Multilinear Polynomials Modulo Composites

Arkadev Chattopadhyay[*]

Abstract

Understanding the power of constant-depth circuits that are allowed to use MOD_m gates, where m is an arbitrary but fixed positive integer, is a fundamental and inviting problem in theoretical computer science. Despite intensive efforts for more than twenty five years, this problem remains wide open.

In this column, we focus our attention on the related, but much simpler, model of computing a boolean function by multilinear polynomials over the ring $\mathbb{Z}_m$, when m is a composite number. As widely known, it is essential to understand this model in order to make progress with constant-depth circuits with MOD gates. We survey some recent results in this natural model that yield superpolynomial lower bounds on the size of some restricted circuits with MOD_m gates. The ingredients that get used in these results are perhaps more interesting. Some natural next steps emerge from these results that are also of independent mathematical interest. It is hoped that progress along these lines is feasible and would provide further insight into the general problem.

1 Introduction

Eric Allender [2] starts his recent survey of the state-of-affairs in proving lower bounds on circuit size by noting that his earlier survey [1] remains depressingly current. While it is true that we cannot pitifully find a function in EXP that cannot be computed by linear size depth-three circuits comprising only MOD_6 gates, the time honored George Polya principle of considering simpler problems seems to again provide ways to making meaningful progress. In this article, we further argue that such considerations

[*] University of Toronto, Department of Computer Science, arkadev@cs.toronto.edu.
The author is partially supported by a postdoctoral fellowship of the Natural Sciences and Engineering Research Council (NSERC) of Canada.

have raised (and sometimes solved) natural and appealing problems that can be stated in pure mathematical terms. This holds the promise that tools from mainstream mathematics can be further exploited in the context of understanding the computational power of mod counting.

A series of interesting works on constant-depth circuits have recently appeared. Here, we just focus on the ones that are motivated by circuits having MOD_m gates, where m is an arbitrary number. Note that MOD_m is a boolean function that is defined below:

Definition 1.1. Let $A \subseteq \mathbb{Z}_m$ be some *accepting* set. Then, for each $x \in \{0,1\}^n$, $\mathrm{MOD}_m^A(x) = 1$ if $\sum_{i=1}^n x_i \equiv a \pmod{m}$ for some $a \in A$, otherwise the function outputs 0.

By default, the accepting set A is $\mathbb{Z}_m - \{0\}$ and in this case it is dropped from the superscript. The class of functions computed by polynomial size and constant-depth circuits[1] of unbounded fan-in having AND, OR and MOD_m gates is called $\mathrm{ACC}^0[m]$. The union of these classes over all fixed positive integer m is defined to be the complexity class ACC^0. As is the convention, we overload these terms to also mean the underlying circuits with no restrictions on size. Understanding the computational limitations of ACC^0 is a major goal of computational complexity that remains unfulfilled.

Smolesnky[40] in the late eighties, building upon the elegant work of Razborov [39], showed that $\mathrm{ACC}^0[p^k]$ circuits require exponential size to compute MOD_q, if p is a prime and k any fixed positive number and q has a prime factor different from p. A simple exercise then shows that MAJORITY cannot be computed in sub-exponential size by such circuits. Indeed, one can very well imagine the excitement this generated back at the time. Smolensky made the following very tempting conjecture:

Conjecture 1.2 (Smolensky). *For any fixed positive integer m, $ACC^0[m]$ circuits needs exponential size to compute MOD_q, if m, q are co-prime numbers.*

At the moment, we seem to be far from proving (or disproving) Smolensky's conjecture. One may be inclined to think that circuits that are restricted to have *only* MOD_m gates (and constant depth, denoted by $\mathrm{CC}^0[m]$) are easier to deal with? Such a thought is especially appealing, given the following fact about prime moduli: for any prime p, circuits of constant-depth having only MOD_p gates cannot compute all functions. In particular, they cannot compute a *high degree* function (over $\mathbb{Z}_p$) like OR, AND and

[1] The input layer of all boolean circuits considered in this article have access to each variable and its negation, in addition to boolean constants 0 and 1.

MOD_q, *no matter how much size is allowed.* Indeed, this is a very strong computational limitation and follows surprisingly easily from the fact that $\mathbb{Z}_p^*$ is a group. In contrast, depth-2 such circuits having only MOD_m gates can compute everything:

Fact 1.3 (Folklore, (see [7])). *Let m be any number that has at least two distinct prime factors. Then, every n-variate boolean function f can be computed by a depth-two circuit of size 2^n having only MOD_m gates.*

In a recent work, Hansen and Koucký [32] observe that one can combine Fact 1.3 with the Razborov-Smolensky idea of approximating AND/OR gates by low degree polynomials over any finite field to yield the following interesting result:

Theorem 1.4 (implied in [32]). *Every quasipolynomial size circuit C comprising AND, OR and MOD_m gates of depth d can be approximated very well by a quasipolynomial size circuit C' of depth $O(d)$ comprising only MOD_m gates, i.e. $\Pr_x\left[C'(x) \neq C(x)\right] \leq 1/qpoly(n)$.*

This hints that proving Smolensky's conjecture for circuits with *only* MOD_m gates may be as hard as proving the general case. Indeed, Smolensky obtains his result for $\mathrm{ACC}^0[p]$ by showing the stronger result that they cannot even approximate well MOD_q. This strengthening is crucial to his argument. Theorem 1.4, on the other hand, shows that such a strengthened result (against MOD_q) for the special case of $\mathrm{CC}^0[m]$ circuits is sufficient to deal with general $\mathrm{ACC}^0[m]$ circuits.

Nevertheless, the intuition that $\mathrm{CC}^0[m]$ circuits are weaker and hence easier to deal with, may not be entirely lost. For a boolean function f, let the *support set* of f, denoted by $\mathrm{supp}(f)$, be the set of points in the cube where f evaluates to 1. The support set of a MOD_m gate is large in size and is in some sense uniformly spread out in the cube. Can the following be true?

Conjecture 1.5 (Large Support Set[2], appears in [17]). *There exists a function $h : \mathbb{N} \rightarrow \mathbb{N}$, such that any non-constant function computed by a $CC^0[m]$ circuit of size s and depth d has a support set of size at least $\frac{2^n}{2^{\Omega(\log s)^{h(d)}}}$.*

[2] In the thesis [17], where this conjecture originates, it is called the Small Support Set Conjecture referring to the fact that functions with a small support set are difficult for $\mathrm{CC}^0[m]$ circuits.

Indeed, the Large Support Set Conjecture is true in a very strong sense when m is a prime p (or a prime power). The argument goes through polynomials over $\mathbb{Z}_p$ and we point this out in Section 2 after the statement of Conjecture 2.7.

Note that in particular, the Large Support Set Conjecture implies that AND (or OR) cannot be computed in small size by $\mathrm{CC}^0[m]$ circuits. This is dual to the celebrated result that MOD_m cannot be computed easily by AC^0 circuits. Such a possibility has long been conjectured by McKenzie, Péladeau and Thérien [33]. The relative hardness of Smolensky's Conjecture and the Large Support Set Conjecture is not clear. Unfortunately, both seem out of hand for the moment.

In this article, we focus our attention on a very basic and natural model of computation: that of multilinear polynomials over the ring $\mathbb{Z}_m$. It is well known that understanding this model is absolutely necessary before significant progress on above conjectures can be made. Indeed, Razborov [39] and Smolensky [40] introduced computation by polynomials over the prime field $\mathbb{Z}_p$ as a key ingredient in their arguments for lower bounds on constant-depth circuits[3]. Unfortunately, as reviewed in the next section, understanding polynomials over $\mathbb{Z}_6$ already presents significant difficulties and several questions remain open. Our study of polynomials is motivated by Smolensky's Conjecture and the Large Support Set Conjecture. In particular, we aim to prove sort of their analogs in the polynomial world.

Before we proceed further, it is important to point out that polynomials over reals are also a very natural and interesting model of computing boolean functions. It is indeed extremely relevant for understanding constant-depth circuits. For lack of space and the sake of focus, we leave out this topic here. The interested reader can consult the excellent survey by Beigel [8] to get pointers to the older literature and more recent works like [37]. Beigel [8] also discusses polynomials over finite rings, but the survey is somewhat dated and broader in scope than ours. Here we survey some recent (and some not so recent) works on polynomials over $\mathbb{Z}_m$ and point out some of the challenges that lie ahead.

2 Computation by Polynomials

An interesting thing to observe is that every function $f : \{0,1\}^n \to \mathbb{Z}_m$ is expressible as a multilinear polynomial over $\mathbb{Z}_m$. To see this one merely has to verify that each so called *delta* function is expressible by such a polynomial.

[3] In fact their methods also work over the ring $\mathbb{Z}_{p^k}$, where p is a prime and k is fixed positive integer.

More precisely, for each $w \in \{0,1\}^n$, define the delta function $\delta_w : \{0,1\}^n \to \mathbb{Z}_m$ as $\delta_w(x) = 1$ if $w = x$ and otherwise $\delta_w(x) = 0$. Consider the set of functions $\Delta = \{\delta_w \,|\, w \in \{0,1\}^n\}$. It is easy to see that every function f can be *uniquely* expressed as a $\mathbb{Z}_m$ linear combination of such functions. On the other hand,

$$\delta_w(x) = \Big(\prod_{i:w_i=1} x_i\Big)\Big(\prod_{i:w_i=0}(1-x_i)\Big).$$

The simple identity above implies that every $\mathbb{Z}_m$-valued function over the boolean cube is expressible as a multilinear polynomial over the ring $\mathbb{Z}_m$. Indeed, a simple counting argument shows that the polynomial corresponding to each such function is unique. This enables us to view each boolean function as an algebraic object. Natural measures of the complexity of this object are its degree and the number of monomials appearing in it. Formalizing things, let $\deg_m(f)$ denote the degree of the polynomial representing the boolean function f over $\mathbb{Z}_m$. In our discussion, polylogarithmic degree will be considered small and $n^{\Omega(1)}$ degree will be high. Exhibiting a function of high degree is not hard. For example,

$$\begin{aligned} \mathrm{AND}(x) &= x_1 x_2 \cdots x_n \\ \mathrm{OR}(x) &= 1 - \prod_{i=1}^{n}(1-x_i) \end{aligned} \tag{2.1}$$

showing that $\deg_m(\mathrm{OR}) = \deg_m(\mathrm{AND}) = n$. On the other hand, demanding a polynomial P to satisfy $P(x) = f(x)$ for each point x in the cube seems too restrictive. A more natural definition, at least from a computational perspective, was introduced in the very interesting work of Barrington, Beigel and Rudich [6]. Let $A \subseteq \mathbb{Z}_m$ be an *accepting* set. Then P represents f w.r.t A if it satisfies the following property for each x in the boolean cube: $P(x) \in A \pmod{m}$ iff $f(x) = 1$. The first thing to note about this model, is that there is not a unique polynomial computing f w.r.t some fixed accepting set A. A straightforward counting argument shows that there are exactly $|A|^{|\mathrm{supp}(f)|}(m-|A|)^{2^n-|\mathrm{supp}(f)|}$ polynomials representing f w.r.t. the accepting set A.

Definition 2.1. Let $\deg_m^A(f)$ denote the minimal degree among degrees of polynomials representing f w.r.t accepting set A. The *generalized degree* of f, denoted by $\text{gen-deg}_m(f)$, is then defined to be the degree of f w.r.t. to the best accepting set, i.e.

$$\text{gen-deg}_m(f) = \min\{\deg_m^A(f) :\ A \subseteq \mathbb{Z}_m\}.$$

While it is immediate that gen-$\deg_m(f) \leq \deg_m(f)$ for every f, it is a central question in the theory of polynomial representations to determine how much degree savings can generalized representation achieve over exact representation in the ring $\mathbb{Z}_m$. For general m, it seems fairly non-trivial to get good estimates of $\deg_m^A(f)$ for even a simple f like OR and AND. However, when m is a prime p (or a prime power), tight bounds can be obtained in a simple and elegant fashion. The fact that $\mathbb{Z}_p^*$ is a group turns out to be very useful:

Fact 2.2 (Fermat's Gift). *Let p be any prime. For every $x \not\equiv 0$ (mod p), $x^{p-1} \equiv 1$ (mod p).*

This gift is great for *booleanization.* Let P be any polynomial and A any accepting set. Let $Q(x) = \sum_{a \in A} 1 - (P(x) - a)^{p-1}$. Using Fermat's Gift, it is easy to verify that $Q(x)$ is 0/1 valued modulo p and $P(x) \in A$ (mod p) iff $Q(x) \equiv 1$ (mod p). Thus, if P represented f w.r.t A, then Q is the *unique polynomial* corresponding to f. Noting that degree of Q is larger than P by a factor of at most $p-1$, one gets linear lower bounds on the degree of P if the function represented is a hard function like OR and AND (recall equation (2.1)):

Fact 2.3. *For any prime p, gen-$\deg_p(f) \geq \deg_p(f)/(p-1)$. In particular,*

$$\textit{gen-}deg_p(OR),\ \textit{gen-}deg_p(AND) \geq \frac{n}{p-1}.$$

Unfortunately, when m contains two distinct prime factors, Fermat's gift stops working. One could hope that given any accepting set $A \subset \mathbb{Z}_m$, there is some univariate 0/1 valued polynomial R over $\mathbb{Z}_m$ corresponding to the characteristic function of the set A. Indeed, Fermat's gift yields such a polynomial when m is prime. Having some such R would be enough for proving lower bounds on the generalized degree of f over $\mathbb{Z}_m$. This hope gets killed for the following reason: let $m = p_1p_2$ be a product of two distinct primes. Recall, via chinese remaindering, the map $a \mapsto ((a \mod p_1), (a \mod p_2))$ forms a bijection between $\mathbb{Z}_m$ and $\mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2}$. Thus, 0 and 1 in $\mathbb{Z}_m$ correspond to tuples $(0,0)$ and $(1,1)$ in $\mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2}$.

Fact 2.4. *Let $m = p_1p_2$ be a product of two distinct primes. Then the characteristic function of the set $A = \{1\}$ (and the set $A = \{0\}$) has no (univariate) polynomial representation over $\mathbb{Z}_m$.*

Proof. Assume for the sake of contradiction that R is such a polynomial. Applying the Chinese Remaindering Theorem, R gives rise to two polynomials, R_{p_1} over $\mathbb{Z}_{p_1}$ and R_{p_2} over $\mathbb{Z}_{p_2}$ with the property that $R(x) \mapsto (R_{p_1}(x \mod p_1), R_{p_2}(x \mod p_2))$. Now $R(0) \equiv 0 \pmod{m}$. Hence, $R_{p_1}(0) \equiv 0 \pmod{p_1}$. Similarly, $R_{p_2}(0) \equiv 0 \pmod{p_2}$. Observing that $R(1) \equiv 1 \pmod{m}$ and applying a similar argument yields the following: $R_{p_1}(1) \equiv 1 \pmod{p_1}$ and $R_{p_2}(1) \equiv 1 \pmod{p_2}$. Thus, combining things back via chinese remaindering, $R((0,1)) \equiv (0,1) \pmod{m}$ and $R((1,0)) \equiv (1,0) \pmod{m}$. However, as R is the exact representation of the characteristic function of $A = \{1\}$, $R((0,1)) \equiv R((1,0)) \equiv (0,0) \pmod{m}$, leading us to the required contradiction. □

Fact 2.4 has turned out to be somewhat of a serious blow to proving lower bounds on the composite degree of boolean functions. To some extent, this is explained by a surprising upper bound discovered by Barrington, Beigel and Rudich [6].

Theorem 2.5 (Barrington, Beigel and Rudich). *Let m have t distinct prime factors. Let $A = \{1\}$ and $A' = \mathbb{Z}_m - \{0\}$. Then, $deg_m^A(AND) = O(n^{1/t})$ and $deg_m^{A'}(OR) = O(n^{1/t})$.*

The above theorem shows that composite moduli can obtain non-trivial computational advantage over their primal counterparts when the accepting set is carefully chosen. Even more surprisingly, the above theorem has been exploited in explicit constructions in combinatorics [29, 22] and very recently in obtaining efficient locally decodable codes by Efremenko [20].

Tardos and Barrington [42] obtained the following lower bound on the generalized degree of the OR function.

Theorem 2.6 ([42]). *Let m have $t \geq 2$ distinct prime factors, and let q be the smallest maximal prime power divisor of m. Then, gen-deg$_m(OR)$ is at least $\left(\left(\frac{1}{q-1} - o(1)\right)\log n\right)^{\frac{1}{t-1}}$.*

The above lower and upper bounds on the degree for OR and AND has not been improved in more than ten years and it is an important challenge to narrow down the gap between them. On the other hand, we speculate the following:

Conjecture 2.7. *Let P be a multilinear polynomial of degree d over $\mathbb{Z}_m$. Let $a \in \mathbb{Z}_m$ be such that there exists an $x_0 \in \{0,1\}^n$ with $P(x_0) \equiv a \pmod{m}$. Then the number of points in the cube at which P evaluates to a is at least $2^{n-O(d^t)}$, where $m = p_1 \cdots p_t$ and each p_i is a distinct prime.*

It is simple to verify that this conjecture implies that for such square-free m, gen-deg$_m$(OR), gen-deg$_m$(AND) $= n^{\Omega(1/t)}$. The conjecture above admits a natural modification to composites with repeated prime factors. We do not state that formally to keep the discussion simple and focussed on the essential problem that lies ahead. Before we end this section, it is worth mentioning that the above conjecture is known to be true for prime moduli (see for example [5]). Using Ramsey Theory, Péladeau and Thérien [38] prove a result that easily implies this conjecture for arbitrary m as long as the degree d is a constant.

2.1 Computing MOD$_q$

The advantage of composites over primes is not limited to computing AND and OR. Among other things, Bhatnagar et.al.[19] showed that one can compute the THRESHOLD$_k$ function by polynomials of degree $O(n^{1/t+\epsilon})$ over $\mathbb{Z}_m$, if m has t distinct prime factors and k is a constant. This is a generalization of the upper bound due to Barrington et.al. as OR is just THRESHOLD$_1$. Bhatnagar et.al. wondered if interesting degree upper bounds could be proved for the simple function MOD$_q$. Hansen [31], disproving a conjecture of Bhatnagar et.al. [19], showed the following:

Theorem 2.8 (Hansen). *Assume $m = p_1 \cdots p_t$ and q are co-prime satisfying the following condition: there exists positive integers $b_1, \ldots, b_t$ such that $\sum_{i=1}^{t} \frac{1}{b_i} < 1$ and $p_i \geq qb_i$ for all i. Then gen-deg$_m$(MOD$_q$) $= O(n^{1/t})$.*

Tardos and Barrington's [42] technique can be adapted (see for example [13]) to prove an $\Omega\big((\log n)^{1/(t-1)}\big)$ lower bound on gen-deg$_m$(MOD$_q$). Such bounds degrade with the number of distinct prime factors of m. In a breakthrough work, Bourgain [11] proved an $\Omega(\log n)$ lower bound on the generalized MOD$_m$-degree of MOD$_q$. Bourgain's method is interesting due to several reasons. First, it proves something stronger, showing that the *correlation* between the boolean function computed by a sub-logarithmic degree polynomial over $\mathbb{Z}_m$, w.r.t. an accepting set, and MOD$_q$ is exponentially small. Such a correlation bound was not know even for polynomials modulo primes, a model which one typically assumes we understand well. The result, very significantly, improves upon a long line of work (see, for example, [21, 12, 24, 4, 26]). Second, Bourgain's method boils down to estimating certain exponential sums. This is an elementary but powerful technique that has spawned more recent progress [15, 30, 18]. Due to its importance, we include a proof of Bourgain's result. Our treatment follows that of Chattopadhyay [14, 16], that is very close to the method of [11, 27] but is slightly simpler and sharper.

Definition 2.9. For any $b \in \{0, \dots, q-1\}$, define the bth MOD$_q$-residue class of $\{0,1\}^n$, denoted by $M_q(b)$, as the following:

$$M_q(b) = \{x = (x_1, \dots, x_n) \in \{0,1\}^n \mid \sum_{i=1}^{n} x_i = b \ (\mathrm{mod}\ q)\} \tag{2.2}$$

Definition 2.10. For any polynomial P over $\mathbb{Z}_m$ and $a \in \mathbb{Z}_m$, let $P^{-1}(a)$ define the set of points in $\{0,1\}^n$ where P evaluates to a.

An intuition about a random and uniform set is that each of the $M_q(b)$ residue classes are equally represented in such a set. Bourgain's result essentially shows that if P has low degree, then $P^{-1}(a)$ appears pseudorandom[4] to the MOD$_q$ function. In other words, either each of the MOD$_q$ residue classes are almost equally represented in $P^{-1}(a)$ or the set is a very small fraction of the cube.

Lemma 2.11 (Bourgain's Uniformity Lemma). *For all positive coprime integers m, q, there exists a positive constant $\gamma = \gamma(q) < 1$ such that for every polynomial P of degree d over $\mathbb{Z}_m$ and every $a \in \mathbb{Z}_m$, the following holds:*

$$\left| \Pr\left[x \in \left(P^{-1}(a) \cap M_q(b)\right)\right] - \frac{1}{q} \Pr\left[x \in P^{-1}(a)\right] \right| \leq exp\left(- \frac{\gamma n}{\left(m 2^{m-1}\right)^d} \right). \tag{2.3}$$

Before we start the proof, let us recall an elementary fact about the primitive roots of unity that we make repeated use of henceforth. Let $\mathrm{e}_m(y)$ denote the primitive m-th root of unity raised to the yth power, i.e. $\exp(\frac{2\pi j y}{m})$, where j is the complex square root of -1. Then,

Fact 2.12. *If $y = 0$ then,$\frac{1}{m} \sum_{\alpha=0}^{m-1} \mathrm{e}_m(\alpha y)$ is 1 and the expression is 0 otherwise.*

Armed with this basic fact, we prove the Uniformity Lemma below:

Proof of Uniformity Lemma. We write $\Pr\left[x \in \left(P^{-1}(a) \cap M_q(b)\right)\right]$ as an exponential sum. Thus,

[4] The method employed by Bourgain to prove this result is closely related to method employed commonly in communication complexity to estimate the discrepancy of a function. Indeed, the quantity in the LHS of (2.3) is closely related to the discrepancy of MOD$_q$ function w.r.t. polynomial mappings modulo m . The interested reader can find more details on this point of view in [16, 17]

$$\Pr_x \left[x \in \left(P^{-1}(a) \cap M_q(b)\right)\right]$$
$$= \mathbb{E}_{x \in \{0,1\}^n} \left[\left(\frac{1}{m} \sum_{\alpha=0}^{m-1} \mathrm{e}_m\big(\alpha(P(x)-a)\big) \right) \left(\frac{1}{q} \sum_{\beta=0}^{q-1} \mathrm{e}_q\big(\beta(x_1 + \cdots + x_n - b)\big) \right) \right] \tag{2.4}$$

Expanding the sum inside the second multiplicand and treating the case of $\beta = 0$ separately, one gets

$$\begin{aligned}(2.4) = \frac{1}{q}\,\mathbb{E}_x &\left[\frac{1}{m} \sum_{\alpha=0}^{m-1} \mathrm{e}_m\big(\alpha(P(x)-a)\big) \right] \\ &+ \frac{1}{mq} \sum_{\alpha \in [m], \beta \in [q]-\{0\}} S^{m,q}(\alpha, \beta, P)\mathrm{e}_m(-a\alpha)\mathrm{e}_q(-b\beta) \end{aligned} \tag{2.5}$$

where,

$$S^{m,q}(\alpha, \beta, P) = \mathbb{E}_{x \in \{0,1\}^n} \left[\mathrm{e}_m\big(\alpha P(x)\big) \cdot \mathrm{e}_q\big(\beta(x_1 + \cdots + x_n)\big) \right] \tag{2.6}$$

Observing that the first term in (2.5) is simply $\frac{1}{q} \Pr\left[x \in P^{-1}(a)\right]$ and $|\mathrm{e}_m(-a\alpha)| = |\mathrm{e}_q(-b\beta)| = 1$, we get :

$$\left| \Pr_x \left[x \in \left(P^{-1}(a) \cap M_q(b)\right) - \frac{1}{q} \Pr_x \left[x \in P^{-1}(a)\right] \right| \leq \frac{1}{mq} \sum_{\alpha \in [m], \beta \in [q]-\{0\}} |S^{m,q}(\alpha, \beta, P)| \tag{2.7}$$

The Uniformity Lemma 2.11 gets proved by the bound on $|S^{m,q}(\alpha, \beta, P)|$ provided below. The bound below is the main technical contribution of Bourgain. □

Lemma 2.13. *For each pair of co-prime integers $m, q > 1$ there exists a constant $\gamma = \gamma(q)$ such that for every polynomial P of degree $d > 0$ in $\mathbb{Z}_m$ and numbers $\alpha \in [m]$, $\beta \in [q] - \{0\}$, the following holds :*

$$|S^{m,q}(\alpha, \beta, P)| \leq exp\bigg(- \frac{\gamma n}{(m 2^{m-1})^d} \bigg). \tag{2.8}$$

Before we begin our formal calculations, we note that a slightly weaker estimate of $|S^{m,q}(\alpha, \beta, P)|$ was first obtained by Bourgain [11] and later generalized by Green et al [27]. The case when P is a linear polynomial was essentially dealt with in [12] and forms our base case[5] just as in [11, 27].

In order to explain the intuition behind our calculations, we develop some definitions and notations. Let $f : \{0,1\}^n \to \mathbb{Z}_m$ be any function. Consider any set $I \subseteq [n]$. Note that each binary vector v of length $|I|$ can be thought of as a partial assignment to the input variables of f by assigning v to the variables in I in a natural way. Let $f^{I(v)}$ be the subfunction of f on variables not indexed in I induced by the partial assignment v to variables indexed in I. For any sequence $Y = \{y_1, \ldots, y_t\}$ having t boolean vectors from $\{0,1\}^n$, let f_Y be the function defined by $f_Y(x) = f(x) + \sum_{i=1}^{t} f(x \oplus y_i)$, where the sum is taken in Z_m. Let $I[Y] \subseteq [n]$ be the set of those indices on which every vector in Y is zero and $J[Y]$ be just the complement of $I[Y]$. Then, the following observation will be very useful in the ensuing calculation :

Observation 2.14. *Let P be a polynomial of degree d in n variables over $\mathbb{Z}_m$. Then, for each sequence Y of $m-1$ boolean vectors in $\{0,1\}^n$, the polynomial $P_Y^{J[Y](v)}$ is a polynomial of degree $d-1$ in variables from $I[Y]$ for each vector $v \in \{0,1\}^{|J[Y]|}$.*

Proof of Lemma 2.13. We drop the superscript from $S^{m,q}$ to avoid clutter in the following discussion.We shall induce on the degree d of the polynomial. Our IH is that there exists a positive real constant $\mu_{d-1} < 1$ such that for all polynomials R of degree at most $d-1$ and for all $n \geq 0$ we have $|S(\alpha, \beta, R)| \leq 2^n \mu_{d-1}^n$. The base case of $d = 0$ is easily verified and is dealt with in earlier works on correlation. Note that μ_0 depends only on q. Our inductive step will yield a relationship between μ_{d-1} and μ_d that will also give us our desired explicit bound of (2.8).

As in [11, 27], we raise S to its mth power. Our point of departure from these work, is to write $(S)^m$ in a slightly different way.

$$
\begin{aligned}
(S)^m = \mathbb{E}_{y^1,\ldots,y^{m-1} \in \{0,1\}^n} \mathbb{E}_x \Bigg[\mathrm{e}_m \left(P(x) + \sum_{j=1}^{m-1} P(x \oplus y^j) \right) \times \\
\times\, \mathrm{e}_q \left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} (x_i \oplus y_i^1) + \cdots + \sum_{i=1}^{n} (x_i \oplus y_i^{m-1}) \right) \Bigg] \qquad (2.9)
\end{aligned}
$$

Let Y be the sequence of length $m-1$ formed by a given set of vectors $y^1, \ldots, y^{m-1}$. We denote by u and v respectively the projection of x to $I[Y]$

[5] We revisit this base case later in Section 3.4.1.

and $J[Y]$. Let n_I and n_J be the cardinality of $I[Y]$ and $J[Y]$ (note that $n_I + n_J = n$) . Then, one can verify

$$(2.9) = \mathbb{E}_{y^1,\ldots,y^{m-1}\in\{0,1\}^n}\mathbb{E}_{v\in\{0,1\}^{n_J}}\Bigg[\mathrm{e}_m\big(Q^{y^1,\ldots,y^{m-1}}(v)\big)\mathrm{e}_q(n_J)\times \\ \times\,\mathbb{E}_{u\in\{0,1\}^{n_I}}\Bigg[\mathrm{e}_m\big(P_Y^{I[Y](v)}(u)\big)\mathrm{e}_q\big(m\sum_{i=1}^{n_I}u_i\big)\Bigg]\Bigg] \tag{2.10}$$

where $Q^{y^1,\ldots,y^{m-1}}$ is some polynomial that is determined by $y^1,\ldots,y^{m-1}$ and polynomial P.

The key thing to note is that Observation 2.14 implies $P_Y^{I[Y](v)}$ to be a polynomial of degree at most $d-1$ over u for every sequence $Y = y^1,\ldots,y^{m-1}$ and every vector v. Thus, the inside sum of (2.10) over the variable u can be estimated using our inductive hypothesis. Noting that the number of sequences Y for which $|I_Y| = k$ is exactly $\binom{n}{k}(2^{m-1}-1)^{n-k}$ and using the triangle inequality with the binomial theorem, we get.

$$|S|^m \le \sum_{k=0}^{n}\binom{n}{k}(2^{m-1}-1)^{n-k}2^{n-k}2^k\mu_{d-1}^k \;=\; 2^{nm}\left(1-\frac{1-\mu_{d-1}}{2^{m-1}}\right)^n \tag{2.11}$$

The rest of the calculation proceeds exactly as in Green et. al. [27]. We repeat it here for the sake of self-containment. Taking the mth root of both sides of (2.11), using the inequality $(1-x)^{1/m} \le 1 - x/m$ if $0 \le x < 1$ amd $m > 1$ after rearranging, we obtain

$$1-\mu_d \ge \frac{1-\mu_{d-1}}{m2^{m-1}} \ge \frac{1-\mu_0}{\left(m2^{m-1}\right)^d} \tag{2.12}$$

Substituting $\gamma = 1-\mu_0$, one gets $\mu_d \le \exp\big(-\frac{\gamma}{(m2^{m-1})^d}\big)$. This immediately yields (2.8) in Lemma 2.13. □

3 Computation by a System of Polynomials

It is natural to extend the notion of computation of a boolean function by a single polynomial to the notion of computation by a system of polynomials. Apart from the fact that systems of polynomials are central objects of interest in branches of pure mathematics like algebraic geometry, the study of their computational power is motivated from proving lower bounds in both boolean and arithmetic circuits. As before, the fact that our polynomials are over a ring $\mathbb{Z}_m$ (rather than a field) and that we are interested in their behavior over the boolean cube, presents difficulties

Let $\mathcal{P}$ be a system of polynomials $P_1, \ldots, P_s$, each over $\mathbb{Z}_m$ and let $A_1, \ldots, A_s$ be their respective accepting sets. The boolean function computed by $\mathcal{P}$, denoted by $f^{\mathcal{P}}$, is simply given by the following: for any $x \in \{0,1\}^n$, $f^{\mathcal{P}}(x) = 1$ if $P_i(x) \in A_i \pmod{m}$ for each $1 \leq i \leq s$, otherwise $f^{\mathcal{P}}(x) = 0$. The degree of the system $\mathcal{P}$, denoted by $\deg(\mathcal{P})$, is the degree of a maximal degree polynomial in $\mathcal{P}$, i.e. $max\{\deg(P_i) : i \leq s\}$.

Definition 3.1. The s-simultaneous MOD$_m$-degree of a boolean function f, denoted by $\deg_m^s(f)$, is the degree of a minimal degree system of s polynomials computing f.

Of course, making progress on proving degree lower bounds for a system of polynomials in general is a harder problem than proving lower bounds on the degree of a single polynomial. It may thus seem pointless to work with systems of polynomials before resolving questions from the previous section. However, consider the following: we know that a *linear polynomial* over $\mathbb{Z}_m$ cannot represent any of AND, OR and MOD$_q$ function. In fact, from results in the previous section, we know that one provably needs almost logarithmic degree to represent them. Thus, one may hope to answer questions of the following type: How large a lower bound on s can we prove so that $\deg_m^s(f) > 1$? As we will see that even for this case, proving strong lower bounds on s can be non-trivial. Additionally, such lower bounds yield new lower bounds on the size of some restricted circuits for which no other methods are currently known.

3.1 Linear Systems

Let $\mathcal{L} = \{\ell_1, \ldots, \ell_t\}$ be a set of n-variate linear forms over $\mathbb{Z}_m$. Such a set forms a linear map $\mathcal{L} : \mathbb{Z}_m^n \to \mathbb{Z}_m^t$. Conversely, given such a linear map, there exists a corresponding set of linear forms. For $v \in \mathbb{Z}_m^t$, let $K^{\mathcal{L}}(v)$ represent the set of points in $\{0,1\}^n$, that satisfy $\ell_i = v_i$ for all $1 \leq i \leq t$. Then, we show the following:

Theorem 3.2 (Chattopadhyay, Goyal, Pudlák and Thérien [15]). *For every positive integer m, there exists a positive constant c such that the following holds. Let $\mathcal{L} : \mathbb{Z}_m^n \to \mathbb{Z}_m^t$ be a linear map. For any $v \in \mathbb{Z}_m^t$, if $K^{\mathcal{L}}(v)$ is non-empty, then*

$$|K^{\mathcal{L}}(v)| \geq \frac{2^n}{c^t}. \tag{3.1}$$

A simple averaging argument shows that for every $\mathcal{L} : \mathbb{Z}_m^n \to \mathbb{Z}_m^t$, there exists a $v \in \mathbb{Z}_m^t$ such that $K^{\mathcal{L}}(v)$ has size at least $2^n/m^t$. Theorem 3.2 is a kind of concentration result in the sense that it shows that every $K^{\mathcal{L}}(v)$ is of size close to the average size if it is non-empty. We note that the results in [43], based on methods introduced in [7], imply a lower bound of $(\frac{\alpha}{\alpha-1})^n \cdot \frac{1}{\alpha^t}$ on the size of $K^{\mathcal{L}}(v)$ when it is non-empty, and α is an increasing function of m. This is still exponentially weaker than what is given by (3.1).

3.2 An Excursion

Before we prove Theorem 3.2, we draw on a notion from combinatorial group theory. Consider a fixed finite 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 of length at least k, has a non-empty subsequence that sums to zero. The pigeon-hole-principle shows that $s(G)$ is finite if G is finite. This is because if we have a sequence of length larger than $|G|^2$, then some element a of G is repeated at least $|G|$ times. The sub-sequence formed by the first $|G|$ instances of a indeed sums to zero as the order of every element in G divides $|G|$. Thus, $s(G) \leq |G|^2$, which gives a quadratic upper bound on the Davenport constant w.r.t. the size of the group.

For specific groups, one can show much better bounds. For instance, if the group is $\mathbb{Z}_p$, then one can show, using the polynomial method, that $s(\mathbb{Z}_p)$ is p. Clearly, the lower bound follows by considering the sequence of $(p-1)$ occurrences of the identity element. Such a sequence has no non-empty subsequence summing to zero. The upper bound can be established as follows: Let $a_1, \ldots, a_p$ be a sequence of elements from $\mathbb{Z}_p$. Assume that no zero-sum subsequence of it exists. In other words, the polynomial $a_1x_1 + \cdots + a_px_p$ over $\mathbb{Z}_p$ evaluates to zero only at one point in the boolean cube $\{0,1\}^p$, which is the all zero point. Thus, applying Fermat's Gift, the polynomial $P \equiv 1 - (a_1x_1 + \cdots + a_px_p)^{p-1}$, *is* exactly the OR function of p boolean variables over $\mathbb{Z}_p$. However, recall that equation (2.1) shows that the degree of the OR polynomial is p. This contradiction finishes the argument.

Olson [35] showed a more general statement: Let G be an abelian p-group of the form $\mathbb{Z}_{p^{k_1}} \oplus \mathbb{Z}_{p^{k_2}} \oplus \cdots \oplus \mathbb{Z}_{p^{k_r}}$, where $\oplus$ denotes direct sum. He shows that $s(G) = 1 + \sum_{i=1}^{r} \left(p^{k_i} - 1\right)$ in this case. We show a little later that $s(\mathbb{Z}_m^t)$ is at most $c(m)t$, where $c(m)$ is a constant that just depends on m. Before doing that, we recall another result by Olson [36] that connects $s(G)$ with 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 3.3 (Olson's Theorem). $|K(G,n)| \geq \max\{1, 2^{n+1-s(G)}\}$.

Proof adapted from [36]. We prove this by induction of n. For $n \leq s(G) - 1$, the theorem is vacuously true. Assuming it is true for n, we prove it for $n+1$. Let the equation be $g_1 x_1 + \cdots + g_{n+1} x_{n+1} = 0$. By the definition of $s(G)$, there is a subsequence of $g_1, \ldots, g_{s(G)}$ that has a subsequence that sums to zero. W.l.o.g., assume this subsequence to be $g_1, \ldots, g_t$. Then consider the equation $(-g_2)x_2 + \cdots + (-g_t)x_t + g_{t+1}x_{t+1} + \cdots + g_{n+1}x_{n+1} = 0$. By our hypothesis, this equation on n variables has at least $2^{n+1-s(G)}$ solutions. For each such solution point u, we obtain a solution to the original equation over $n+1$ variables in which the value of x_1 is set to 1 in the following way: $x_1 = 1$, for $2 \leq i \leq t$, x_i is set to the value that is the complement of its value in u, and for $t < i \leq n+1$, x_i is set to its corresponding value in u. Finally, extend the solutions of $g_2 x_2 + \cdots + g_{n+1} x_{n+1} = 0$ to our original equation by simply fixing $x_1 = 0$ to obtain at least another $2^{n+1-s(G)}$ solutions. Thus, we have at least $2^{n+2-s(G)}$ solutions in total, proving the theorem. □

3.3 A Simple Fourier Analytic Argument

The usefulness of Olson's Theorem for our purpose is evident from its following immediate corollary:

Corollary 3.4. *Let $\mathcal{L} : \mathbb{Z}_m^n \rightarrow \mathbb{Z}_m^s$ be a linear map. Then, for all $v \in \mathbb{Z}_m^s$ such that $K^{\mathcal{L}}(v)$ is non-empty, we have $|K^{\mathcal{L}}(v)| \geq 2^{n+1-s(\mathbb{Z}_m^s)}$.*

Proof. Let $\mathcal{L} \equiv \{\ell_1, \ldots, \ell_t\}$ be the underlying linear forms, where $\ell_i = a_{i,1}x_1 + \cdots + a_{i,n}x_n$. As $K^{\mathcal{L}}(v)$ is non-empty, there exists $b \in \{0,1\}^n$ such that $\ell_i(b) = v_i$. Consider $\ell'_i = a'_{i,1}x_1 + \cdots + a'_{i,n}x_n$, where $a'_{i,j} = -a_{i,j}$ if $b_j = 1$ and otherwise $a'_{i,j} = a_{i,j}$, for each $1 \leq j \leq n$ and $1 \leq i \leq t$. Define $\mathcal{L}' \equiv \{\ell'_1, \ldots, \ell'_t\}$. Then, it is straight-forward to verify that sets $K^{\mathcal{L}}(v)$ and $K^{\mathcal{L}'}(0^s)$ are in one-to-one correspondence with each other. The result follows by observing that Olson's Theorem implies $K^{\mathcal{L}'}(0^s)$ has size at least $2^{n+1-s(\mathbb{Z}_m^s)}$. □

In view of Corollary 3.4, it is sufficient to establish an $O(t)$ upper bound on $s(\mathbb{Z}_m^t)$ for proving Theorem 3.2. Although, to the best of our knowledge, determining the exact bound on $s(\mathbb{Z}_m^t)$ is still open, the linear upper bound that we seek follows from the independent work of Meshulam [34] and Therien [43]. We inlcude a proof of this, using simple Fourier analysis over groups of the form $\mathbb{Z}_m^t$. Recall, from the proof of Bourgain's Theorem in Section 2.1, $\mathrm{e}_m(y)$ denotes the primitive m-th root of unity raised to the yth power.

Theorem 3.5. *If m is even, $s(\mathbb{Z}_m^t) \leq ct$, where $c = \frac{\log m}{\log m - \log(m-1)}$ is a constant.*

Proof. Let $\mathcal{L} \equiv \{\ell_1, \ldots, \ell_t\}$ be a linear map from $\mathbb{Z}_m^s$ to $\mathbb{Z}_m^t$, such that $K^{\mathcal{L}}(0^t)$ is a singleton set, i.e. contains only the point 0^s. Let $\lambda_S : \mathbb{Z}_m^s \to \{0,1\}$ denote the characteristic function for any set $S \subseteq \mathbb{Z}_m^s$. Then, using Fact 2.12, one writes

$$\begin{aligned}
\lambda_{\{0,1\}^s}(x) &\equiv \frac{1}{m^s} \prod_{j=1}^{s} \left[\sum_{a=0}^{m-1} \mathrm{e}_m\big(ax_j\big) + \sum_{a=0}^{m-1} \mathrm{e}_m\big(a(x_j - 1)\big) \right] \\
&= \frac{1}{m^s} \prod_{j=1}^{s} \left[\sum_{a=0}^{m-1} \big(1 + \mathrm{e}_m(-a)\big)\mathrm{e}_m\big(ax_j\big) \right].
\end{aligned}$$

Let $m = 2\ell$. Then clearly for $a = \ell$, we have $(1+\mathrm{e}_m(a)) = 1+\mathrm{e}_m(\pi) = 0$ using a basic trigonometric identity. Thus, noting that $|\mathrm{supp}(\widehat{fg})| \leq |\mathrm{supp}(\hat{f})| \cdot |\mathrm{supp}(\hat{g})|$, we see that $|\mathrm{supp}(\widehat{\lambda_{\{0,1\}^s}})| \leq (m-1)^s$. Further,

$$\lambda_{K^{\mathcal{L}}(0^t)}(x) \equiv \left[\prod_{j=1}^{t} \left(\frac{1}{m} \sum_{a=0}^{m-1} \mathrm{e}_m\big(a\ell_j(x)\big) \right) \right] \lambda_{\{0,1\}^s}(x).$$

Thus, one concludes

$$\left| \mathrm{supp}\big(\widehat{\lambda_{K^{\mathcal{L}}(0^t)}}\big) \right| \leq m^t \left| \mathrm{supp}\big(\widehat{\lambda_{\{0,1\}^s}}\big) \right| \leq m^t (m-1)^s.$$

Applying the Uncertainty Principle from Fourier Analysis, we get

$$m^t (m-1)^s \geq |\mathbb{Z}_m^s| = m^s$$

whence the result follows. □

The case of an odd m can be dealt with by the following simple trick. Multiply each linear form ℓ_i by 2. Viewing each modified linear form to be over $\mathbb{Z}_{2m}$ (instead of over $\mathbb{Z}_m$), we obtain a new map $\mathcal{L}' : \mathbb{Z}_{2m}^s \to \mathbb{Z}_{2m}^t$. It is easily verified that sets $K^{\mathcal{L}}(0^t)$ and $K^{\mathcal{L}'}(0^t)$ are in one-to-one correspondence with each other. Hence, applying Theorem 3.5 to $K^{\mathcal{L}'}(0^t)$ yields bounds on $K^{\mathcal{L}}(0^t)$ as well, though with a very slight worsening of the constant c.

Corollary 3.6. *For every m, $s(\mathbb{Z}_m^t) \leq ct$, where $c = \frac{\log(2m)}{\log(2m) - \log(2m-1)}$ is a constant that just depends on m.*

Combining Corollary 3.4 with bounds on $s(\mathbb{Z}_m^t)$ as given above, we immediately derive Theorem 3.2 which states that the size of each non-empty $K^{\mathcal{L}}(v)$ is at least $\frac{2^n}{c^t}$.

Remark 3.7. Here, we point out a consequence of Theorem 3.2 for $\mathrm{CC}^0[m]$ circuits. It easily yields a linear lower bound on the size of such circuits[6] for computing AND. Such a bound was first obtained by Thérien [43]. It also makes some progress toward the Large Support Set Conjecture (see Conj. 1.5). While there it is conjectured that the size of the support set of a function computed by a $\mathrm{CC}^0[m]$ circuit decays polynomially w.r.t. the size of the circuit, Theorem 3.2 yields an exponential decay. Recently, Allender and Koucký [3] have shown that a lower bound of the form $n^{1+\gamma}$ on the size of a $\mathrm{CC}^0[m]$ circuit computing AND (MOD_q), for any constant $\gamma > 0$ that does not depend on the depth of the circuit, is enough to imply a superpolynomial lower bound on $\mathrm{CC}^0[m]$ circuits computing AND (MOD_q).

3.4 Computing MOD_q

Until recently, it was not known if a linear system $\mathcal{L} = \{\ell_1, \ldots, \ell_t\}$ over $\mathbb{Z}_m$ with arbitrary accepting sets $\{A_1, \ldots, A_t\}$ could compute MOD_q, even for $t = o(n)$. A stronger result of [15] (and implicit in the independent work of Hansen [30]), showed that even polynomial systems of low degree and small size fail to correlate well with MOD_q.

Definition 3.8. The $\mathbb{Z}_q$-discrepancy of a boolean function f, denoted by $\mathrm{disc}_q(f)$, is given by the following:

$$\mathrm{disc}_q(f) \;\equiv\; \left| \Pr\big[f(x) = 1 \wedge x \in M_q(b)\big] - \frac{1}{q} \Pr\big[[f(x) = 1\big] \right|$$

The theorem below, first obtained in [15] and independently in [30], shows that low-degree polynomial systems of small size have exponentially small $\mathbb{Z}_q$-discrepancy.

Theorem 3.9 (Polynomial Uniformity). *For all positive co-prime integers m, q, there exists a positive constant $\gamma = \gamma(m, q) < 1$ such that the following holds: let $\mathcal{P} = \{P_1, \ldots, P_t\}$ be a n-variate polynomial system of degree d over $\mathbb{Z}_m$, with accepting sets $\{A_1, \ldots, A_t\}$. Then,*

$$disc_q\big(f^{\mathcal{P}}\big) \;\le\; (m-1)^t exp\big(-n/\gamma^d\big). \tag{3.2}$$

The above result follows from a simple use of exponential sums, hinting at their untapped potential in this context.

[6] In fact, as the bound is information theoretic, one need not impose any restriction on the depth of a circuit.

Remark 3.10. The special case of the Polynomial Uniformity Theorem, obtained by restricting the system to be linear, already leads to an interesting consequence for circuits with MOD_m gates. Using this, [15] shows that circuits (of arbitrary depth) comprising only MOD_m gates cannot compute MOD_q in sub-linear size, if (m, q) are co-prime. This significantly improves upon the earlier result of Smolensky [41] that showed such circuits need $\Omega(\log n)$ size. Further, [15] combine this special case of the Polynomial Uniformity Theorem with graph-theoretic arguments to prove that such circuits of *bounded depth* need superlinear number of wires to compute MOD_q. This, in some sense, is the strongest known lower bound for general $\mathrm{CC}^0[m]$ circuits.

Very recently, Chattopadhyay and Wigderson [18] have been able to significantly improve Theorem 3.9 for the case of linear systems under the condition that m is precisely a product of two primes.

Theorem 3.11 (Two-Prime Uniformity). *Let m, q be coprime positive integers, with $m = p_1p_2$ and each p_i is a prime. There exists a positive constant $\gamma = \gamma(m, q) < 1$ such that the following holds: let $\mathcal{L} = \{\ell_1, \dots, \ell_t\}$ be a n-variate linear system over $\mathbb{Z}_m$, with accepting sets $\{A_1, \dots, A_t\}$. Then,*

$$disc_q\big(f^{\mathcal{L}}\big) \;\le\; exp\big(-\gamma n\big). \tag{3.3}$$

An interesting thing to note is that the constant γ in equation (3.3) above is independent of the size t of the system. The argument of [18] is complicated and combines ideas of using exponential sums from [15], estimates of Bourgain (Lemma 2.13 in this article) with the notion of matrix rigidity from the ingenious work of Grigoriev and Razborov [28]in *arithmetic circuits.* While space constraints will not allow us to cover the entire argument, we describe some details of the main ideas involved in proving the Two-Prime Uniformity Theorem.

3.4.1 Singleton Accepting Sets

To begin with, let us assume that each accepting set is a singleton set. In this case, w.l.o.g each $A_i \equiv \{0\}$. Then, as before, one can write

$$\begin{aligned}&\Pr_x\big[f_{\mathcal{L}}(x) = 1 \wedge x \in M_q(b)\big]\\&= \mathbb{E}_{x\in\{0,1\}^n}\left[\left\{\prod_{i=1}^{t}\left(\frac{1}{m}\sum_{\alpha=0}^{m-1}\mathrm{e}_m\big(\alpha(\ell_i(x) - a_i)\big)\right)\right\}\left(\frac{1}{q}\sum_{\beta=0}^{q-1}\mathrm{e}_q\big(\beta(x_1 + \cdots + x_n - b)\big)\right)\right]\end{aligned} \tag{3.4}$$

Mimicking arguments used in the proof of Bourgain's Uniformity Lemma to go from (2.4) to (2.7), we obtain,

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

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

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

for some $0<\alpha<1$, where the last inequality is a simple exercise to derive using the fact that m, q are co-prime. Thus, in the singleton case there is no dependence on t the number of polynomials in $\mathcal{L}$.

For general accepting sets, the first thing to do is to break down our original system into all possible singleton accepting set systems: we write $f^{\mathcal{L}} \equiv \sum_{j=1}^{s} f^{\mathcal{L}_j}$, where $\mathcal{L}_j$ is a singleton system verifying if x satisfies $\ell_i(x) = a_{i,j}$ for $1 \leq i \leq t$ and $a_{i,j} \in A_i$. Here $s \leq (m-1)^t$ as we may assume that each A_i is a proper subset of $\mathbb{Z}_m$. This decomposition of $f^{\mathcal{L}}$, along with an application of triangle inequality allows us to deal with individual $f^{\mathcal{L}_j}$ in the manner prescribed above for singleton accepting sets. It is straightforward to verify that it proves Theorem 3.9 for the restricted case of linear systems.

Remark 3.12. The careful reader may have noted that fortified with Bourgain's estimates from (2.8) in Lemma 2.13 for degree d polynomials, each step of the above argument readily adapts to polynomial systems of degree d yielding the Polynomial Uniformity Theorem. Further, it is worth pointing out that this technique yields much stronger results for *singleton polynomial systems* just as in the case of singleton linear systems described above. These stronger bounds yield exponential lower bounds for depth-four circuits of type $\mathrm{MAJ} \circ \mathrm{AND} \circ \mathrm{MOD}_m^{\{0\}} \circ \mathrm{AND}_{o(\log n)}$ (see Theorem 6 in [18]).

3.4.2 Low Rank Systems

The first thing to note is that arguments in the previous section for linear systems of small size almost instantaneously generalize to systems of low rank. Of course, we have to define rank properly as we are over the ring $\mathbb{Z}_m$

with zero divisors. The definition we need is simply the following: the $\mathbb{Z}_m$-rank of $\mathcal{L}$ is the smallest positive integer r such that there exists r linear forms in $\mathcal{L}$ that generate every other linear form in the system as some $\mathbb{Z}_m$-linear combination of them. W.l.o.g, let these basis forms be $\ell_1, \ldots, \ell_r$.

Observation 3.13. *Let $\mathcal{L}$ be a linear system of rank r. Then, $disc_q\big(f^{\mathcal{L}}\big) \leq (m-1)^r exp\big(-\gamma n\big)$, where $\gamma = \gamma(m,q)$ is a constant.*

Proof. Assume w.l.o.g., that $\ell_1, \ldots, \ell_r$ span the remaining $t-r$ forms in $\mathcal{L}$. Thus, the r-tuple $\big(\ell_1(x), \ldots, \ell_r(x)\big)$ at any point x, determines $\ell_j(x)$ for any $\ell_j \in \mathcal{L}$. Hence, we can write $f^{\mathcal{L}} \equiv \sum_{j \in J} f^{\mathcal{L}_j}$, as before, 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 the first 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 (linear subcase of) Theorem 3.9. □

Hence, if our system has sublinear rank we can prove very good bounds on the discrepancy. A tempting intuition from linear algebra suggests that systems with high (i.e. linear) rank should be almost unsatisfiable and hence their solution set cannot correlate well with a nearly balanced function like MOD_q. This may not be true because our domain of interest is the *boolean cube* and not $\mathbb{Z}_m^n$. Indeed, the following example confirms this fear: let $\mathcal{L}$ have n linear forms, with the ith linear form being just x_i. Each accepting set $A_i \equiv \{0,1\}$. Thus, the rank of this system is n, but every point in our boolean domain satisfies it!

On the other hand, this counter example represents a natural class of systems, those that are *sparse*. We say $\mathcal{L}$ is *k-sparse* if each $\ell_i \in \mathcal{L}$ has at most k non-zero coefficients (out of the possible n) appearing in it. The following shows that sparse systems have low $\mathbb{Z}_q$ discrepancy.

Lemma 3.14. *Let $\mathcal{L}$ be a k-sparse linear system in $\mathbb{Z}_m$. Then, $disc_q\big(f^{\mathcal{L}}\big) \leq exp\big(-n/\gamma^k\big)$ for some constant $\gamma(m,q)$, if m, q are co-prime.*

Proof. Consider any linear form ℓ_i in the system, with its accepting set A_i. As $\mathcal{L}$ is k-sparse, the boolean function f^{ℓ_i} depends on at most k variables. Hence, there is a polynomial P_i of degree at most k over $\mathbb{Z}_m$ that exactly represents it, i.e. $P_i(x) = f^{\ell_i}(x)$ for all $x \in \{0,1\}^n$. Replacing each ℓ_i by its corresponding P_i thus yields a singleton polynomial system $\mathcal{P}$ of degree at most k. The argument gets finished by mimicking the arguments in Section 3.4.1 (see also Remark 3.12 in that section). □

3.4.3 Low Rigid Rank

It turns out that we can combine low rank and sparsity such that we can handle linear systems which can be made to have low rank after a sparse change to each linear form. This is inspired by Valiant's famous notion of rigidity [45, 46], used to attack (so far unsuccessfully) size-depth trade-offs for computing linear systems over fields. We use the following definition:

We say $\mathcal{L}$ is (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 $\{L_i | 1 \le i \le t\}$ has rank r and every ℓ'_i is k-sparse.

Lemma 3.15. *Let $\mathcal{L}$ be a linear system that is (k,r)-sparse. Then, there exists a constant γ such that $disc_q\big(f^{\mathcal{L}}\big) \le m^r exp\big(-n/\gamma^k\big)$, when m,q are co-prime numbers.*

Proof. As before, we look at the possible evaluations of the various linear forms. Let t be the size of $\mathcal{L}$, 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 in $\mathbb{Z}_m^r$. Given u, we know what each L_i evaluates to in $\mathbb{Z}_m$, for all $i \le t$. 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. 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 the following:

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

Expanding out the product of sums into sum of products,

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

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 Bourgain's estimate (2.8), we are done. □

At this point, could we hope that systems that are not (k, r)-sparse, i.e. do not have low rigid rank are hardly satisfiable over the cube? Indeed, such a hope is generated from a beautiful result of Grigoriev and Razborov [28]: they manage to show that if a linear system $\mathcal{L}$ *over a field* $\mathbb{F}_q$ has high rigid rank, then an exponentially small fraction of the set of points in the boolean cube satisfy the system. To show this, they introduce an ingenious notion of rank called communication rank. Porting their argument to our setting raises an obvious difficulty: they work over a field and we work over the ring $\mathbb{Z}_m$.

However, in [18], we show that their argument can be generalized to our setting in the following sense: let $m = p_1 \cdots p_s$ be a product of s distinct primes. Let $\mathcal{L} \equiv \{\ell_1, \ldots, \ell_t\}$ be a linear system having t linear forms in $\mathbb{Z}_m$. Via chinese remaindering, any linear form ℓ_i projects to s linear forms $\ell_i^1, \ldots, \ell_i^s$, where ℓ_i^j is in the field $\mathbb{Z}_{p_j}$. Hence, $\mathcal{L}$ naturally projects to s linear systems $\mathcal{L}^1, \ldots, \mathcal{L}^s$, with $\mathcal{L}^j$ in $\mathbb{Z}_{p_j}$. Indeed, one could consider the rank and sparsity of each $\mathcal{L}^j$ in the field $\mathbb{Z}_{p_j}$. Motivated by this, we say $\mathcal{L}$ in $\mathbb{Z}_m$ is r-simple if the set of linear forms can be partitioned into s sets $J_1, \ldots, J_s$ with the following property: the projection of the set of linear forms in J_j to $\mathbb{Z}_{p_j}$ forms a $\big(sm, (sm+1)r\big)$-sparse system.

Theorem 3.16 (Chattopadhyay and Wigderson [18], extending Grigoriev and Razborov [28]). *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 with no repeated prime factors. If $\mathcal{L}$ is not r-simple, then*

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

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

We remark that the proof of the above theorem uses very different techniques than any that we have covered here. In particular, it involves an interesting combination of elementary additive combinatorics and linear algebra. Theorem 3.16 provides a rank-sparsity condition under which the system becomes highly unsatisfiable. It is worth noting that apart from assuming that m is square-free, it does not limit the number of prime factors of m. Extending ideas from the proof of Lemma 3.15, [18] complements the above Theorem by the following:

Lemma 3.17. *Let $\mathcal{L}$ be a linear system over $\mathbb{Z}_m$ with $m = p_1 p_2$. Let linear forms in $\mathcal{L}$ admit a partition into sets J_1 and J_2 such that the set of linear* *forms in J_i are (k, r)-sparse over $\mathbb{Z}_{p_i}$ for each $i \leq 2$. Then, if m, q are co-prime,*

$$disc_q^b\big(f^{\mathcal{L}}\big) \leq m^{2r} exp\bigg(- \frac{n}{\gamma^{k+m-1}}\bigg).$$

where $\gamma = \gamma(m, q)$ is a constant.

Unfortunately, the argument in [18] for proving the above works only for the case when m is precisely a product of two primes. It is not hard to combine Lemma 3.17 and Theorem 3.16 to prove the Two-Prime Uniformity Theorem. We do not waste space filling in more details as the interested reader can find the full argument in [18].

4 Conclusion

We argued that the world of low-degree multilinear polynomials modulo a composite is a very natural and fascinating setting to explore the power of modular counting. Fundamental questions on the degree needed to represent simple functions remain wide open. No serious bottleneck is known that prevents us from making progress on them. We believe that with more efforts these problems can be solved in the not too distant future.

On the other hand, mysteriously $\log n$ comes up as a common barrier in different settings. For instance, it shows up in the argument of Tardos and Barrington [42] seemingly for one reason and in Bourgain's [11] argument for seemingly another. Is it just coincidental? More intriguingly, by the result of Beigel and Tarui [10] (improving upon an earlier work of Yao [48]), we know that every function in ACC^0 can be written as $f\big(P(x_1, \ldots, x_n)\big)$, where f is a symmetric function and P is an *integer* polynomial of polylogarithmic degree with coefficients of magnitude at most quasipolynomial. Again $\Omega(\log n)$-degree bounds can be proven for P (via multiparty communication complexity) to decompose a simple function like GIP that can be trivially computed in ACC^0[2]. Improving over $\log n$ is wide open! While it is conceivable that going past $\log n$ degree is difficult for a general symmetric function f, it is remarkable that we are stuck, more or less, at the same place even when f is a very special symmetric function like MOD_m^A. Viola and Wigderson [47], using the language of Gower's norm [23], try to suggest an answer. In a related work, Chattopadhyay [16] argues that troubles on both fronts emanate from the technique of repeatedly raising the sum in question to a fixed power, until the degree of the polynomial crashes to linear. While these provide some clue, we feel that the mystery is not entirely solved.

Assuming that going past $\log n$ degree is a difficult task, we have explored questions about systems of polynomials of degree well below $\log n$. Even understanding linear systems over a modulus that is just a product of two distinct primes has proved non-trivial and has generated interesting mathematics. It is hard to believe that this cannot be pushed to three and more prime factors. Finally, can one generalize Bourgain's result to the setting of a system of polynomials over $\mathbb{Z}_m$, where each polynomial has appropriately low degree. We know this is true at least when m is a prime power by techniques discussed in this article.

References

[1] E. Allender. Circuit complexity before the dawn of the new millenium. *16th Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)* LNCS 1180, 1–18, 1996.

[2] E. Allender. Cracks in the Defenses: Scouting Out Approaches on Circuit Lower Bounds. *3rd International Computer Science Symposium in Russia(CSR)* LNCS 5010, 3–10, 2008.

[3] E. Allender and M. Koucký. Amplifying lower bounds by means of self-reducibility. *IEEE Conference on Computational Complexity* 31–40, 2008.

[4] N. Alon and R. Beigel. Lower bounds for approximations by low degree polynomials over $\mathbb{Z}_m$. *16th Annual IEEE Conference on Computational Complexity*, 184–187, 2001.

[5] D. A. M. Barrington. Some problems involving Razborov-Smolensky polynomials. *Boolean function complexity* London Math.Soc.Lec.Note.(169), Cambridge University Press, 109–128, 1992.

[6] D. A. M. Barrington and R. Beigel and S. Rudich. Representing boolean functions as polynomials modulo composites. *Computational Complexity*(4), 367–382, 1994.

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

[8] R. Beigel. The polynomial method in circuit complexity. *Structure in Complexity*, 82–95, 1993.

[9] R. Beigel and A. Maciel. Upper and lower bounds for some depth-3 circuit classes. *Computational Complexity*, 6(3):235–255, 1997.

[10] R. Beigel and J. Tarui. On ACC. *Computational Complexity*, 4:350–356, 1994.

[11] J. Bourgain. Estimates of certain exponential sums arising in complexity theory. *C.R.Acad.Sci.Paris*(9), Ser I 340, 627–631, 2005.

[12] J. Y. Cai and F. Green and T. Thierauf. On the correlation of symmetric functions. *Mathematical Systems Theory*, 29(3), 245–258, 1996.

[13] A. Chattopadhyay and K. A. Hansen Lower bounds for circuits with few modular and symmetric gates. *International Colloquium on Automata, Languages and Programming (ICALP)*, Lisbon, Portugal 994–1005, 2005.

[14] A. Chattopadhyay. An improved bound on correlation between polynomials over $\mathbb{Z}_m$ and MOD_q. *Electronic Colloquium on Computational Compelxity*, TR06-107, 2006.

[15] A. Chattopadhyay and N. Goyal and P. Pudl'ak and D. Thérien. Lower bounds for circuits with MOD_m gates. *IEEE Symposium on Foundations of Computer Science (FOCS)*, Berkeley, 709–718, 2006.

[16] A. Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. *IEEE Symposium on Foundations of Computer Science (FOCS)* Providence, RI, 449–458, 2007.

[17] A. Chattopadhyay. Circuits, communication and polynomials. *Ph.D. Thesis* School of Computer Science, McGill University, Montreal, 2008.

[18] A. Chattopadhyay and A. Wigderson. Linear systems over composite moduli. *IEEE Symposium on Foundations of Computer Science (FOCS)*, Georgia, 709–718, 2009.

[19] N. Bhatnagar and P. Gopalan and R. J. Lipton. Symmetric polynomials over $\mathbb{Z}_m$ and simultaneous communication protocols. *J.Comput.System.Sciences* 72(2):252–285, 2006.

[20] K. Efremenko. 3-query locally decodable codes of subexponential length. In *41st Annual Symposium on Theory of Computing (STOC)*, 39–44 2009.

[21] M. Goldmann. A note on the power of Majority gates and Modular gates. *Inf.Process.Lett.*, 53(6):321–327, 1995.

[22] P. Gopalan. Constructing Ramsey graphs from boolean function representations. *IEEE Conference on Computational Complexity*, 115–128, 2006.

[23] T. Gowers. A new proof of Szemeredi's theorem. *Geometric and Functional Analysis*, 11(3):465–558, 2001.

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

[25] F. Green. A complex-number Fourier technique for lower bounds on the MOD_m-degree. *Computational Complexity*, 9, 16–38, 2000.

[26] F. Green. The correlation between parity and quadratic polynomials mod 3. *J.Computer.Systems.Sciences* 69(1), 28–44, 2004.

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

[28] 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*, 10(6):465–487, 2000.

[29] V. Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit ramsey graphs. *Combinatorica*, 20(1):71–86, 2000.

[30] K. A. Hansen. Lower bounds for circuits with few modular gates using exponential sums. Technical Report TR06-079, Electronic Colloquium on Computational Complexity, 2006.

[31] K. A. Hansen. On modular counting with polynomials. In *IEEE Conference on Computational Complexity*, pages 202–212, 2006.

[32] K. A. Hansen amd M. Koucký. A new characterization of ACC^0 and probabilistic CC^0. In *IEEE Conference on Computational Complexity*, pages 27–34, 2009.

[33] P. McKenzie, P. Péladeau, and D. Thérien. NC^1: The automata-theoretic viewpoint. *Computational Complexity*, 1:330–359, 1991.

[34] R. Meshulam. An uncertainty inequality and zero subsums. *Discrete Mathematics*, 84:187–200, 1990.

[35] J. E. Olson. A combinatorial problem on finite abelian groups, I. *J. Number Theory*, 1:8–10, 1969.

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

[37] R. O'Donnell and R. Servedio. Extremal properties of polynomial threshold functions. In *18th Annual IEEE Conference on Computational Complexity*, pages 3–12. IEEE, 2003.

[38] P. Péladeau and D. Thérien. Sur les langages reconnus par des groupes nilpotents. *C.R. Acad. des Sci. Paris Sér. I Math.*, 306(2):93–95, 1988. English translation by A. Russell and S. Russell appears as TR01-040 of ECCC.

[39] 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*, volume 41, pages 333–338. 1987.

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

[41] R. Smolensky. On interpolation by analytic functions with special properties and some weak lower bounds on the size of circuits with symmetric gates. In *31st IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 628–631, 1990.

[42] G. Tardos and D. A. M. Barrington. A lower bound on the MOD 6 degree of the OR function. *Computational Complexity*, 7(2):99–108, 1998.

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

[44] S. C. Tsai. Lower bounds on representing boolean functions as polynomials in $\mathbb{Z}_m$. *SIAM J. Discrete Math*, 9:55–62, 1996.

[45] L. Valiant. Some conjectures relating to superlinear complexity. Technical Report 85, University of Leeds, 1976.

[46] L. Valiant. Graph-theoretic arguments in low-level complexity. In *The 6th Mathematical Foundations of Computer Science (MFCS)*, volume 53 of *LNCS*, pages 162–176, 1977.

[47] E. Viola and A. Wigderson. Norms, XOR Lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *22nd Annual IEEE Conference on Computational Complexity*, pages 141–154. IEEE Computer Society, 2007.

[48] A. C. C. Yao. On ACC and Threshold circuits. In *37th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 619–627. IEEE Computer Society, 1990.