "Lower bounds for (MOD-p-MOD-m) circuits," *SIAM J. Computing*, vol. 29, no. 4, pp. 1209–1222, 2000.

[19] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán, "Threshold circuits of bounded depth," *J.Computer.System.Sciences*, vol. 46, no. 2, pp. 129–154, 1993.

[20] K. A. Hansen, "On modular counting with polynomials," in *IEEE Conference on Computational Complexity*, 2006, pp. 202–212.

[21] K. A. Hansen and M. Koucký, "A new characterization of ACC^0 and probabilistic CC^0," in *The 24th Annual IEEE Conference on Computational Complexity*, 2009.

[22] M. Krause and P. Pudlák, "On the computational power of depth 2 circuits with threshold and modulo gates," in *26th Annual Symposium on Theory of Computing (STOC)*. ACM, 1994, pp. 48–57.

[23] A. A. Razborov, "Lower bounds on the size of bounded-depth networks over a complete basis with logical addition," in *Math. Notes of the Acad. of Sci. of USSR*, 1987, vol. 41, no. 4, pp. 333–338.

[24] R. Smolensky, "Algebraic methods in the theory of lower bounds for Boolean circuit complexity," in *19th Symposium on Theory of Computing (STOC)*, 1987, pp. 77–82.

[25] ——, "On interpolation by analytic functions with special properties and some weak lower bounds on the size of circuits with symmetric gates," in *31st Annual Symposium on Foundations of Computer Science (FOCS)*, 1990, pp. 628–631.

[26] H. Straubing and D. Thérien, "A note on MOD_p-MOD_m-circuits," *Theory of Computing Systems*, vol. 39, no. 5, pp. 699–706, 2006.

[27] D. Thérien, "Circuits constructed with MOD_q gates cannot compute "And" in sublinear size," *Computational Complexity*, vol. 4, pp. 383–388, 1994.

[28] L. Valiant, "Some conjectures relating to superlinear complexity," University of Leeds, Tech. Rep. 85, 1976.

[29] ——, "Graph-theoretic arguments in low-level complexity," in *The 6th Mathematical Foundations of Computer Science (MFCS)*, ser. LNCS, vol. 53, 1977, pp. 162–176.

[30] E. Viola and A. Wigderson, "Norms, XOR lemmas, and lower bounds for $gf(2)$ polynomials and multiparty protocols," *Theory of Computing*, vol. 4, pp. 137–168, 2008.