Lower Bounds for Matrix Product

Amir Shpilka
Institute of Computer Science
Hebrew University
Jerusalem, Israel
amirs@cs.huji.ac.il

Abstract

We prove lower bounds on the number of product gates in bilinear and quadratic circuits that compute the product of two $n \times n$ matrices over finite fields. In particular we obtain the following results:

1. *We show that the number of product gates in any bilinear (or quadratic) circuit that computes the product of two $n \times n$ matrices over $GF(2)$ is at least $3n^2 - o(n^2)$.*
2. *We show that the number of product gates in any bilinear circuit that computes the product of two $n \times n$ matrices over $GF(p)$ is at least $(2.5 + \frac{1.5}{p^3-1})n^2 - o(n^2)$.*

These results improve the former results of [3, 1] who proved lower bounds of $2.5n^2 - o(n^2)$.

1. Introduction

The problem of computing the product of two matrices is one of the most studied computational problems: We are given two $n \times n$ matrices $x = (x_{i,j})$, $y = (y_{i,j})$, and we wish to compute their product, i.e. there are n^2 outputs where the (i, j)'th output is

$$(x \cdot y)_{i,j} = \sum_{k=1}^{n} x_{i,k} \cdot y_{k,j} \ .$$

In 69' Strassen surprised the world by showing an upper bound of $O(n^{\log_2 7})$ [12]. This bound was later improved and the best upper bound today is $O(n^{2.376})$ [5] (see also [7] for a survey). The best lower bound is a lower bounds of $2.5n^2 - o(n^2)$, on the number of products needed to compute the function [3, 1]. Thus the following problem is still open: Can matrix product be computed by a circuit of size $O(n^2)$?

The standard computational model for computing polynomials is the model of arithmetic circuits, i.e. circuits over the base $\{+, \cdot\}$ over some field F. This is indeed the most general model, but for matrix product two other models are usually considered, *quadratic circuits* and *bilinear circuits*. In the quadratic model we require that product gates are applied only on two linear functions. In the bilinear model we also require that product gates are applied only on two linear functions, but in addition we require that the first linear function is linear in the variables of x and that the second linear function is linear in the variables of y. These models are more restricted than the general model of arithmetic circuits. However it is interesting to note that over infinite fields we can always assume w.l.o.g. that any circuit for matrix product is a quadratic circuit [13]. In addition we note that the best circuits that we have today for matrix product are bilinear circuits.

In this paper we prove that any *quadratic* circuit that computes matrix product over the field $GF(2)$ has at least $3n^2 - o(n^2)$ product gates, and that any *bilinear* circuit for matrix product over the field $GF(p)$ must have at least $(2.5 + \frac{1.5}{p^3-1})n^2 - o(n^2)$ product gates.

From now on we will use the notation MP_n to denote the problem of computing the product of two $n \times n$ matrices.

1.1. Known Lower Bounds

In contrast to the major advances in proving upper bound, the attempts to prove lower bounds on the size of bilinear circuits that compute MP_n were less successful. Denote by $q_*(MP_n)$ and $bl_*(MP_n)$ the number of product gates in a smallest quadratic circuit for MP_n, and in a smallest bilinear circuit for MP_n respectively. We also denote by $bl_{tot}(MP_n)$ the total number of gates in a smallest bilinear circuit for MP_n. In 78' Brocket and Dobkin proved that $bl_*(MP_n) \geq 2n^2 - 1$ over any field [10]. This lower bound was later generalized by Lafon and Winograd to a lower bound on $q_*(MP_n)$ over any field [8]. In 89' Bshouty showed that over $GF(2)$, $q_*(MP_n) \geq 2.5n^2 - O(n \log n)$ [3]. Recently Bläser proved a lower bound of $2n^2 + n - 3$ on $q_*(MP_n)$ over any field [2]. In [1] Bläser proved that

$bl_*(MP_n) \geq 2.5n^2 - 3n$ over any field.

In [9] it is shown that any bounded depth circuit for MP_n, over any field, has a super linear (in n^2) size. Notice however, that the best known circuits for MP_n have depth $\Omega(\log n)$.

1.2. Bilinear Rank

An important notion that is highly related to the problem of computing matrix product in bilinear circuits is the notion of bilinear rank.

A bilinear form in two sets of variables x, y is a polynomial in the variables of x and the variables of y, which is linear in the variables of x and linear in the variables of y. Clearly each output of MP_n is a bilinear form in $x = \{x_{i,j}\},\ y = \{y_{i,j}\}$. The bilinear rank of a set of bilinear forms $\{\ b_1(x,y),\ \ldots,\ b_m(x,y)\ \}$ is the smallest number of rank 1 bilinear forms that span $b_1,\ \ldots,\ b_m$, where a rank 1 bilinear form is a product of a linear form in the x variables and a linear form in the y variables. We denote by $R_F(b_1,\ \cdots,\ b_m)$ the bilinear rank of $\{\ b_1,\ \ldots,\ b_m\ \}$ over the field F. For further background see [4, 7].

We denote by $R_F(MP_n)$ the bilinear rank over F of the n^2 outputs of matrix product, i.e. it is the bilinear rank of the set $\{\sum_{k=1}^{n} x_{i,k} \cdot y_{k,j}\}_{i,j}$ over F.

The following inequalities are obvious (over any field).

- $q_*(MP_n) \leq bl_*(MP_n) \leq 2q_*(MP_n)$.
- $R_F(MP_n) = bl_*(MP_n)$.
- The following inequality is less obvious, but also not so hard to see.

  $$bl_*(MP_n) \leq bl_{tot}(MP_n) \leq \mathrm{poly}(\log n) \cdot bl_*(MP_n)\ .$$

  I.e. up to polylogarithmic factors, the number of product gates in a smallest bilinear circuit for MP_n, over any field F, is equal to the total number of gates in the circuit.

1.3. Results and Methods

We prove that any *quadratic* circuit that computes MP_n over the field $GF(2)$ has at least $3n^2 - o(n^2)$ product gates (i.e. $q_*(MP_n) \geq 3n^2 - o(n^2)$ over $GF(2)$). We also prove that over the field $GF(p)$ every *bilinear* circuit for MP_n must have at least $(2.5 + \frac{1.5}{p^3-1})n^2 - o(n^2)$ product gates (i.e. $bl_*(MP_n) \geq (2.5 + \frac{1.5}{p^3-1})n^2 - o(n^2)$ over $GF(p)$). Both of these results actually hold for the bilinear rank as well.

The proof of the lower bound over $GF(2)$ is based on techniques from the theory of linear codes. However, we cannot use known results from coding theory in a straightforward way, since we are not dealing with codes in which every two words are distant, but rather with codes on matrices in which the distance between two code words, of two matrices, is proportional to the rank of the difference of the matrices. The reduction from circuits to codes and the proof of the bound are given in section 4.

The proof of the second bound is based on a lemma proved by Bläser in [1]. We prove that in the case of finite fields we can use the lemma with better parameters than those used by Bläser. This result is proved in section 5.

1.4. Organization of the paper

In section 2 we present the models of bilinear circuits and quadratic circuits. In section 3 we present some algebraic and combinatorial tools that we need for the proofs of our lower bounds.

In section 4 we introduce the notion of linear codes of matrices, and prove our lower bound on bilinear and quadratic circuits that compute MP_n over $GF(2)$. In section 5 we prove our lower bound on bilinear circuits that compute MP_n over $GF(p)$.

2. Arithmetic Models

In this section we present the models of quadratic circuits and bilinear circuits. These are the models for which we prove our lower bounds. We first give the definition of a general arithmetic circuit. An arithmetic circuit over a field F is a directed acyclic graph as follows. Nodes of in-degree 0 are called inputs and are labeled with input variables. Nodes of out-degree 0 are called outputs. Each edge is labeled with a constant from the field and each node other than an input is labeled with one of the following operations $\{\ +\ ,\ \cdot\ \}$, in the first case the node is a plus gate and in the second case a product gate. The computation is done in the following way. An input just computes the value of the variable that labels it. Then, if $v_1,\ \ldots,\ v_k$ are the vertices that fan into v then we multiply the result of each v_i with the value of the edge that connects it to v. If v is a plus gate we sum all the results, otherwise v is a product gate and we multiply all the results. Obviously the value computed by each node in the circuit is a polynomial over F in the input variables.

We are interested in the problem of computing the product of two $n \times n$ matrices, MP_n. The input consists of two $n \times n$ matrices $x,\ y$. The output is the matrix $x \cdot y$, i.e., there are n^2 outputs, and the (i,j)'th output is:

$$(x \cdot y)_{i,j} = \sum_{k=1}^{n} x_{i,k} \cdot y_{k,j}\ .$$

Each output $(x \cdot y)_{i,j}$ is hence a bilinear form in x and y.

Since each output of MP_n is a bilinear form, it is natural to consider bilinear arithmetic circuits for it. A bilinear arithmetic circuit is an arithmetic circuit with the additional restriction that product gates are applied only on two linear functions, one function is linear in the variables of x and the other function is linear in the variables of y. Thus, bilinear circuits have the following structure. First, there are many plus gates computing linear forms in x and linear forms in y. Then there is one level of product gates that compute bilinear forms, and finally there are many plus gates that eventually compute the outputs. We will be interested in bounding from below the number of products in any bilinear circuit for MP_n. This model is more restricted than the general model of arithmetic circuits but we note that all the known upper bounds (over any field) for MP_n are by bilinear circuits.

Another model that we will consider is the model of quadratic circuits. A quadratic circuit is an arithmetic circuit with the additional restriction that product gates are applied only on two linear functions. Notice that the only difference between quadratic circuits and bilinear circuits is that in the quadratic model the product gates compute *quadratic forms* in $x,\ y$, whereas in the bilinear model the product gates compute *bilinear forms* in $x,\ y$. This model is more general than the model of bilinear circuits, but it is still more restricted than the general model. However it is interesting to note that over infinite fields we can assume w.l.o.g. that any arithmetic circuit for MP_n is a quadratic circuit [13].

3. Algebraic and Combinatorial tools

In this section we present some algebraic and combinatorial tools that we will use.

The following lemma is an extremely weak variant of the famous Schwartz-Zippel lemma which shows that every non zero polynomial (non zero as a formal expression) over a large enough field has a non zero assignment in the field (see [11, 15]).

Lemma 1 *Let P be a polynomial of degree d in $x_1,\ \ldots,\ x_n$ over some field F, such that $d < |F|$, and such that at least one of the coefficients of P is not zero. Then we can find an assignment, $\rho \in F^n$, to the x_i's, such that $P(\rho_1, \ldots, \rho_n) \neq 0$.*

We say that two polynomials p, q in n variables are equivalent over a field F, if $p(x_1, \ldots, x_n) = q(x_1, \ldots, x_n)$ for any $x_1,\ \ldots,\ x_n \in F$. We denote $p \equiv q$ if p and q are equivalent over F (we omit F from the notation as the field that we deal with will be clear from the context).

Lemma 2 *Let P be a polynomial of degree d in the variables $x_1,\ \ldots,\ x_n$ over a field F. If $P \not\equiv 0$ then we can find an assignment, $\rho \in F^n$, to the x_i's such that at most d of the ρ_i's get a nonzero value, and such that $P(\rho_1, \ldots, \rho_n) \neq 0$.*

Proof: P is equal (as a function) to a polynomial $\bar{P}$ in which the degree of each variable is at most $|F| - 1$. We call $\bar{P}$ the reduction of P. Consider some monomial M in $\bar{P}$ whose coefficient is not zero. We assign all the variables that do not appear in M to zero. The resulting polynomial (after the assignment), is a polynomial in the variables of M, which is not the zero polynomial as it is a reduced polynomial which has a monomial with a non zero coefficient (M of course). Therefore according to lemma 1 there is some assignment to the variables of M, that gives this polynomial a nonzero value. Therefor we have found an assignment which gives nonzero values only to the variables of M (and there are at most d such variables) under which $P \neq 0$. ♣

The following useful lemma, which is a straightforward implication of the previous lemma, is the key lemma in most of our proofs. The lemma deals with linear forms in n^2 variables. From now on we shall think about such linear forms as linear forms in the entries of $n \times n$ matrices.

Lemma 3 *Let $\mu_1,\ \ldots,\ \mu_{n^2}$ be n^2 linearly independent linear forms in n^2 variables over some field F. Let P be a polynomial of degree d in kn^2 variables over F, i.e. we can view P as a polynomial $P(x_1, ..., x_k)$ in the entries of k matrices, $x_1, ..., x_k$, of size $n \times n$ each. Assume that $P \not\equiv 0$. Then we can find k matrices $a_1,\ \ldots,\ a_k \in M_n(F)$ such that $P(a_1, ..., a_k) \neq 0$ and such that there exist $n^2 - d$ linear forms among $\mu_1, ..., \mu_{n^2}$'s that vanish on all the a_i's.*

Proof: The idea of the proof is the following. Let $b_1,\ \ldots,\ b_{n^2}$ be the dual basis of $\mu_1, ..., \mu_{n^2}$, i.e. it is a basis of $M_n(F)$ satisfying $\forall i, j \quad \mu_i(b_j) = \delta_{i,j}$. We wish to find k matrices, $a_1, ..., a_k$, such that $P(a_1, ..., a_k) \neq 0$, and such that there exist $b_{i_1}, ..., b_{i_d}$ that span all of them. If we manage to find such matrices, then since the b_i's are the dual basis to the μ_i's we will get that $n^2 - d$ of the μ_i's vanish on $a_1, ..., a_k$. The way to find such matrices that are contained in the span of a small subset of the b_i's, is based on lemma 2.

So let $b_1,\ \ldots,\ b_{n^2}$ be the dual basis to $\mu_1,\ \ldots,\ \mu_{n^2}$, i.e. $\forall i, j \ \ \mu_i(b_j) = \delta_{i,j}$. We now change the variables of P. Let $\alpha_{i,j} \quad j = 1...k,\ i = 1...n^2$, be a set of kn^2 variables. Denote $x_j = \sum_{i=1}^{n^2} \alpha_{i,j} b_i$. Thus $P(x_1, ..., x_k)$ can be viewed as a polynomial of degree d in the kn^2 variables $\alpha_{i,j}$. Therefore $P \not\equiv 0$ as a polynomial in the $\alpha_{i,j}$'s. Hence, according to lemma 2 there exists an assignment, ρ, to the $\alpha_{i,j}$'s such that at most d of them get a nonzero value. Define $a_j = \sum_{i=1}^{n^2} \rho_{i,j} b_i$. Clearly $P(a_1, \ldots, a_k) \neq 0$. Since at most d of the $\rho_{i,j}$'s got non zero values, we see that there are at most d b_i's such that all the a_j's are linear combinations of them. Since the b_i's are the dual basis to $\mu_1, \ldots, \mu_{n^2}$ we get that there are at least $n^2 - d$ of the μ_i's that vanish on all the a_j's. Therefore $a_1, \ldots, a_k$ satisfy the requirements of the lemma. ♣

The next lemma will enable us to translate properties of matrices over large fields of characteristic p to properties of matrices (of higher dimension) over $GF(p)$.

Lemma 4 *There exist an embedding, $\phi : GF(p^n) \hookrightarrow M_n(GF(p))$. That is there exist a mapping $\phi : GF(p^n) \mapsto M_n(GF(p))$ such that*

- *ϕ is a one to one linear transformation.*
- *$\phi(1) = I$, where I is the $n \times n$ identity matrix.*
- *ϕ is multiplicative, i.e. $\forall x, y \in GF(p^n)$ we have that $\phi(xy) = \phi(x) \cdot \phi(y)$.*

This embedding also induces an embedding $M_k(GF(p^n)) \hookrightarrow M_{nk}(GF(p))$.

This lemma is a standard tool in algebra, but for completeness we give the proof.

Proof: $GF(p^n)$ is an n dimensional vector space over $GF(p)$. Each element $x \in GF(p^n)$ can be viewed as a linear transformation $x : GF(p^n) \mapsto GF(p^n)$ in the following way:

$$\forall y \in GF(p^n) \;\; x(y) = x \cdot y \; .$$

Clearly this is a linear transformation of $GF(p^n)$ into itself, as a vector space over $GF(p)$. Therefore, by picking a basis to $GF(p^n)$ we can represent the linear transformation corresponding to each $x \in GF(p^n)$ by a matrix $a_x \in M_n(GF(p))$. Thus, we have defined a mapping $\phi : GF(p^n) \mapsto M_n(GF(p))$ such that $\phi(x) = a_x$, and it is easy to verify that this mapping is an embedding of $GF(p^n)$ into $M_n(GF(p))$. The way to generalize it to an embedding of $M_k(GF(p^n))$ into $M_{nk}(GF(p))$ is the following. Let $a = (a_{i,j}) \in M_k(GF(p^n))$ be some matrix. Every entree of $a_{i,j}$ of a, is some element of $GF(p^n)$. We can now replace $a_{i,j}$ with the matrix $\phi(a_{i,j})$. Thus the resulting matrix will be a $kn \times kn$ matrix whose entries are in $GF(p)$. Again it is easy to verify that this is indeed an embedding of $M_k(GF(p^n))$ into $M_{nk}(GF(p))$. ♣

In addition to the algebraic lemmas we also need the following combinatorial tools.

Definition 1 *Let F be a field, and let v, u be two vectors in F^m. We denote by $\mathbf{weight}(v)$ the number of nonzero coordinates of v. Let $\mathbf{d_H}(v,u) = \mathbf{weight}(v-u)$, i.e. $\mathbf{d_H}(v,u)$ is the number of coordinates on which u and v differ. $\mathbf{d_H}(v,u)$ is also known as the Hamming distance of u and v. We also denote by $\mathbf{agree}(u,v)$ the number of coordinates on which u and v are equal, i.e. $\mathbf{agree}(u,v) = m - \mathbf{d_H}(v,u)$.*

The next lemma shows that if a vector space contains a set of vectors such that every pair/triplet of them don't agree on many coordinates (i.e. their Hamming distance is large) then it is of large dimension. There are numerous similar lemmas in coding theory, and in particular the first part of our lemma is the famous Plotkin bound (see [14]).

Lemma 5 *1. In every set of k vectors in $GF(p)^t$, such that $p < k$, there are two vectors that agree on at least $(\frac{t}{p} - \frac{t}{k})$ coordinates.*

2. In every set of k vectors in $GF(p)^t$, such that $2p < k$, there are three vectors that agree on at least $(\frac{t}{p^2} - \frac{3t}{pk})$ coordinates.

Proof: We begin by proving the first claim. Let $v_1, \ldots, v_k$ be k vectors in $GF(p)^t$. We are going to estimate $\sum_{i<j} \mathbf{agree}(v_i, v_j)$ in two different ways. On the one hand this sum is at most $\binom{k}{2}$ times the maximum of $\mathbf{agree}(v_i, v_j)$. On the other hand consider a certain coordinate. For every $\alpha \in GF(P)$ denote by n_α the number of vectors among the v_i's that are equal to α on this coordinate. Clearly $\sum_{\alpha=0}^{p-1} n_\alpha = k$. The contribution of this coordinate to $\sum_{i<j} \mathbf{agree}(v_i, v_j)$ is exactly $\sum_{\alpha=0}^{p-1} \binom{n_\alpha}{2}$. By convexity

$$\sum_{\alpha=0}^{p-1} \binom{n_\alpha}{2} \geq p \cdot \frac{1}{2} \cdot \frac{k}{p} \left(\frac{k}{p} - 1 \right) = \frac{k(k-p)}{2p} \; .$$

We get that

$$\binom{k}{2} \cdot \max_{i<j} (\mathbf{agree}(v_i, v_j)) \geq$$

$$\sum_{i<j} \mathbf{agree}(v_i, v_j) \geq t \cdot \frac{k(k-p)}{2p} \; .$$

Therefore

$$\max_{i<j} (\mathbf{agree}(v_i, v_j)) \geq \frac{t}{p} \cdot \frac{k-p}{k-1} \geq \frac{t}{p} \cdot \frac{k-p}{k} = \frac{t}{p} - \frac{t}{k} \; .$$

The proof of the second claim is similar. We give two different estimates to $\sum_{i<j<l} \mathbf{agree}(v_i, v_j, v_l)$ (the number of coordinates on which v_i, v_j, and v_l are the same). In the same manner as before we get that

$$\max_{i<j<l} (\mathbf{agree}(v_i, v_j, v_l)) \geq \frac{t}{p^2} \cdot \frac{k-p}{k-1} \cdot \frac{k-2p}{k-2} \geq \frac{t}{p^2} - \frac{3t}{pk} \; .$$

♣

Corollary 1 *If $\{0,1\}^t$ contains k vectors $v_1, \ldots, v_k$, such that $2 < k$ and $\forall i \neq j$ $\mathbf{d_H}(v_i, v_j) \geq N$, then $t \geq 2N - 4\frac{N}{k+2}$.*

Proof: According to lemma 5 there are two vectors, w.l.o.g. v_1 and v_2, such that $\mathbf{agree}(v_1, v_2) \geq \frac{t}{2} - \frac{t}{k}$. Since $\mathbf{d_H}(v_1, v_2) = t - \mathbf{agree}(v_1, v_2)$ we get that

$$t - (\frac{t}{2} - \frac{t}{k}) \geq \mathbf{d_H}(v_1, v_2) \geq N$$

and the result follows. ♣

4. Lower bound over GF(2)

In this section we prove our main theorems.

Theorem 1 $bl_*(MP_n) \geq 3n^2 - O(n^{\frac{5}{3}})$ *(in other words $R_{GF(2)}(MP_n) \geq 3n^2 - O(n^{\frac{5}{3}})$).*

The second theorem that we shall prove is a lower bound for quadratic circuits.

Theorem 2 $q_*(MP_n) \geq 3n^2 - O(n^{\frac{5}{3}})$. *I.e. the number of product gates in any quadratic circuit that computes the product of two $n \times n$ matrices over $GF(2)$ is at least $3n^2 - O(n^{\frac{5}{3}})$.*

Clearly theorem 2 imply theorem 1, but we first prove of theorem 1 as it is more intuitive and simple. We begin by introducing the notion of linear codes of matrices.

4.1. Linear Codes of Matrices

Definition 2 *A linear code of matrices is a mapping,*

$$\Gamma : M_n(GF(2)) \mapsto \{0,1\}^m ,$$

(for some m) with the following properties:

- *Γ is linear.*
- *For any matrix a, $\mathbf{weight}(\Gamma(a)) \geq n \cdot \mathrm{rank}(a)$.*

From the linearity of Γ and the requirement on $\mathbf{weight}(\Gamma(a))$ we get the following corollary.

Corollary 2 *Γ is a one to one mapping, and for any two matrices a and b, $\mathbf{d_H}(\Gamma(a), \Gamma(b)) \geq n \cdot \mathrm{rank}(a - b)$.*

The following theorem shows that the dimension of the range of any linear code of matrices is large (i.e. m must be large).

Theorem 3 *Let $\Gamma : M_n(GF(2)) \mapsto \{0,1\}^m$ be a linear code of matrices, then $m \geq 3n^2 - O(n^{\frac{5}{3}})$.*

Proof: Denote

$$\Gamma(a) = (\ \mu_1(a),\ \ldots,\ \mu_m(a)\ ) .$$

The proof is based on the following lemma that shows that we can find $k = n^{\frac{1}{3}}$ matrices, $a_1, \ldots, a_k \in M_n(GF(2))$, with the following properties.

- $\forall i \neq j$, $a_i - a_j$ is an invertible matrix.
- There are $n^2 - \binom{k}{2}n$ linear forms among the μ_i's that vanish on all the a_i's.

We state the lemma for every $k < 2^n$ but we apply it only to $k = n^{\frac{1}{3}}$.

Lemma 6 *For every n, k such that $k < 2^n$, and any $\mu_1, \ldots, \mu_{n^2}$ linearly independent linear forms in n^2 variables, over $GF(2)$, there are k matrices, $a_1, \ldots, a_k \in M_n(GF(2))$, such that for every $i \neq j$, $a_i - a_j$ is an invertible matrix, and such that $n^2 - \binom{k}{2}n$ of the μ_i's vanish on them.*

Proof: Consider the following polynomial P in k matrices:

$$P(a_1, \ldots, a_k) = \text{determinant}\left(\prod_{i<j}(a_i - a_j)\right) .$$

Clearly a set of k matrices $a_1, \ldots, a_k$ satisfy $P(a_1, \ldots, a_k) \neq 0$ iff all the matrices $a_i - a_j$ are invertible. In addition, it is easy to see that $deg(P) = \binom{k}{2}n$. Therefore if we show that $P \not\equiv 0$ over $GF(2)$, then according to lemma 3 we will get what we wanted to prove.

In order to show that $P \not\equiv 0$ we just have to prove the existence of k matrices, such that the difference of every two of them is invertible. Lemma 4 assures us that we can embed the field $GF(2^n)$ into $M_n(GF(2))$. Denote this embedding by $\Phi : GF(2^n) \hookrightarrow M_n(GF(2))$. We take k distinct elements in $GF(2^n)$, $x_1, \ldots, x_k$. Their images, $\Phi(x_1), \ldots, \Phi(x_k)$, are matrices in $M_n(GF(2))$ such that the difference of every two of them, $\Phi(x_i) - \Phi(x_j) = \Phi(x_i - x_j)$, is an invertible matrix. This is because the x_i's are distinct (i.e. $x_i - x_j \neq 0$), and every nonzero element in $GF(2^n)$ is invertible. Thus, $\Phi(x_1), \ldots, \Phi(x_k)$ are exactly the k matrices that we were looking for. This concludes the proof of the lemma. ♣

We proceed with the proof of the theorem. Let $k = n^{\frac{1}{3}}$. Since Γ is a one to one mapping, there are n^2 independent linear forms among $\mu_1, \ldots, \mu_m$. Therefore we can use lemma 6 and get that there are k matrices $a_1, \ldots, a_k$ such that for every $i \neq j$ $a_i - a_j$ is invertible, and such that, w.l.o.g., $\mu_{m-r+1}, \ldots, \mu_m$ vanish on $a_1, \ldots, a_k$ for some $r \geq n^2 - \binom{k}{2}n \geq n^2 - n^{\frac{5}{3}}$.

Since the last r linear forms vanish on all the a_i's, we are going to restrict our attention only to the first $m-r$ linear forms. So from now on we only consider $\Gamma(a_i)$ restricted to its first $m-r$ coordinates.

Since each of the differences, $a_i - a_j$ ($\forall i \neq j$), is an invertible matrix, we get that $\mathbf{d_H}(\Gamma(a_i), \Gamma(a_j)) \geq n^2$. Thus, $\Gamma(a_1), \ldots, \Gamma(a_k)$ are k vectors contained in $\{0,1\}^{m-r}$ (we consider only their first $m-r$ coordinates !) such that the hamming distance of every pair of them is at least n^2. Therefore according to corollary 1 we get that

$$m - r \geq 2n^2 - 4\frac{n^2}{k+2} .$$

Since $r \geq n^2 - n^{\frac{5}{3}}$ and $k = n^{\frac{1}{3}}$, we get that

$$m \geq 3n^2 - O(n^{\frac{5}{3}})$$

which is what we wanted to prove. This concludes the proof of the theorem. ♣

4.2. Proof of Theorem 1

Assume that $bl_*(MP_n) = m$. Let C be a smallest bilinear circuit for MP_n. Let

$$\mu_1(x) \cdot \eta_1(y), \ \ldots, \ \mu_m(x) \cdot \eta_m(y)$$

be the m bilinear forms computed in the product gates of C. We will show that these bilinear forms define in a very natural way a code on $M_n(GF(2))$. The code thus defined, will have the property that the dimension of the space into which the code maps $M_n(GF(2))$ is exactly m. Thus, according to theorem 3 we will get that $m \geq 3n^2 - O(n^{\frac{5}{3}})$, which is what we wanted to prove.

So we begin by defining a mapping from $M_n(GF(2))$ to $\{0,1\}^m$. Let $\Gamma : M_n(GF(2)) \mapsto \{0,1\}^m$ be the following mapping.

$$\Gamma(x) = (\mu_1(x), \ \ldots, \ \mu_m(x)) .$$

Notice that we ignore the η_i's in the definition of Γ. The next lemma shows that Γ is a linear code of matrices.

Lemma 7 *Γ is a linear transformation with the property that for every matrix $x \in M_n(GF(2))$, $\mathbf{weight}(\Gamma(x)) \geq n \cdot \mathrm{rank}(x)$.*

Proof: Clearly Γ is a linear transformation from $M_n(GF(2))$ to $\{0,1\}^m$. So we only have to prove the claim about the weights. Let x be a matrix of rank r. Assume w.l.o.g. that $\mu_1(x) = \ldots = \mu_k(x) = 1$ and that $\mu_{k+1}(x) = \ldots = \mu_m(x) = 0$, i.e. $\mathbf{weight}(\Gamma(x)) = k$. We shall show that $k \geq nr$. For every $y \in M_n(GF(2))$, the n^2 entries of $x \cdot y$ are functions of $\mu_1(x) \cdot \eta_1(y), \ \ldots, \ \mu_m(x) \cdot \eta_m(y)$. Since $\mu_{k+1}(x) = \ldots = \mu_m(x) = 0$, we get that $x \cdot y$ is a function of $\eta_1(y), \ \ldots, \ \eta_k(y)$. Therefore there at most 2^k different matrices of the form $x \cdot y$. Since $\mathrm{rank}(x) = r$ we get that there are exactly 2^{nr} different matrices of the form $x \cdot y$. Therefore $k \geq nr$. This concludes the proof of the lemma. ♣

Therefore Γ is a linear code of matrices, so according to theorem 3 we get that $m \geq 3n^2 - O(n^{\frac{5}{3}})$ which is what we wanted to prove. This concludes the proof of theorem 1. ♣

4.3. Proof of Theorem 2

As in the proof of theorem 1 we will show that every quadratic circuit for MP_n, defines a code on $M_n(GF(2))$. The code thus defined, will have the property that m (i.e the dimension of the space into which the code maps $M_n(GF(2))$) is exactly the number of product gates in the circuit. Thus, according to theorem 3 we will get that $m \geq 3n^2 - O(n^{\frac{5}{3}})$, which is what we wanted to prove.

Let C be a quadratic circuit for MP_n. Assume that the product gates of C compute the quadratic forms $\mu_1(x,y) \cdot \eta_1(x,y), \ \ldots, \ \mu_m(x,y) \cdot \eta_m(x,y)$. Thus, each of the outputs $(x \cdot y)_{i,j}$ can be written as a sum of these quadratic forms:

$$(x \cdot y)_{i,j} = \sum_{k=1}^{m} \alpha_{i,j}^{(k)} \cdot \mu_k(x,y) \cdot \eta_k(x,y) ,$$

where $\alpha_{i,j}^{(k)} \in \{0,1\}$.

We would like to have a proof similar to the proof of theorem 1. In that proof we defined a code of matrices using the linear transformation $\mu_1, ..., \mu_m$. Unfortunately this method will fail here as μ_i is a linear function in both the variables of x and the variables of y and not just in the variables of x as in the proof of theorem 1. In order to overcome this obstacle we introduce a new set of variables $z = \{z_{i,j}\}_{i,j=1\ldots n}$. We think about z as an $n \times n$ matrix. Define the following m linear forms in z:

$$\gamma_k(z) = \sum_{i,j} \alpha_{i,j}^{(k)} z_{i,j} , \quad k = 1, \ldots, m .$$

We get that

$$\sum_{k=1}^{m} \mu_k(x,y) \cdot \eta_k(x,y) \cdot \gamma_k(z) =$$

$$\sum_{k=1}^{m} (\sum_{i,j} z_{i,j} \cdot \alpha_{i,j}^{(k)}) \cdot \mu_k(x,y) \cdot \eta_k(x,y) =$$

$$\sum_{i,j} z_{i,j} \sum_{k=1}^{m} \alpha_{i,j}^{(k)} \cdot \mu_k(x,y) \cdot \eta_k(x,y) = \tag{1}$$

$$\sum_{i,j} z_{i,j} \cdot (x \cdot y)_{i,j} = \mathbf{trace}(x \cdot y \cdot z^t) \ ,$$

where $(z^t)_{i,j} = z_{j,i}$. The computation that we just performed shows that the γ_k's that we introduced are quite natural. We also notice that z plays the same role in $\mathbf{trace}(x \cdot y \cdot z^t)$ as x and y. These observations motivate us to try to repeat the proof of theorem 1 using the γ_k's instead of the μ_i's.

So define a linear mapping $\Gamma : M_n(GF(2)) \mapsto \{0,1\}^m$ by

$$\Gamma(z) = (\gamma_1(z), \ \ldots, \ \gamma_m(z)) \ .$$

The following lemma shows that Γ is indeed a linear code of matrices.

Lemma 8 Γ *is a linear mapping and it has the property that for every matrix z,* $\mathbf{weight}(\Gamma(z)) \geq n \cdot \mathrm{rank}(z)$.

Proof: Clearly Γ is a linear mapping. So we only have to prove the claim about the weights. Let z_0 be a matrix of rank r, and assume w.l.o.g. that $\gamma_1(z_0) = \ldots = \gamma_k(z_0) = 1$ and $\gamma_{k+1}(z_0) = \ldots = \gamma_m(z_0) = 0$. We wish to prove that $k \geq nr$. From equation 1 we get that

$$\mathbf{trace}(x \cdot y \cdot {z_0}^t) = \sum_{i=1}^{k} \mu_i(x,y) \cdot \eta_i(x,y) \ .$$

We now consider the discrete derivatives of this equation. Let $e_{i,j}$ be the matrix of all zeros but 1 in the (i,j)'th place. Define

$$\frac{\partial}{\partial x_{i,j}} \mathbf{trace}(x \cdot y \cdot {z_0}^t) \stackrel{\mathrm{def}}{=}$$

$$\mathbf{trace}((x + e_{i,j}) \cdot y \cdot {z_0}^t) - \mathbf{trace}(x \cdot y \cdot {z_0}^t) \ .$$

On the one hand

$$\mathbf{trace}((x + e_{i,j}) \cdot y \cdot {z_0}^t) - \mathbf{trace}(x \cdot y \cdot {z_0}^t) =$$

$$\mathbf{trace}(e_{i,j} \cdot y \cdot {z_0}^t) = (z_0 \cdot y^t)_{i,j} \ .$$

On the other hand we have that

$$\mathbf{trace}((x + e_{i,j}) \cdot y \cdot {z_0}^t) - \mathbf{trace}(x \cdot y \cdot {z_0}^t) =$$

$$\sum_{i=1}^{k} (\mu_i(x + e_{i,j}, y) \cdot \eta_i(x + e_{i,j}, y) - \mu_i(x,y) \cdot \eta_i(x,y)) =$$

$$\sum_{i=1}^{k} (\mu_i(e_{i,j}, 0) \cdot \eta_i(x,y) + \mu_i(x,y) \cdot \eta_i(e_{i,j}, 0)) + \tag{2}$$

$$+ \sum_{i=1}^{k} \mu_i(e_{i,j}, 0) \cdot \eta_i(e_{i,j}, 0) \ ,$$

where the last equality follows from the linearity of the μ_i's and the η_i's. Since $(z_0 \cdot y^t)_{i,j}$ is a linear form in y, we actually get that

$$(z_0 \cdot y^t)_{i,j} = \frac{\partial}{\partial x_{i,j}} \mathbf{trace}(x \cdot y \cdot {z_0}^t) =$$

$$\sum_{i=1}^{k} (\mu_i(e_{i,j}, 0) \cdot \eta_i(x,y) + \mu_i(x,y) \cdot \eta_i(e_{i,j}, 0))$$

$$\subset \mathrm{span}(\mu_i(x,y), \ \eta_i(x,y))$$

(since $(z_0 \cdot y^t)_{i,j}$ is a linear form the third summand of equation 2 sums to 0). In the same manner we define

$$\frac{\partial}{\partial y_{i,j}} \mathbf{trace}(x \cdot y \cdot {z_0}^t) \stackrel{\mathrm{def}}{=}$$

$$\mathbf{trace}(x \cdot (y + e_{i,j}) \cdot {z_0}^t) - \mathbf{trace}(x \cdot y \cdot {z_0}^t) \ .$$

We get that

$$(x^t \cdot z_0)_{i,j} = \frac{\partial}{\partial y_{i,j}} \mathbf{trace}(x \cdot y \cdot {z_0}^t) =$$

$$\sum_{i=1}^{k} (\mu_i(0, e_{i,j}) \cdot \eta_i(x,y) + \mu_i(x,y) \cdot \eta_i(0, e_{i,j}))$$

$$\subset \mathrm{span}(\mu_i(x,y), \ \eta_i(x,y)) \ .$$

Denote by PD the set of all the discrete partial derivatives

$$\left\{ \frac{\partial}{\partial x_{i,j}} \mathbf{trace}(x \cdot y \cdot {z_0}^t), \frac{\partial}{\partial y_{i,j}} \mathbf{trace}(x \cdot y \cdot {z_0}^t) \right\}_{i,j} \ .$$

We just proved that PD is contained in the linear span of

$$\{ \ \mu_i(x,y), \ \eta_i(x,y) \ \}_{i=1}^{k} \ ,$$

in the vector space of all linear forms in $x, \ y$. Therefore

$$\dim(\mathrm{span}(\mathrm{PD})) \leq$$

$$\leq \dim(\mathrm{span}\{ \ \mu_i(x,y), \ \eta_i(x,y) \ \}_{i=1}^{k}) \leq 2k \ . \tag{3}$$

We also showed that

$$\left\{ \frac{\partial}{\partial x_{i,j}} \mathbf{trace}(x \cdot y \cdot {z_0}^t), \frac{\partial}{\partial y_{i,j}} \mathbf{trace}(x \cdot y \cdot {z_0}^t) \right\}_{i,j} =$$

$$= \left\{ \ (x^t \cdot z_0)_{i,j}, \ (z_0 \cdot y^t)_{i,j} \ \right\}_{i,j} \ .$$

Therefor, using our assumption that $\mathrm{rank}(z_0) = r$, we get that

$$\dim(\mathrm{span}(\mathrm{PD})) =$$

$$= \dim(\mathrm{span} \left\{ \ (x^t \cdot z_0)_{i,j}, \ (z_0 \cdot y^t)_{i,j} \right\}_{i,j}) = 2nr \ . \tag{4}$$

Combining equations 3 and 4 we get that $2k \geq 2nr$. ♣

Theorem 2 now follows from applying theorem 3 on the linear code of matrices Γ. ♣

5. Other Finite Fields

In this section we prove the following theorem.

Theorem 4 *The number of product gates in any bilinear circuit that computes the product of two $n \times n$ matrices over $GF(p)$ is at least $(2.5 + \frac{1.5}{p^3-1})n^2 - O(n^{\frac{7}{4}})$ (i.e. $bl_*(MP_n) \geq (2.5 + \frac{1.5}{p^3-1})n^2 - O(n^{\frac{7}{4}})$ over $GF(p)$).*

Let C be a bilinear circuit for MP_n over $GF(p)$. Assume that $\mu_1(x) \cdot \eta_1(y), \ \ldots, \ \mu_m(x) \cdot \eta_m(y)$ are the bilinear forms computed in the product gates of C. The following lemma of Bläser is the main tool in the proof of the theorem.

Lemma 9 *[1] Let $[a, b] = ab - ba$. If there are two matrices $a, \ b$ such that $[a, b]$ is an invertible matrix, and such that there are t linear forms among $\mu_1, \ \ldots, \ \mu_m$ such that each of them vanish on $I, \ a, \ b$ then*

$$m \geq t + 1.5n^2 \ .$$

We are going to prove that we can find a, b such that $(1 - \frac{1}{p^3})n^2 + \frac{m}{p^3} - O(n^{\frac{7}{4}})$ linear forms among $\mu_1, \ \ldots, \ \mu_m$ vanish on $I, \ a, \ b$, and such that $[a, b]$ is invertible.

Proof of Theorem 4: We begin by proving that (w.l.o.g.) many of the μ_i's vanish on I. The following lemma shows that we can always find an invertible matrix such that many of the μ_i's vanish on it. As before we assume that $\mu_1, \ldots, \mu_{n^2}$ are independent linear forms.

Lemma 10 *There exists an invertible matrix c, such that at least $(1 - \frac{1}{p})n^2 + \frac{m}{p} - O(n^{\frac{5}{3}})$ of the μ_i's vanish on it, where $n^2 - O(n^{\frac{5}{3}})$ of the μ_i's that vanish on it are among $\mu_1, \ \ldots, \ \mu_{n^2}$.*

Proof: An analog of lemma 6 over $GF(p)$ guarantees that we can find $k = n^{\frac{1}{3}}$ matrices, $a_1, \ \ldots, \ a_k \in M_n(GF(p))$, such that for $i \neq j$ $a_i - a_j$ is invertible, and such that $n^2 - \binom{k}{2}n$ linear forms among $\mu_1, \ \ldots, \ \mu_{n^2}$ vanish on all of them. Denote $r = n^2 - \binom{k}{2}n$. And assume w.l.o.g. that $\mu_1, \ \ldots, \ \mu_r$ vanish on all the a_i's.

Let us consider the following k vectors in $GF(p)^{m-r}$:

$$\Gamma(a_i) \stackrel{\rm def}{=} (\mu_{r+1}(a_i), \ldots, \mu_m(a_i)) \ , \ i = 1 \ldots k \ .$$

As in the proof of theorem 1, we get that since $\forall i \neq j$ $a_i - a_j$ is an invertible matrix, then $\mathbf{d_H}(\Gamma(a_i), \Gamma(a_j)) \geq n^2$. According to lemma 5, two of these vectors agree on at least $\frac{m-r}{p} - \frac{m-r}{k}$ coordinates. Assume that $\Gamma(a_1)$ and $\Gamma(a_2)$ are these vectors. Denote $c = a_1 - a_2$. We have that c is an invertible matrix, such that the first $r = n^2 - \binom{k}{2}n$ linear forms (which are independent) vanish on it, and such that all the linear forms that $\Gamma(a_1)$ and $\Gamma(a_2)$ agree on, vanish on it as well. Therefore there are at least

$$r + \frac{m-r}{p} - \frac{m-r}{k}$$

linear forms that vanish on c. Since $r = n^2 - \binom{k}{2}n$, and $k = n^{\frac{1}{3}}$, we get that at least

$$(1 - \frac{1}{p})n^2 + \frac{m}{p} - O(n^{\frac{5}{3}})$$

linear forms vanish on c, $n^2 - O(n^{\frac{5}{3}})$ of them are among $\mu_1, \ \ldots, \ \mu_{n^2}$ (we assume for simplicity that $n^2 < m < 10n^2$, as it will not change the results). This completes the proof of the lemma. ♣

The lemma doesn't tell us who c is, but using the *sandwiching method* we can assume that $c = I$: We know that $x \cdot y$ is computed using the bilinear forms

$$\mu_1(x) \cdot \eta_1(y), \ \ldots, \ \mu_m(x) \cdot \eta_m(y) \ .$$

We now do the following trick: $x \cdot y = (x \cdot c) \cdot (c^{-1} \cdot y)$, therefore $x \cdot y$ can be computed using the bilinear forms

$$\tilde{\mu_1}(x) \cdot \tilde{\eta_1}(y), \ \ldots, \ \tilde{\mu_m}(x) \cdot \tilde{\eta_m}(y) \ ,$$

where

$$\tilde{\mu_i}(x) \stackrel{\rm def}{=} \mu_i(x \cdot c) \ \text{ and } \ \tilde{\eta_i}(y) \stackrel{\rm def}{=} \eta_i(c^{-1} \cdot y) \ .$$

Thus, if $\mu_i(c) = 0$ then we get that $\tilde{\mu_i}(I) = \mu_i(I \cdot c) = 0$. This trick is called sandwiching, for further background see [1, 6].

So by combining the sandwiching method and lemma 10 we get that we can assume w.l.o.g. that $(1 - \frac{1}{p})n^2 + \frac{m}{p} - O(n^{\frac{5}{3}})$ of the μ_i's, where $n^2 - O(n^{\frac{5}{3}})$ of them are among $\mu_1, \ \ldots, \ \mu_{n^2}$, vanish on I. The next lemma now assures us that we can find two matrices a, b that satisfy the requirements of lemma 9.

Lemma 11 *There are two matrices $a, \ b$ such that $[a, b]$ is an invertible matrix and such that at least $(1 - \frac{1}{p^3})n^2 + \frac{m}{p^3} - O(n^{\frac{7}{4}})$ of the μ_i's vanish on $I, \ a, \ b$.*

Proof: The proof of this lemma is similar to the proof of lemma 10. Let $k = n^{\frac{1}{4}}$. The following lemma shows that we can find k matrices such that many of the μ_i's vanish on all of them and such that among their differences there are matrices satisfying the requirements of lemma 9.

Lemma 12 *For every n, k, such that $p^{\frac{n}{2}} > 4\binom{k}{3}$, and any $\mu_1, \ \ldots, \ \mu_{n^2}$ linearly independent linear forms, in n^2 variables, over $GF(p)$, there are k matrices, $a_1, \ \ldots, \ a_k$, such that $\forall i < j < l$, $[a_i - a_l, a_j - a_l]$ is invertible, and such that $n^2 - 2\binom{k}{3}n$ of the μ_i's vanish on all the a_i's.*

Proof: Again we use lemma 3. Let P be the following polynomial.

$$P(a_1,...,a_k) = \text{determinant}\left(\prod_{i<j<l}[a_i - a_l, a_j - a_l]\right).$$

Clearly $deg(P) = 2\binom{k}{3}n$ (as a polynomial in the entries of the a_i's). Therefore if we will prove that $P \not\equiv 0$, i.e. that there exist k matrices on which P is not zero, then according to lemma 3 we are done. This is guaranteed by the following lemma.

Lemma 13 *If $p^{\frac{n}{2}} > 4\binom{k}{3}$ then there exist k matrices in $M_n(GF(p))$, $a_1, \ldots, a_k$ such that $\forall i < j < l$, $[a_i - a_l, a_j - a_l]$ is invertible.*

We prove the lemma only for n even. Clearly this will not affect theorem 4, as the lower bound for odd n follows from the lower bound for even n.

Proof: Consider the following polynomial in **4k** variables (i.e. it is a polynomial in k matrices over $\mathbf{M_2(GF(p))}$!).

$$Q(x_1,...,x_k) = \text{determinant}\left(\prod_{i<j<l}[a_i - a_l, a_j - a_l]\right).$$

Q is a polynomial of degree $d = 4\binom{k}{3}$ over $GF(p)$, in the entries of the a_i's. Clearly Q is not the zero polynomial (as it is a product of non zero polynomials). Consider the field $F = GF(p^{\frac{n}{2}})$. Since $d < |F|$ we get by lemma 1 that there are k matrices $\rho_1,...,\rho_k \in M_2(F)$ such that $Q(\rho_1,...,\rho_k) \neq 0$. That is, $\forall i < j < l$ $[\rho_i - \rho_l, \rho_j - \rho_l]$ is an invertible matrix. According to lemma 4 we can embed $M_2(F)$ in $M_n(GF(p))$. Therefore there are k matrices in $M_n(GF(p))$ satisfying $\forall i < j < l$ $[a_i - a_l, a_j - a_l]$ is an invertible matrix, which is what we wanted to prove. ♣

This concludes the proof of lemma 12. ♣

We proceed with the proof of lemma 11. We now restrict our attention to the linear forms among $\mu_{n^2+1}, \ldots, \mu_m$ that vanish on I. We shall prove that three of the matrices guaranteed by lemma 12 agree on many of these linear forms (more formally on $\frac{m-n^2}{p^3} - O(n^{\frac{7}{4}})$ of them). Thus, if a_1, a_2, a_3 are these three matrices, then we get that $(1 - \frac{1}{p^3})n^2 + \frac{m}{p^3} - O(n^{\frac{7}{4}})$ linear forms vanish on I, $(a_1 - a_3)$, $(a_2 - a_3)$, and that $[(a_1 - a_3), (a_2 - a_3)]$ is an invertible matrix, which is what we wanted to prove.

So assume w.l.o.g. that the linear forms $\mu_{n^2+1}, \ldots, \mu_{n^2+r}$, vanish on I (beside those among $\mu_1, \ldots, \mu_{n^2}$ that vanish on it) where $r \geq \frac{m-n^2}{p} - O(n^{\frac{5}{3}})$. Let $a_1, \ldots, a_k$ be the matrices guaranteed by lemma 12. Consider the following vectors: $\forall 1 \leq i \leq k$,

$$v_i = (\mu_{n^2+1}(a_i), \ldots, \mu_{n^2+r}(a_i)) \in GF(p)^r .$$

According to lemma 5 three of these vectors, namely v_1, v_2, v_3, agree on at least $\frac{r}{p^2} - \frac{3r}{pk}$ coordinates. Therefore there are $\frac{r}{p^2} - \frac{3r}{pk}$ linear forms among $\mu_{n^2+1}, \ldots, \mu_{n^2+r}$ that vanish on $a_1 - a_3$ and $a_2 - a_3$. In addition there are $n^2 - 2\binom{k}{3}n$ linear forms among $\mu_1, \ldots, \mu_{n^2}$ that vanish on a_1, a_2, a_3, hence there are $n^2 - 2\binom{k}{3}n$ linear forms among $\mu_1, \ldots, \mu_{n^2}$ that vanish on $a_1 - a_3$ and on $a_2 - a_3$. Let $a = a_1 - a_3$, $b = a_2 - a_3$.

We get that there are $\frac{r}{p^2} - \frac{3r}{pk}$ linear forms among $\mu_{n^2+1}, \ldots, \mu_{n^2+r}$ that vanish on I, a, b. Since $n^2 - O(n^{\frac{5}{3}})$ of the first n^2 μ_i's vanish on I, we get that at least $n^2 - 2\binom{k}{3}n - O(n^{\frac{5}{3}})$ of the first n^2 μ_i's vanish on I, a, b. Putting it all together we get that at least

$$n^2 - 2\binom{k}{3}n - O(n^{\frac{5}{3}}) + \frac{r}{p^2} - \frac{3r}{pk}$$

linear forms among $\mu_1, \ldots, \mu_m$ vanish on I, a, b. Since $r = \frac{m-n^2}{p} - O(n^{\frac{5}{3}})$, and $k = n^{\frac{1}{4}}$, we get that at least

$$(1 - \frac{1}{p^3})n^2 + \frac{m}{p^3} - O(n^{\frac{7}{4}})$$

of the μ_i's vanish on them. This concludes the proof of lemma 11. ♣

Putting everything together we get by lemma 9 and lemma 11 that:

$$m \geq 1.5n^2 + (1 - \frac{1}{p^3})n^2 + \frac{m}{p^3} - O(n^{\frac{7}{4}}) .$$

Therefore

$$m \geq (2.5 + \frac{1.5}{p^3 - 1})n^2 - O(n^{\frac{7}{4}}) .$$

This concludes the proof of theorem 4. ♣

6. Acknowledgment

I would like to thank Michael Ben-Or and Avi Wigderson for helpful conversations. I would also like to thank Ran Raz for commenting on an earlier draft of the paper (in his words: Would you like to hear my comments or do you just want to rewrite everything).

References

[1] M. Bläser. A $5/2n^2$-lower bound for the rank of $n \times n$-matrix multiplication over arbitrary fields. In IEEE, editor, *40th Annual Symposium on Foundations of Computer Science: October 17–19, 1999, New York City, New York,*, pages 45–50, 1109 Spring Street, Suite 300, Silver Spring, MD 20910, USA, 1999. IEEE Computer Society Press.

[2] M. Bläser. Lower bound for the multiplicative complexity of matrix multiplication. *Journal of computational complexity*, 9:73–112, 2000.

[3] N. H. Bshouty. A lower bound for matrix multiplication. *SIAM Journal on Computing*, 18(4):759–765, Aug. 1989.

[4] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. *Algebraic Complexity Theory*. Springer, 1997.

[5] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progression. *Journal of Symbolic Computation*, 9:251–280, 1990.

[6] H. F. de Groote. *Lectures on the complexity of bilinear problems*. Number 245 in LNCS. Springer, 1986.

[7] J. v. z. Gathen. Algebraic complexity theory. *Ann. Rev. Computer Science*, pages 317–347, 1988.

[8] J.-C. Lafon and S. Winograd. A lower bound for the multiplicative complexity of the product of two matrices. Technical report, Centre de Calcul de L'Esplanade, U.E.R. de Mathematique Univ. Louis Pasteur, Strasbourg, France, 1978.

[9] R. Raz and A. Shpilka. Lower bounds for matrix product, in bounded depth circuits with arbitrary gates. In *Thirty-Third Annual ACM Symposium on Theory of Computing, Crete, Greece, July 6-8, 2001*, pages 409–418. ACM, 2001.

[10] D. D. Roger W. Brocket and. On the optimal evaluation of a set of bilinear forms. *Linear Algebra and its Applications*, 9:207–235, 1978.

[11] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM*, 27(4):701–717, 1980.

[12] V. Strassen. Gaussian elimination is not optimal. *Numer. Math*, 13:354–356, 1969.

[13] V. Strassen. Vermeidung von divisionen. *Journal of Reine Angew. Math.*, 264:182–202, 1973.

[14] J. H. van Lint. *Introduction to coding theory*. Springer-Verlag, Berlin, second edition, 1992.

[15] R. Zippel. Probabilistic algorithms for sparse polynomials. In *Symbolic and algebraic computation (EUROSAM '79, Internat. Sympos., Marseille, 1979)*, pages 216–226. Springer, Berlin, 1979.