Exponential Complexity Lower Bounds for Depth 3 Arithmetic Circuits in Algebras of Functions Over Finite Fields

D. Grigoriev[1]

Departments of Mathematics and Computer Science
The Pennsylvania State University
University Park, PA 16802
dima@cse.psu.edu

Alexander A. Razborov[2]

Steklov Mathematical Institute
Gubkina 8, 117966, GSP-1
Moscow, Russia

A depth 3 arithmetic circuit can be viewed as a sum of products of linear functions. We prove an exponential complexity lower bound on depth 3 arithmetic circuits computing some natural symmetric functions over a finite field F. Also, we study the complexity of the functions $f : D^n \to F$ for subsets $D \subset F$. In particular, we prove an exponential lower bound on the complexity of a depth 3 arithmetic circuit which computes the determinant or the permanent of a matrix considered as functions $f : (F^*)^{n^2} \to F$.

[1]Partially supported by NSF Grant CCR-9424358.
[2]Supported by Russian Basic Research Foundation grant 96-01-01222.

Introduction.

We study depth 3 arithmetic circuits computing a function over a finite field (see (1) in section 1 below). Basically, a depth 3 arithmetic circuit can be viewed as a sum of products of linear functions. Despite this clear structure, it appears to be quite difficult to obtain for them complexity lower bounds. Let us mention that the boolean theory of depth 3 (and even bounded-depth) circuits is much more developed (see e.g., [6], [7], [11], [12], [13], [14], [16]). On the other hand, there were also considered *arithmetic* bounded-depth circuits over the field $GF(2)$ (see [1] and the further literature there).

It is somewhat easier to obtain exponential complexity lower bounds for depth 3 arithmetic circuits under the assumption that they involve just (homogeneous) linear forms, rather than arbitrary linear functions, by the same token, if products in a circuit contain a bounded number of linear functions (see [3], [10]). A breakthrough for arbitrary depth 3 arithmetic circuits was made in [5] where an exponential complexity lower bound was proved for a circuit computing the determinant in the *algebra of polynomials* over a finite field.

First, in this paper an exponential complexity lower bound is proved for circuits computing symmetric functions over finite fields (theorem 1 in section 1), in particular, the symmetric functions MOD_{p_1} over the field $GF(p)$ for distinct primes p, p_1. Thus, a depth 3 arithmetic circuit in section 1 is treated in the algebra of functions over a finite field.

Afterwards in the next sections we study depth 3 arithmetic circuits which compute "quasi-boolean" functions $f : D^n \to F$ where $D \subset F$. This setting extends the approach of [14] where the boolean case $D = \{0, 1\}$ was considered. In section 2 we give some basic properties of the algebra of all functions $f : D^n \to F$.

In section 3 we introduce a m-communication rank and m-rigid rank of a matrix and relate them (lemma 3).

In section 4 we state that a product of linear functions has only *few nonzeroes* on D^n, provided that this product has a large communication (or thereby, rigid as well) rank (lemma 4). This allows to approximate products of linear functions with large communication ranks from a depth 3 arithmetic circuit by a zero function, and in the sequel to deal only with the products having small communication ranks.

In section 5 we provide an approximation of a function $f : D^n \to F$ with a small depth 3 arithmetic circuit complexity by means of a function having

some special form (theorem 2). This allows in a particular case $D = F^*$ (or more generally when $|D| = |F| - 1$) to provide an approximation by means of a sparse polynomial (lemma 5). Moreover, the support (the set of monomials) of this polynomial lies in a union of few balls (w.r.t. the Hamming metric) each of a small radius.

Continuing this topic in section 6, we prove that if the support of a function $f : (F^*)^n \to F$ has a large coding distance then a lower bound of the complexity of computing f by a depth 3 arithmetic circuit is exponential (theorem 3). As a consequence, we obtain exponential complexity lower bounds for computing the determinant or the permanent of a matrix, treating them as the functions $f : (F^*)^{n^2} \to F$. Thereby, it gives, in particular, the exponential lower bound for the determinant or the permanent in the algebra of functions $f : F^{n^2} \to F$, which strengthens the result of [5].

1 Exponential lower bound for depth 3 arithmetic circuits for symmetric functions over a finite field.

We study depth 3 arithmetic circuits, so representations of a function in the following form:

$$f = \sum_{1 \le i \le N} \prod_j L_{ij} \tag{1}$$

where $L_{ij} = \sum_{1 \le \ell \le n} \alpha_{ij}^{(\ell)} X_\ell + \alpha_{ij}^{(0)}$ are linear functions. We fix a prime p, denote by $F = GF(p)$ a finite field and in this section we consider the identity (1) for functions $f : F^n \to F$ over the field F. The purpose of this section is to obtain exponential lower bounds on the complexity (in fact, on N) in the representations (1) for quite natural symmetric functions f.

Viewing each element $x \in F$ as an integer $0 \le x \pmod p \le p - 1$, one can define for any prime q (similar to [14]) a function $MOD_q : F^n \to F$ as follows:

$$MOD_q(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } \sum\limits_{1 \le i \le n} x_i \pmod p \equiv 0 \pmod q \\ 0 & \text{else} \end{cases}$$

Theorem 1 *For the complexity of depth 3 arithmetic circuit (1) representing MOD_q (provided $q \neq p$) the lower bound $N \geq \exp(\Omega(\sqrt{n}/\log n))$ holds (in case $p = 2$ one has a stronger bound $N \geq \exp(\Omega(n))$).*

Proof. Suppose the contrary. By the rank $rk(\prod_j L_{ij})$ we mean the rank of the matrix of coefficients $(\alpha_{ij}^{(\ell)})_{j,\ell}$ of the linear functions $\{L_{ij}\}_j$ without their free terms. Take some threshold r being a real number (for the time being it varies, we'll specify it later). We treat separately the products $\prod_j L_{ij}$ with the rank greater or less than r, respectively.

1.1 Products with the large rank.

Let $rk\left(\prod_j L_{ij}\right) \geq r$. Then the probability $Pr\left[x \in F^n : \prod_j L_{ij}(x) \neq 0\right] \leq \left(\frac{p-1}{p}\right)^r$.

1.2 Products with the small rank.

Now let $rk\left(\prod_j L_{ij}\right) < r$. Then the degree $\deg\left(\prod_j L_{ij}\right) \leq r(p-1)$.

Indeed, express each L_{ij} as a linear combination of (less than r) elements of a basis, thereupon open the parenthesis in the product and use the relation $L^p = L$ for any (linear) function L.

Denote by g the sum of all the products $\prod_j L_{ij}$ (from (1)) of (small) ranks less than r.

Then $Pr[x \in F^n : (MOD_q - g)(x) \neq 0] \leq \left(\frac{p-1}{p}\right)^r N$. In case $p = 2$ one can take $r = n/2 - \sqrt{n}$ (see [11])and complete the proof of the theorem. From now on we assume that $p \geq 3$.

Consider the boolean cube $B^n = \{0,1\}^n \subset F^n$. For any vector $c = (c_1, \ldots, c_n) \in F^n$ consider a (shifted) function $MOD_q^{(c)} : B^n \to F$ defined by the formula $MOD_q^{(c)}(x_1, \ldots, x_n) = MOD_q(x_1 + c_1, \ldots, x_n + c_n)$. We call c nondegenerated if at least $n/2$ of its coordinates $c_1, \ldots, c_n$ are distinct from

$p-1$. The number of degenerated vectors c does not exceed

$$\left(\begin{array}{c} n \\ n \end{array} \right) (p-1)^0 + \left(\begin{array}{c} n \\ n-1 \end{array} \right) (p-1)^1 + \cdots + \left(\begin{array}{c} n \\ [n/2] \end{array} \right) (p-1)^{[n/2]}$$
$$< 2^{n-1}(p-1)^{n/2} \leq \frac{1}{2}p^n, \text{ hence } Pr[c \in F^n \text{ is nondegenerated}] > \frac{1}{2}.$$

Since any point from F^n belongs to the same number 2^n of the shifted boolean cubes $B^n + c$ (for both degenerated and nondegenerated vectors c), we conclude that the expectation over all *nondegenerated* vectors $c \in F^n$

$$E(Pr[x \in B^n : (MOD_q^{(c)} - g^{(c)})(x) \neq 0]) \leq 2 \left(\frac{p-1}{p} \right)^r N$$

where the function $g^{(c)}(x) = g(x+c) : B^n \to F$ is represented by a polynomial of the degree less than $r(p-1)$. Pick out a nondegenerated $c \in F^n$ for which

$$Pr[x \in B^n : (MOD_q^{(c)} - g^{(c)})(x) \neq 0] \leq 2 \left(\frac{p-1}{p} \right)^r N. \qquad (2)$$

Let $1 \leq j_1 < \cdots < j_s \leq n$ be such that $0 \leq c_{j_i} \pmod p \leq p-2$, $1 \leq i \leq s$, $s \geq n/2$. Substituting zeroes for all ℓ with $\ell \notin \{j_1, \ldots j_s\}$ into the function $MOD_q^{(c)}(X_1, \ldots, X_n)$, one obtains a function $MOD_{q,t} : B^s \to F$ for a certain $0 \leq t < q$ (recall [14] that $MOD_{q,t}(y_1, \ldots, y_s) = 1$ if the number of ones among $y_1, \ldots, y_s \in \{0,1\}$ has a residue $t \pmod q$, otherwise $MOD_{q,t}(y_1, \ldots, y_s) = 0$). Specifying $r = \sqrt{n/\log n}$ we get a contradiction of (2) with the corollary and lemmas 4,5 from [14], that proves theorem 1.

One can introduce a symmetric function $MAJ_F : F^n \to F$ similar to the customary $MAJ : B^n \to \{0,1\}$ and being universal for all symmetric functions. Namely, $MAJ_F(x_1, \ldots, x_n) = 1$ if $\gamma_0 \leq \gamma_1 \leq \cdots \leq \gamma_{p-1}$ where γ_i equals to the number of $i \pmod p$ among $x_1, \ldots, x_n$ for $0 \leq i \leq p-1$, otherwise $MAJ_F(x_1, \ldots, x_n) = 0$. One can show (similar to theorem 4 [11]) that any symmetric function $f(x_1, \ldots, x_n) : F^n \to F$ could be represented as a F-linear combination of the functions of the form $MAJ_F(X_1, \ldots, X_n, \eta_1, \ldots, \eta_{n(p-1)})$ for suitable $\eta_i \in F$, $1 \leq i \leq n(p-1)$. This entails the following corollary.

Corollary 1 *For the complexity of depth 3 arithmetic circuit (1) representing MAJ_F the lower bound $N \geq \exp(\Omega(\sqrt{n}/\log n))$ holds.*

2 Quasi-boolean functions over finite fields.

In the rest of the paper we deal with the following setting. Let $q = p^\lambda$, $F = GF(q)$, fix a subset $D \subset F$, for nondegenerality we suppose that $|D| \geq 2$. The case $D = F$ was studied in section 1, that is why we shall assume that $D \neq F$. A function $f : D^n \to F$ we call *quasi-boolean* extending the concept from [14] where D was just $\{0, 1\}$ (actually one could extend this concept further considering functions, say $f : D_1 \times \cdots \times D_n \to F$ for $D_1, \ldots, D_n \subset F$, but we restrict overselves with the case $D_1 = \cdots = D_n = D$ for the sake of simplicity of notations). One could also view f as a partial function on the entire space F^n.

Let $g \in F[X]$ be the minimal polynomial with the set of roots equal to D (evidently, $\deg g = |D|$). Then the F-algebra of all functions $f : D^n \to F$ is isomorphic to the quotient algebra $\mathcal{A} = F[X_1, \ldots, X_n]/(g(X_1), \ldots, g(X_n))$.

The main purpose in the next sections is to obtain lower bounds on the complexity of depth 3 arithmetic circuits (1) for certain functions $f \in \mathcal{A}$, thereby equality (1) is viewed in the algebra $\mathcal{A}$.

Let us mention some easy properties of $\mathcal{A}$ which we use below (cf. [14], also [4]).

Lemma 1 *a) $\mathcal{A}$ is an algebra of principal ideals;*

b) for any $f \in \mathcal{A}$ the number of nonzeroes $|\{x \in D^n : f(x) \neq 0\}|$ coincides with the dimension of the principal ideal $(f) \subset \mathcal{A}$.

Clearly, monomials of the form $X_1^{i_1} \cdots X_n^{i_n}$, $0 \leq i_1, \ldots, i_n < |D|$, constitute a basis of $\mathcal{A}$, for an element $f \in \mathcal{A}$ we refer to its degree w.r.t. this basis. In abuse of notations we identify sometimes f with the corresponding polynomial in this basis.

Below it will be sometimes convenient to imagine a metric geometry in the space of monomials $\mathcal{M} = \{X_1^{i_1} \cdots X_n^{i_n}\}_{0 \leq i_1, \ldots, i_n < |D|}$ endowed with the Hamming distance ρ (being equal to the number of distinct coordinates). If the minimal polynomial $g \in F[X]$ of D is a binomial $q = x^\ell - a$ (obviously, $g|(X^{q-1} - 1)$) then $\mathcal{M}$ satisfies the following property.

Definition 1 *We say that the algebra $\mathcal{A}$ has a multiplicative basis of the monomials $\mathcal{M}$ if*

1) for any two monomials $m_1, m_2 \in \mathcal{M}$ there is a monomial $m \in \mathcal{M}$ which equals to their product $m = m_1 m_2$ (in $\mathcal{A}$);

2) the Hamming distance is invariant under multiplication: $\rho(m_1, m_2) = \rho(m'm_1, m'm_2)$.

We note that the algebra of functions on F (see section 1), in other words $D = F$, $g = X^q - X$, satisfies the condition 1), whereas, fails to satisfy 2), it satisfies just the inequality $\rho(m_1, m_2) \geq \rho(m'm_1, m'm_2)$.

Definition 2 *For any integer* $2 \leq d < q$ *a linear d-sweep* $\ell s(d)$ *(w.r.t. the field* $F = GF(q)$*) is the minimal* ℓ *(provided that it does exist) such that for any* ℓ *subsets* $D_1, \ldots, D_\ell \subset F$, $|D_1| = \cdots = |D_\ell| = d$ *any linear function* $L(X_1, \ldots, X_\ell) = a_1X_1 + \cdots + a_\ell X_\ell + a$ *with nonvanishing coefficients* $a_1 \neq 0, \ldots, a_\ell \neq 0$, *sweeps* $L(D_1 \times \cdots \times D_\ell) = F$ *the entire* F.

Lemma 2 *a)* $\ell s(d)$ *is defined if and only if* $d > p^{\lambda-1}$, *in this case* $\ell s(d) \leq q - d + 1$;

b) for $d > q/2$ *we have* $\ell s(d) = 2$.

3 Rigid rank and communication rank of a matrix.

In this section we consider matrices over arbitrary (not necessary finite) fields.

Let $A = (a_{ij})$ be $k \times n$ matrix and $m \geq 0$ be an integer. For subsets $I \subset \{1, \ldots, k\}, J \subset \{1, \ldots, n\}$ we denote by $A_{I,J}$ the submatrix of A formed by its rows from I and the columns from J.

Definition 3 *m-rigid rank* $rrk_m(A)$ *of* A *is defined as the minimal possible rank of matrices which differ from* A *by at most of* m *entries in each row.*

Definition 4 *m-communication rank* $crk_m(A)$ *of* A *is defined as the maximal possible number* r *of its rows* $I \subset \{1, \ldots, k\}$, $|I| = r$ *such that there exists a partition* $\{1, \ldots, n\} = J_0 \cup \cdots \cup J_m$ *of the set of its columns into* $(m + 1)$ *subsets with the property that every submatrix* A_{I,J_ℓ}, $0 \leq \ell \leq m$ *has the rank* r.

Notice that both rrk_m and crk_m are not invariant in general with respect to transposing the matrix A. Obviously, rrk_0 and crk_0 coincide with the usual rank.

The next lemma relates the rigid and communication ranks.

Lemma 3 $rrk_m(A) \leq (m + 2)\, crk_m(A) \leq (m + 2)(m + 1) rrk_m(A)$.

4 Upper bound on the number of nonzeroes of a product of linear functions via the communication rank.

Throughout this section we fix a product of linear functions $\Pi = \prod_j L_j$ (cf. (1)) and a subset $D \subset F$, denote $d = |D|$, viewing $\Pi : D^n \to F$ as a quasi-boolean function (see section 2). Let $L_j = \sum_{1 \le i \le n} a_j^{(i)} X_i + a_j$ where $a_j^{(i)}, a_j \in F$. By the ℓ_1-communication rank of Π we mean the ℓ_1-communication rank (see section 3) of the matrix $A = (a_j^{(i)})$ of its coefficients, L_j being treated as the rows of A (excluding the free coefficients of L_j). Denote $\ell = \ell s(d)$ (we assume that ℓ is defined, i.e. $d > p^{\lambda-1}$, then lemma 2a) entails that $\ell \le q-1$, see section 2).

The purpose of this section is to bound from above the number of nonzeroes of Π on D^n via the $(\ell-1)$-communication rank of Π.

Lemma 4 *Let the $(\ell-1)$-communication rank of Π equal to r. Then the probability*

$$Pr[x \in D^n : \Pi(x) \neq 0] \le \exp(-\Omega(r)).$$

5 Approximating depth 3 arithmetic circuits by sparse polynomials

In this section we show that if a quasi-boolean function $f : D^n \to F$, $|D| = d > p^{\lambda-1}$ can be computed by a depth 3 arithmetic circuit (1) with a small complexity N then f can be approximated by a polynomial of a special type (see theorem 2 below), in case $D = F - \{0\}$ this polynomial is sparse (see lemma 5 below). For this purpose it suffices to approximate a product $\Pi = \prod L_j$ of linear functions.

Let $\ell = \ell s(d)$, recall that $\ell \le q-1$ due to lemma 2 (cf. also section 4). Again (cf. section 1) we fix a threshold r (which could be varied). The proof of the following theorem uses lemmas 3,4.

Theorem 2 *For any r a product $\prod_j L_j$ of linear functions can be approximated on D^n by a suitable function of the form*

$$g_1 = \sum_{1 \le i \le \exp(O(r))} \prod_m \mathcal{L}_{i,m} \tag{3}$$

where $\mathcal{L}_{i,m}$ are linear functions such that for every i the number of m for which $\mathcal{L}_{i,m}$ has at least one zero in D^n, is bounded by $O(r)$. The approximation means that the probability

$$Pr\left[x \in D^n : \prod_j L_j(x) \neq g_1(x)\right] \leq \exp(-\Omega(r)).$$

If one would be able to prove that a particular function $f : D^n \to F$ for a certain r can not be approximated by a function of the form (3), then theorem 2 would imply a lower bound $N \geq \exp(\Omega(r))$ on the complexity of computing f by a depth 3 arithmetic circuit (1).

We prove a complexity exponential lower bound in the next section for an explicitly given function f in case when $|D| = q - 1$, from now on we study this particular case.

Let $\{a\} = F - D$. Observe that the only linear functions $L \in F[X_1, \ldots, X_n]$ without zeroes in D^n, are $\{X_i - a\}_i$ (up to constant factors), in particular, $\ell = \ell s(|D|) = 2$ (cf. lemma 2b). For the sake of simplicity of notations we assume in the sequel that $D = F^* = F - \{0\}$, nevertheless, all the further results still hold for any $D = F - \{a\}$ by means of the linear transformation of the coordinates $X_i \to X_i + a$, $1 \leq i \leq n$.

Lemma 5 *In case $D = F^*$ for any r a product $\Pi = \prod_j L_j$ of linear functions can be approximated on D^n by an appropriate function of the form $\widetilde{g} = \prod_{1 \leq i \leq n} X_i^{\mu_i} \cdot g$, $0 \leq \mu_i \leq p - 2$, $1 \leq i \leq n$ where $\deg g \leq O(r)$. Again the approximation means that the probability*

$$Pr[x \in D^n : \Pi(x) \neq \widetilde{g}(x)] \leq \exp(-\Omega(r))$$

As we mentioned already a similar statement is valid for any $D = F - \{a\}$ by means of replacing $\widetilde{g}$ for $\prod_{1 \leq i \leq n}(X_i - a)^{\mu_i} g$.

One can view $\widetilde{g}$ from lemma 5 as a sparse (when r is relatively small) polynomial with less than $n^{O(r)}$ monomials. Moreover, $\widetilde{g}$ has a special structure and that is why we introduce the following definition.

Definition 5 *A polynomial of the form $\sum_{1 \leq \eta \leq t} X^{I_\eta} g_\eta$ where X^{I_η} is a monomial and $\deg(g_\eta) \leq \widetilde{r}$ is called $(t, \widetilde{r})$-sparse.*

Lemma 5 implies the following proposition.

Proposition 1 *There exists a constant $\delta_0 > 0$ such that if a function $f : (F^*)^n \to F$ can be computed by an arithmetic depth 3 circuit (1) with the complexity $N \leq \exp(\delta_0 r)$ for a certain r then f can be approximated by a pertinent $(N, O(r))$-sparse function g, i.e.*

$$Pr[x \in (F^*)^n : f(x) \neq g(x)] \leq \exp(-\Omega(r))$$

6 Exponential lower bound for arithmetic depth 3 circuits computing the determinant or the permanent of matrices with nonzero entries

The main goal of this section is to prove an exponential lower bound on the complexity of an arithmetic depth 3 circuit (1) which computes either the determinant, either the permanent or the Hamiltonian cycles polynomial treated as a function : $(F^*)^{n^2} \to F$. Also the lower bounds for some other explicitly given functions will be proved.

Although, we are interested mainly in the case $D = F^*$, the next lemma is valid for any $D \subset F$, $d = |D|$ such that the algebra of $\mathcal{A}$ of function $D^n \to F$ has a multiplicative basis of monomials $\mathcal{M} = \{X_1^{i_1} \cdots X_n^{i_n}\}_{0 \leq i_1, \ldots, i_n < d}$ (see definition 1 in section 2). One can rephrase definition 5 using the (Hamming metric ρ) geometric language in $\mathcal{M}$ (cf. section 2): if a polynomial f is (t, r)-sparse then its support $\mathrm{supp}\,(f) \subset \mathcal{M}$ (throughout this section the support is the set of monomials from $\mathcal{M}$ occurring in the polynomial) lies in a union of t balls each of a radius at most r (centered at X^{I_η}).

The following lemma provides a lower bound on the number of nonzeroes of a (t, r)-sparse polynomial.

Lemma 6 *Let the algebra $\mathcal{A}$ of functions : $D^n \to F$ have a multiplicative basis $\mathcal{M}$ and for a certain R the support $\mathrm{supp}\,(f)$ of a (t, r)-sparse polynomial $f \in \mathcal{A}$ (where $t \geq n$) contain a monomial $X^{I^{(0)}}$ such that $\rho(X^{I^{(0)}}, X^I) \geq R$ for any other monomial $X^{I^{(0)}} \neq X^I \in \mathrm{supp}\,(f)$. Then*

$$Pr[x \in D^n : f(x) \neq 0] \geq \exp\left(-O\left(\frac{n}{R^2}(r + n^{4/3}\log^{2/3} t)\right)\right) \qquad (4)$$

Remark 1 *The imposed inequality $t \geq n$ is not essentially restrictive, because when $t < n$ one could treat f as (n, r)-sparse polynomial. Thus, one can replace in (4) the occurrence of t by $t + n$ getting rid of the restriction $t \geq n$.*

Proof of the lemma. Replacing f by $f \cdot (a\, X^{I^{(0)}})^{-1}$ where $a \in F^*$ is the coefficient at the monomial $X^{I^{(0)}}$ in f (and taking into account that $\mathcal{M}$ is a multiplicative basis of $\mathcal{A}$), we assume f is still (t, r)-sparse, $1 \in \mathrm{supp}\,(f)$ and for any $1 \neq X^I \in \mathrm{supp}\,(f)$ we have $\rho(1, X^I) \geq R$. We keep the notation X^{I_η} for the centers of the balls. One can deem that there are $(t-1)$ balls centered at X^{I_η} and the monomial 1 lies in its own ball centered at 1. Taking any I_η and $X^I \in \mathrm{supp}\,(f)$ which lies in the ball of a radius at most r centered at X^{I_η}, we obtain that $\rho(1, X^{I_\eta}) \geq \rho(1, X^*) - \rho(X^*, X^{I_\eta}) \geq R - r$.

Put

$$s = \frac{Cn}{R}(r + n^{2/3} \log^{1/3} t) \tag{5}$$

for an appropriate sufficiently large constant C which will be specified later. Consider the sphere $S \subset \mathcal{M}$ of the radius $\frac{d-1}{d}n - s$ centered at 1 (w.l.o.g. we can assume that $d|n$), i.e. $S = \{X^J : \rho(1, X^J) = \frac{d-1}{d}n - s\}$. Since

$$|S| = (d-1)^{\frac{d-1}{d}n - s} \left(\begin{array}{c} n \\ \frac{d-1}{d}n - s \end{array} \right) \geq d^n \exp\left(-O\left(\frac{s^2}{n}\, \frac{d^2}{d-1} + \log n \right) \right),$$

we notice that the probability

$$Pr[X^J \in \mathcal{M} : X^J \in S] \geq \exp\left(-O\left(\frac{s^2}{n} \right) \right),$$

and the right side of the latter inequality has the same order of growth as the right side of the desired inequality (8).

W.l.o.g. we can assume that $s < \frac{d-1}{d}n$, because otherwise (4) is trivial.

Let us view a polynomial from $\mathcal{A}$ as a row of its d^n coefficients at the monomials from $\mathcal{M}$. We suppose to prove that one can pick out at least half of the elements X^J from S such that the matrix composed of the rows $X^J f$ for these $X^J \in S$ contains the unit submatrix just in the set $\mathcal{J}$ of columns X^J. That means that the dimension of the ideal $(f) \subset \mathcal{A}$ is greater or equal to $|S|/2$. Then lemma 1b) (see section 2) would imply (4) due to the obtained above bound on $|S|$.

We call $X^J \in S$ *remote* if $\rho(X^J, X^{I_\eta}) > \frac{d-1}{d}n - s + r$ for all η. Observe that if we compose the above matrix of the rows $X^J f$ for all remote X^J then

it contains the desired unit submatrix. Indeed, any monomial $1 \neq X^I \in \mathrm{supp}\,(f)$ belongs to a ball centered at X^{I_η} for a certain η, with the radius r. Therefore,

$$\rho(X^J, X^I) \geq \rho(X^J, X^{I_\eta}) - \rho(X^{I_\eta}, X^I) > \frac{d-1}{d}n - s.$$

Hence for any $X^{J_0} \in S$ (even not necessary to be remote) we have

$$\rho(X^J, X^{J_0}X^I) \geq \rho(X^J, X^I) - \rho(X^I, X^{J_0}X^I) = \rho(X^J, X^I) - \left(\frac{d-1}{d}n - s\right),$$

the latter equality invokes the property 2) from definition 1 (see section 2). Thus, $\rho(X^J, X^{J_0}X^I) > 0$ which means that the row $X^{J_0}f$ can not have a nonzero entry in any column X^J for a *remote* X^J (see again definition 1), except for appearance of an entry equal to 1 in the column X^{J_0} (the latter appearance makes sense only if $X^{J_0} \in \mathcal{J}$, i.e., X^{J_0} is remote).

In order to justify the remaining goal, i.e., to show that at least half of the elements $X^J \in S$ are remote, it suffices to prove for every η that the probability

$$Pr\left[X^J \in S : \rho(X^J, X^{I_\eta}) \leq \frac{d-1}{d}n - s + r\right] \leq \frac{1}{2t} \tag{6}$$

Notice that one can assume that $R \geq 3r$, otherwise (4) becomes trivial due to the occurrence of the first term $\frac{nr^2}{R^2}$.

Thus, we fix I_η and denote by $U_0, \ldots, U_{d-1}$ the partition of the set $\{1, \ldots, n\}$ where U_j consists of all i such that i-th coordinate of the vector I_η equals to j. Denote $w_j = |U_j|$, $0 \leq j \leq d-1$. Introduce independent random variables $Y_1, \ldots, Y_n$ each taking the value 0 with the probability $\frac{1}{d} + \frac{s}{n}$ and every value among $1, \ldots, d-1$ with the probability $\frac{1}{d} - \frac{s}{(d-1)n}$.

Denote $\delta = \frac{1}{6}\,\frac{sR}{n^2}$ and $w = \frac{1}{6d}\,\frac{sR}{n}$. We say that U_j is large if $w_j \geq w$. Fix a large U_j for the time being, denote by m_j the number of Y_i among $i \in U_j$ such that $Y_i = j$. Then Chernoff's inequality (see, e.g. [8]) states that

$$Pr\left[\left|\frac{m_j}{w_j} - \left(\frac{1}{d} - \frac{s}{(d-1)n}\right)\right| \geq \delta\right] \leq \exp(-\Omega(\delta^2 w_j)) \leq \exp(-\Omega(\delta^2 w)) \tag{7}_j$$

in case when $j \geq 1$ and

$$Pr\left[\left|\frac{m_0}{w_0} - \left(\frac{1}{d} + \frac{s}{n}\right)\right| \geq \delta\right] \leq \exp(-\Omega(\delta^2 w_0)) \leq \exp(-\Omega(\delta^2 w)) \tag{7}_0$$

in case when $j = 0$

where the probabilities are taken over the random variables $Y_1, \dots, Y_n$.

We claim that one can achieve that

$$Pr\left[n - (m_0 + m_1 + \cdots + m_{d-1}) \le \frac{d-1}{d} n - s + r\right] \le \exp(-C_0 \log t) \quad (8)$$

for an arbitrary constant $C_0 > 0$ (by means of suitable specifying a constant C in (5)).

To estimate the sum $n - (m_0 + m_1 + \cdots + m_{d-1})$ (being the Hamming distance between the monomials $X_1^{Y_1} \cdots X_n^{Y_n}$ and X^{I_η}), we handle separately the case of a large U_j applying the inequality $m_j \le \left(\frac{1}{d} - \frac{s}{(d-1)n}\right) w_j + \delta w_j$ when $j \ge 1$ (see $(7)_j$) or the inequality $m_0 \le \left(\frac{1}{d} + \frac{s}{n}\right) w_0 + \delta w_0$ when $j = 0$ (see $(7)_0$), and the case of a small U_j in which we simply apply the inequality $m_j \le w_j \le w$. Hence (with a probability greater than $1 - \exp(-\Omega(\delta^2 w))$) the following inequality is true:

$$n - (m_0 + m_1 \cdots + m_{d-1}) \ge \frac{d-1}{d} n - \frac{w_0 s}{n} + \frac{(w_1 + \cdots + w_{d-1})s}{(d-1)n} - \delta n - dw.$$

Since $n - w_0 = w_1 + \cdots + w_{d-1} = \rho(1, X^{I_\eta}) \ge R - r \ge \frac{2}{3} R$ (see the beginning of the proof of the lemma), we obtain

$$n - (m_0 + m_1 + \cdots + m_{d-1}) \ge \frac{d-1}{d}\, n - s + \frac{2}{3}\, \frac{d}{d-1}\, \frac{Rs}{n} - \delta n - dw.$$

Taking into account that $\frac{2}{3}\frac{Rs}{n} \ge 4\delta n$, $\frac{2}{3}\frac{Rs}{n} \ge 4dw$, $\frac{2}{3}\frac{Rs}{n} \ge 4r$ (the latter inequality can be secured by choosing a sufficiently large constant C in (5)) we deduce that the required in (8) inequality $n - (m_0 + m_1 + \cdots m_{d-1}) \le \frac{d-1}{d}\, n - s + r$ could be true with a probability less than $\exp(-\Omega(\delta^2 w))$. Again increasing a constant C in (5) if necessary, we can achieve that $\delta^2 w \ge C_1 \log t$ and thereby, prove (8).

Now to complete the proof of (6) we denote for brevity the event that among random values $Y_1, \dots, Y_n$ there are exactly $\frac{n}{d} + s$ zeroes, i.e. the monomial $X_1^{Y_1} \cdots X_n^{Y_n}$ belongs to the sphere S, by $(Y_1, \dots, Y_n) \in S$. Then

$$Pr\left[n - (m_0 + m_1 + \cdots + m_{d-1}) \le \frac{d-1}{d}\, n - s + r\right] \ge$$

$$Pr[(Y_1, \dots, Y_n) \in S] \times$$

$$Pr\left[n-(m_0+m_1+\cdots+m_{d-1})\le\frac{d-1}{d}n-s+r|(Y_1,\ldots,Y_n)\in S\right]$$

$$=Pr\left[(Y_1,\ldots,Y_n)\in S\right]Pr\left[X^J\in S:\rho(X^J,X^{I_\eta})\le\frac{d-1}{d}n-s+r\right]\ge$$

$$\frac{1}{n+1}Pr\left[X^J\in S:\rho(X^J,X^{I_\eta})\le\frac{d-1}{d}n-s+r\right], \qquad (9)$$

the latter inequality follows from the observation that the radius $\frac{d-1}{d}n-s$ of the sphere S is the expectation of radii of $(n+1)$ possible spheres in which the monomial $X_1^{Y_1}\cdots X_n^{Y_n}$ could lie, and the probability for the monomial to lie in a sphere increases along with the radius of the sphere till $\frac{d-1}{d}n-s$ and after that decreases.

Together with (8) for an appropriate constant C_0 the inequality (9) implies (6), taking into account that $t\ge n$, the lemma is proved.

From now on till the end of the paper we assume that $D=F^*$ (cf. the end of the previous section).

Denote by $K=\mathbb{Z}/(q-1)$ the ring of residues. Recall that for a subset $U\subset K^n$ of the free K-module its *coding distance* $m=m(U)$ is defined as the minimum of $\rho(u_1,u_2)$ over all pairs $u_1\ne u_2$, $u_1,u_2\in U$. Denote by $f_U=\sum_{u\in U}X^u:(F^*)^n\to F$ the sum of the monomials with the exponents from U.

Theorem 3 *There exists $\epsilon>0$ such that if the number of the elements $n\le t=|U|\le\exp(\epsilon\, m^6/n^5)$ then for any depth 3 arithmetic circuit (1) computing $f_U:(F^*)^n\to F$ the complexity lower bound $N\ge t$ holds.*

Proof. Suppose the contrary. Set $r=\epsilon_1 m^2/n$ for a pertinent small enough $\epsilon_1>0$ which will be specified later. Applying to f_U proposition 1, provided that $\epsilon<\delta_0\epsilon_1$, we obtain a $(t-1,C_1r)$-sparse polynomial g which approximates f_U:

$$Pr[x\in(F^*)^n:f_U(x)\ne g(x)]\le\exp(-\delta_1 r)$$

for certain $C_1,\delta_1>0$.

Observe that $(2t-1,C_2r)$-sparse polynomial f_U-g contains a monomial $X^{I^{(0)}}\in\mathcal{M}$ from the support $\mathrm{supp}\,(f_U)$ such that $\rho(X^{I^{(0)}},X^I)\ge m/4$ for any $X^{I^{(0)}}\ne X^I\in\mathrm{supp}\,(f_U-g)$, provided that $\epsilon_1<\frac{1}{4C_1}$. Indeed, denote by $X^{I_1},\ldots,X^{I_{t-1}}$ the centers of the balls of radii at most C_1r which cover the support $\mathrm{supp}\,(g)$. Then at least for one of the monomials $X^{I^{(0)}}\in\mathrm{supp}\,(f_U)$ we

have $\rho(X^{I^{(0)}}, X^{I_i}) \geq m/2$ for $1 \leq i \leq t-1$. Hence for any monomial X^I from the ball of the radius $C_1 r \leq m/4$ centered at X^{I_i} we have $\rho(X^{I_0}, X^I) \geq m/4$.

Now we are able to apply lemma 6 to the polynomial $f_U - g$ with $R = m/4$ which gives the bound

$$Pr[x \in (F^*)^n : f_U(x) \neq g(x)] \geq \exp\left(-C_2 r\left(\epsilon_1 + \frac{\epsilon^{2/3}}{\epsilon_1}\right)\right)$$

for a suitable constant C_2 appearing from the right side of (4). Now specifying first $\epsilon_1 < \frac{\delta_1}{3C_2}$ and thereupon $\epsilon < \left(\frac{\epsilon_1 \delta_1}{3C_2}\right)^{3/2}$ (and also satisfying $\epsilon < \delta_0 \epsilon_1$, see above), we get a contradiction which proves the theorem.

Take any prime $p_0 | (q-1)$ and consider a linear code $U_0 \subset (GF(p_0))^n$ with a basis $u_1, \ldots, u_k$ and the coding distance m. Then the subset $U = \frac{(q-1)}{p_0} U_0 \subset K^n$ obtained from U_0 by multiplying every its element by $\frac{q-1}{p_0} \in K$, has also the coding distance m. Moreover, the following formula is valid:

$$f_U = \prod_{1 \leq i \leq k} \left(\sum_{0 \leq j \leq p_0 - 1} X^{j \frac{q-1}{p_0} u_i} \right) : (F^*)^n \to F \tag{10}$$

Using Gilbert-Varshamov bound from the coding theory [9], which supplies us with a linear code U of the coding distance $m \geq \delta_0 n$ for appropriate $\delta_0 > 0$ and the dimension $k = \epsilon_0 n$, where $0 < \epsilon_0 < \epsilon \delta_0^6$ for the constant ϵ from theorem 3, we get the following corollary.

Corollary 2 *There exists a linear code $U_0 \subset (GF(p_0))^n$ such that for any depth 3 arithmetic circuit (1) computing $f_U : (F^*)^n \to F$ the exponential lower bound on its complexity $N \geq \exp(\Omega(n))$ holds.*

If we would like to stick with explicitly constructed functions, we can consider a BCH code [9] with the coding distance $m \geq \Omega(n/\log n)$, which has a linear dimension $\Omega(n)$, and as $U_0^{(\mathrm{BCH})}$ we take an arbitrary its linear subcode (i.e. a subspace) of the dimension $\Omega(n/\log^6 n)$. Again applying theorem 3, we get the following corollary.

Corollary 3 *For an explicitly constructed function $f_{U^{(BCH)}} : (F^*)^n \to F$ (obtained from a BCH code) we have the exponential lower bound $N \geq \exp(\Omega(n/\log^6 n))$ on the complexity of any computing it depth 3 arithmetic circuit (1).*

Finally, based on formula (10) which has a linear size $O(n)$ and applying the construction from [15] (see also [2]) to the function f_U in the algebra $\mathcal{A}$, we conclude with the following corollary.

Corollary 4 *For each of the following three functions* : $(F^*)^{n^2} \to F$

a) Determinant $\sum_{\pi \in S_n} (-1)^{sgn(\pi)} \prod_{1 \le i \le n} X_{i,\pi(i)}$*;*

b) Permanent $\sum_{\pi \in S_n} \prod_{1 \le i \le n} X_{i,\pi(i)}$*;*

c) Hamiltonian cycles polynomial $\sum_{\pi} \prod_{1 \le i \le n} X_{i,\pi(i)}$*, where the latter summation is taken over all permutations* π *which consist of a single cycle, any computing it depth 3 arithemtic circuit (1) has the exponential complexity* $N \ge \exp(\Omega(n))$*.*

References

[1] M.Agrawal, E.Allender, S.Datta, On TC^0, AC^0, and arithmetic circuits. ECCC Report TR97-016.

[2] J. von zur Gathen. Feasible arithmetic computations: Valiant's hypothesis. J. Symp. Comput., 4, 1987, p. 137-172.

[3] D. Grigoriev. Lower bounds in algebraic complexity. J. Soviet Math., 29, 1985, p. 1388–1425.

[4] D. Grigoriev. Testing shift-equivalence of polynomials by deterministic, probabilistic and quantum machines. Theor. Comput. Sci., 180, 1997, p. 217–228.

[5] D. Grigoriev, M. Karpinski. An exponential lower bound for depth 3 arithmetic circuits. Proc. ACM Symp. Th. Comput., Dallas, 1998.

[6] J. Hastad, M. Goldmann. On the power of small-depth threshold circuits. Comput. Complexity, 1, 1991, p. 113–129.

[7] M. Krause, S. Waack. Variation ranks of communication marices and lower bounds for depth two circuits having symmetric gates with unbounded fan-in. Proc. IEEE Symp. Found. Comput. Sci., 1991, p. 777–782.

[8] E. Kushilevitz, N. Nisan. Communication complexity. Cambridge University Press, 1997.

[9] F. MacWilliams, N. Sloane. The theory of error-correcting codes. North-Holland, 1977.

[10] N. Nisan, A. Wigderson. Lower bound on arithmetic circuits via partial derivatives. Proc. IEEE Symp. Found. Comput. Sci., 1995, p. 16–25.

[11] A. Razborov. Lower bounds on the size of bounded depth ciruits over a complete basis with logical addition. Math. Notes, 41, 1987, p. 333–338.

[12] A. Razborov. On small depth threshold circuits. Lect. Notes Comput. Sci., 621, 1992, p. 42–52.

[13] A. Razborov, A. Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth 3 threshold circuits with AND gates at the bottom. Inform. Proc. Lett., 45, 1993, p. 303–307.

[14] R. Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. Proc. ACM Symp. Th. Comput., 1987, p. 77–82.

[15] L. Valiant. Completeness classes in algebra. Proc. ACM Symp. Th. Comput. 1979, p. 259–261.

[16] A. Yao. On ACC and threshold circuits. Proc. IEEE Symp. Found. Comput. Sci., 1990, p. 619–627.