Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 29 (2000)

Lower Bounds for Matrix Product, in Bounded Depth Circuits with Arbitrary Gates

Ran Raz [*] Amir Shpilka [†]

Abstract

We prove super-linear lower bounds for the number of edges in constant depth circuits with n inputs and up to n outputs. Our lower bounds are proved for all types of constant depth circuits, e.g., constant depth arithmetic circuits and constant depth Boolean circuits with arbitrary gates. The bounds apply for several explicit functions, and, most importantly, for matrix product. In particular, we obtain the following results:

1. We show that the number of edges in any constant depth arithmetic circuit for **matrix product** (over any field) is super-linear in m^2 (where $m \times m$ is the size of each matrix). That is, the lower bound is super-linear in the number of input variables. Moreover, if the circuit is bilinear the result applies also for the case where the circuit gets for free any product of two linear functions.
2. We show that the number of edges in any constant depth arithmetic circuit for the trace of the product of 3 matrices (over fields with characteristic 0) is super-linear in m^2. (Note that the trace is a **single-output** function).
3. We give explicit examples for n Boolean functions $f_1, ..., f_n$, such that any constant depth **Boolean circuit with arbitrary gates** for $f_1, ..., f_n$ has a super-linear number of edges. The lower bound is proved also for **circuits with arbitrary gates over any finite field**. The bound applies for matrix product over finite fields as well as for several other explicit functions.

1 Introduction

Exponential lower bounds are well known for constant depth Boolean circuits over the base $\{\mathrm{AND}, \mathrm{OR}, \mathrm{NOT}\}$ [Ajt83, FSS81, Yao85, Hås86]. However, for many other types of constant depth circuits almost nothing is known. In this work, we prove super-linear lower bounds for the number of edges in constant depth circuits with n inputs and up to n outputs. Our lower bounds are proved for all models of Boolean and arithmetic circuits and, in particular, for

[*] ranraz@wisdom.weizmann.ac.il, Department of Computer Science, Weizmann Institute, Rehovot 76100, ISRAEL

[†] amirs@cs.huji.ac.il, Department of Computer Science, Hebrew University, Givat-Ram, Jerusalem, ISRAEL

ISSN 1433-8092

Boolean circuits with arbitrary gates. The bounds apply for several explicit functions and, in particular, for matrix product.

In general, our lower bound for circuits of depth $d \geq 2$ is $\Omega(n \cdot \lambda_d(n))$, where $\lambda_d(n)$ is a slowly growing function. A description of the functions $\lambda_d(n)$ is given below in Subsection 1.5. Our main method is a graph theoretic argument that analyzes certain superconcentration properties of the circuit as a graph. Hence, the same lower bounds are obtained for all types of circuits. Our results and proof methods are related to the works of [DDPW83, Pud94], where lower bounds of $\Omega(n \cdot \lambda_d(n))$ were proved for the size of superconcentrators. Pudlak used similar methods to prove lower bounds of $\Omega(n \cdot \lambda_d(n))$ for the number of edges in constant depth arithmetic circuits with n inputs and up to n outputs over fields with characteristic 0 [Pud94]. Pudlak's results hold for the parallel prefix problem as well as for other explicit functions.

In all that comes below, the size of a circuit means the number of edges in it.

1.1 Matrix Product

Matrix product is among the most studied computational problems. Surprising upper bounds of $O(m^{2+\epsilon})$ (where $\epsilon < 1$, and $m \times m$ is the size of each matrix) were obtained by Strassen in [Str69], and improved in many other works (see [Gat88] for a survey). The only known lower bound, however, is a lower bound of $2.5 \cdot m^2$ for the number of products needed [Bsh89, Bla99]. In particular, the following problem is still open: Can matrix product be computed by circuits of size $O(m^2)$? Non-trivial size-depth tradeoffs for matrix product are also not known. In particular, the following problem is still open: Can matrix product be computed by circuits of size $O(m^2)$ and logarithmic depth ? In this work we prove that matrix product cannot be computed by circuits of size $O(m^2)$ and constant depth.

The standard computational model for matrix product is by arithmetic circuits over some field F. Usually, it is assumed that the circuits are bilinear, that is, product gates are applied only on two linear functions, where the first function is linear in the variables of the first matrix and the second function is linear in the variables of the second matrix. Such an assumption can be made w.l.o.g. if the field F is of characteristic 0. For fields of characteristic different than 0, the non-bilinear case is also interesting. Note, however, that all known upper bounds for matrix product (over any field) are by bilinear circuits.

In the bilinear case, our lower bound proof works also if the circuit gets for free any product of two linear functions. That is, the lower bound is proven for the number of edges above the product gates. We prove that if the circuit is of depth 1 above these products (i.e., total depth 3) it is of size $\Omega(m^3)$. For $d \geq 2$, we prove that if the circuit is of depth d above the products (i.e., total depth $d+2$) it is of size $\Omega(m^2 \cdot \lambda_d(m))$. In the general (non-bilinear) case for fields of characteristic different than 0, our lower bound is $\Omega(m^3)$ for circuits of depth 1 and $\Omega(m^2 \cdot \lambda_d(m))$ for circuits of depth $d \geq 2$. The last lower bound is a special case of a more general lower bound for circuits with arbitrary gates over finite fields. That lower bound is discussed in Subsection 1.2.

1.2 Circuits with Arbitrary Gates

In a *Boolean circuit with arbitrary gates*, we allow each gate (of fanin k) to compute an arbitrary function $g : \{0,1\}^k \to \{0,1\}$. In this work, we give explicit examples for (up to) n Boolean functions $f_1, ..., f_n$, such that any constant depth Boolean circuit with arbitrary gates for $f_1, ..., f_n$ is of super-linear size (as before, the bound for depth $d \geq 2$ is $\Omega(n \cdot \lambda_d(n))$). The bound holds for matrix product over $GF[2]$ (where the dimension of each matrix is $m = \sqrt{n/2}$), as well as for matrix product over other finite fields (where, say, each field element is represented by its bits). The bound also holds for the parallel prefix problem and for other problems from [Pud94].

One important special case of this model is *constant depth threshold circuits.* Note, however, that for constant depth threshold circuits super-linear lower bounds are already known [IPS97].

The above results for Boolean circuits with arbitrary gates can be generalized to circuits over larger domains. Let F be some fixed finite set (e.g., some fixed finite field), and assume that the input variables range over F (or over a subset of F). A *circuit with arbitrary gates over F* allows each gate (of fanin k) to compute an arbitrary function $g : F^k \to F$. By a reduction to the Boolean case, we get explicit examples for (up to) n functions $f_1, ..., f_n$, such that any constant depth circuit with arbitrary gates over F for $f_1, ..., f_n$ is of super-linear size (as before, the bound for depth $d \geq 2$ is $\Omega(n \cdot \lambda_d(n))$). In particular, this gives lower bounds for circuits with arbitrary gates over any finite field F. The bound holds for matrix product over F as well as for many other functions.

1.3 Arithmetic Circuits

In *arithmetic circuits* the allowed gates are product and addition over a field F. Constants in the field are also allowed. Arithmetic circuits compute polynomials in the ring $F[x_1, ..., x_n]$ (where $x_1, ..., x_n$ are the input variables for the circuit), and we would like to give explicit examples for polynomials that are hard to compute. Note that for finite fields the representation of a function $f : F^n \to F$ as a polynomial is not unique (since for every i we have the equation $x_i^p = x_i$, where $p \neq 0$ is the characteristic of the field). Usually, it is only required that the circuit compute the given polynomials as functions, that is, the circuit may compute other polynomials that represent the same functions.

Lower bounds for the size of arithmetic circuits for explicit polynomials are known only if we allow polynomials with large degree or large coefficients (e.g., [Str73, BS82]). However, if we limit the degree and the coefficients to be of size $O(1)$ then no non-trivial lower bound is known. For constant depth arithmetic circuits, exponential lower bounds are known for fields F with characteristic $p = 2$ [Razb87, Smo87]. For other finite characteristics, exponential lower bounds are known only for depth 3 [GK98, GR98] (and for depth larger than 3 no non-trivial lower bound was known). For characteristic 0, the best lower bounds for depth 3 are the almost quadratic bounds of $\Omega(n^{2-\epsilon})$ [SW99].

In this work, we get (for any field F) explicit examples for (up to) n polynomials $f_1, ..., f_n$ such that any constant depth arithmetic circuit (over F) for $f_1, ..., f_n$ is of super-linear size. One such example is matrix product (over F). For finite fields (and hence also for any field with characteristic different than 0), this follows by the general lower bound for circuits with arbitrary gates over F, as discussed in Subsection 1.2. For fields with characteristic 0, this follows from the bilinear lower bound for matrix product, as discussed in Subsection 1.1. Similar bounds for fields with characteristic 0 were previously proved by Pudlak [Pud94]. Pudlak gives explicit examples for n linear functions $f_1, ..., f_n$ such that any constant depth arithmetic circuit with linear gates (i.e., products are not allowed) for $f_1, ..., f_n$ is of super-linear size. (Over fields with characteristic 0, the assumption that all the gates in the circuit are linear can be made w.l.o.g.)

For fields with characteristic 0, our results (as well as Pudlak's results) also give explicit examples for **one** polynomial $h = f_1 \cdot y_1 + \cdots + f_n \cdot y_n$ (in the input variables $x_1, ..., x_n, y_1, ..., y_n$) such that any constant depth arithmetic circuit for h is of super-linear size. This follows easily by the result of [BS82] and was noted to us by Toni Pitassi and Avi Wigderson.

1.4 Methods and Related Work

Our main lemma gives an analysis of the structure of a constant depth circuit as a graph. Let G be a directed acyclic graph. Denote by V_G the set of all nodes of G. Denote by I_G the set of all nodes of indegree 0 (inputs), and by O_G the set of all nodes of outdegree 0 (outputs). The depth of G is the length of the longest directed path in G. Roughly speaking, the main lemma shows that if G is of depth d and has less than $n \cdot \lambda_d(n)$ edges then one can remove from G a set of $\epsilon \cdot n$ inputs and $\epsilon \cdot n$ outputs (for some small constant ϵ) and a small number of intermediate nodes, such that in the new graph the total number of directed paths from I_G to O_G is small.

Lemma 1.1 *For any $0 < \epsilon < 1/400$ and any directed acyclic graph G of depth d, with more than n vertices and less than $\epsilon \cdot n \cdot \lambda_d(n)$ edges, the following is satisfied:*

For some k, s.t., $\sqrt{n} \leq k = o(n)$, there exist subsets $I \subset I_G$, $O \subset O_G$ and $V \subset V_G$, s.t., $|I|, |O| \leq 5\epsilon \cdot d \cdot n$, and $|V| = k$, and such that the total number of directed paths from $I_G \setminus I$ to $O_G \setminus O$, that do not pass through nodes in V, is at most $\epsilon \cdot n^2/k$.

Lemma 1.1 is restated (in a slightly more general form) as Corollary 3.12.

The main lemma is used to transform any circuit of depth d and size less than $\epsilon n \lambda_d(n)$ into a new circuit of depth 1 (and relatively small size). This is done by removing from the original circuit $5\epsilon dn$ inputs, $5\epsilon dn$ outputs, and a small number of intermediate nodes. The lower bounds then follow by a *rigidity argument*, in the spirit of Valiant's approach [Val77].

As mentioned before, similar methods were previously used to prove lower bounds for superconcentrators [DDPW83, Pud94] and for constant depth arithmetic circuits over fields with characteristic 0 [Pud94]. In particular, methods similar to our main lemma are implicit in [Pud94] (although the presentation there is different). Versions of these methods appeared already in [DDPW83]. One can think of Lemma 1.1 also as a generalization of the lower bounds for superconcentrators given in [DDPW83, Pud94]. In fact, all these lower bounds follow easily by a reduction to Lemma 1.1. Our proof for Lemma 1.1 heavily relies on [Pud94].

1.5 The Functions $\lambda_d(n)$

We will now give a description of the functions $\lambda_d(n)$. The exact definition is given below in Section 2. Let $\theta(n)$ be any function on the natural numbers, such that for any $k > 0$, $\theta(k) < k$. We denote by $\theta^*(n)$ the following function: $\theta^*(n) = \min\{i \mid \theta^{(i)}(n) \leq 1\}$, that is, the smallest integer i such that i-times iteration of θ gives a value smaller than 1. The functions $\lambda_d(n)$ will satisfy:

1. $\lambda_2(n) = \Theta(\log n)$,
2. $\lambda_3(n) = \Theta(\log\log n)$,
3. for any even $d > 2$, $\lambda_{d+1}(n), \lambda_d(n) = \Theta(\lambda^*_{d-2}(n))$
   (e.g., $\lambda_4(n), \lambda_5(n) = \Theta(\log^*(n)))$.

1.6 Organization of the Paper

In Section 2 we give the definition of the functions $\lambda_d(n)$ and prove some simple properties of these functions. In Section 3 we give the proof of Lemma 1.1. In Section 4 we prove our results for matrix product by bilinear arithmetic circuits. In Section 5 we prove our results for Boolean circuits with arbitrary gates and for circuits with arbitrary gates over finite fields.

2 Slowly Growing Functions

In this section, we define the functions $\lambda_d(n)$ and we prove some easy properties of them. We start with a definition of the "star" operator.

Definition 2.1 *For a function f, define $f^{(i)}$ to be the composition of f with itself i times, i.e., $f^{(i)} = f \circ f \circ ... \circ f$ i times, $f^{(1)} = f$.*
For a function $f : N \to N$ such that $f(n) < n$ for $n > 0$, define:

$$f^*(n) = \min\{\ i \mid \ s.t., \ \ f^{(i)}(n) \leq 1\} \ .$$

We will need the following properties of f^* (taken from [Pud94]).

Claim 2.2 *[Pud94] Suppose $f(n) \leq \lfloor\sqrt{n}\rfloor$. For every $n \geq 0$ we have:*

1. *$\frac{f^{(i)}(n)}{f^{(i+1)}(n)} \geq f^{(i+1)}(n)$, for every $i > 0$, (provided that the denominator is not 0).*
2. *$f^{(i)}(n) \geq \frac{f^*(n)}{2}$, for every $i \leq \frac{f^*(n)}{2}$.*

The proof is taken from [Pud94] as well.

Proof:

1. $$\frac{f^{(i)}(n)}{f^{(i+1)}(n)} \geq \frac{f^{(i)}(n)}{\lfloor\sqrt{f^{(i)}(n)}\rfloor} \geq \sqrt{f^{(i)}(n)} \geq f^{(i+1)} \quad .$$

2. From (1) it follows that if $f^{(i+1)}(n) > 1$ then $f^{(i)}(n) > f^{(i+1)}(n)$. Therefore
$$f(n) > f^{(2)}(n) > ... > f^{(f^*(n))}(n) \quad .$$
Since the values of f are integers, the result follows.

□

Our lower bounds will be expressed in terms of the following set of slowly growing functions.

Definition 2.3 *Let*
$$\lambda_1(n) = \lfloor\sqrt{n}\rfloor \quad ,$$
$$\lambda_2(n) = \lceil\log n\rceil \quad ,$$
$$\lambda_d(n) = \lambda^*_{d-2}(n) \quad .$$

Some easy to verify properties of these functions are:

Claim 2.4 *1. Each $\lambda_i(n)$ is a monotone increasing function tending to infinity with n.*

2. for $i \geq 2$, $\lambda_{2i}(n) = \theta(\lambda_{2i+1}(n))$.

3. for $i \geq 2$ and n large enough, $\lambda_i(n) \leq \lfloor\sqrt{\frac{n}{2}}\rfloor$.

Proof:

1. The fact that λ_i is increasing is immediate from the fact that λ_1 and λ_2 are.

2. Notice that $\lambda_3(n) = \theta(\log\log n)$. Since $\log\log n = \log^{(2)}(n)$, we have $\lambda_4(n) = \theta(\lambda_5(n))$. Using induction we get the desired result.

3. Clearly $\lambda_2(n), \lambda_3(n) \leq \sqrt{\frac{n}{2}}$, for n large enough. Assume that $\lambda_j(n) \leq \sqrt{\frac{n}{2}}$. We have
$$\sqrt{\frac{n}{2}} \geq \lambda_j(n) \geq (\lambda_j^{(2)}(n))^2 \geq \frac{1}{4}(\lambda_{j+2}(n))^2 \ ,$$
hence
$$\lambda_{j+2}(n) \leq (8n)^{\frac{1}{4}} \leq \sqrt{\frac{n}{2}}$$
for n large enough.

□

3 Superconcentration Properties of Graphs

In this section we prove our main lemma on graphs, and several stronger versions of it. The lemma will be used to analyze the structure of a constant depth circuit as a graph. For simplicity, we prove here the lemma for leveled graphs. The general case follows easily by a reduction to the leveled case. Let $G = (V_G, E_G)$ be a leveled graph of depth d. The number of levels in G is hence $d+1$ and all edges in the graph are between vertices of adjacent levels. In all the following we allow all graphs to be multi-graphs.

We will use the following notations: We denote by $L_0, ..., L_d$ the levels of G, that is, L_i is the set of vertices at level i. The set of vertices L_0 is also denoted by I_G (and we call these vertices *inputs*). The set of vertices L_d is also denoted by O_G (and we call these vertices *outputs*).

Let $U \subset V_G$ be a set of vertices. We denote by $E(U)$ the set of edges that touch vertices in U. We denote by $\Gamma(U)$ the set of neighbors of U. We denote by $maxdeg(U)$ the maximal degree of a vertex in U. For subsets $I \subset I_G$, $O \subset O_G$, $V \subset V_G$ we denote by $P_G[I, O, V]$ the total number of paths of length d between I and O that do not pass through vertices in V.

3.1 Depth 2

We will first prove the main lemma for graphs of depth 2. We prove first a stronger lemma (Lemma 3.1). The main lemma (for depth 2) will then follow as Corollary 3.3. These lemmas are not needed for the proofs for higher depth. Nevertheless, the methods used here give some hints for the proofs needed for the general case.

Lemma 3.1 *Let G be a leveled graph of depth 2 with at most $\epsilon n \lambda_2(\frac{n}{k})$ edges, for some $1 \leq k = o(n)$, and some $\epsilon > 0$. Assume that $|L_1| \geq k$. Then, there exists a set $V \subset L_1$ of size $k \leq |V| = o(n)$ such that*

$$P_G[I_G, O_G, V] \leq \frac{100\epsilon^2 n^2}{|V|} \ .$$

For the proof of Lemma 3.1 we will need the following Lemma 3.2. For the proof of Lemma 3.2, the reader is referred to [Pud94] (Lemma 4).

Lemma 3.2 *[Pud94] Let $c_1 \geq c_2 \geq ... \geq c_t \geq 0$ be a sequence of real numbers, and let p, q be two integers such that $1 \leq p \leq q \leq t$. If for every $p \leq l \leq q$,*

$$\sum_{i=l}^{t} {c_i}^2 \geq \frac{1}{l} \ ,$$

then

$$\sum_{i=1}^{t} c_i \geq \frac{1}{2} \log\left(\frac{q}{p}\right) \ .$$

Proof of Lemma 3.1:
Denote $m = |L_1|$. Let $v_1, v_2, ..., v_m$ be the vertices of L_1, ordered according to their degree,

from highest to lowest. That is, for every i, $deg(v_i) \geq deg(v_{i+1})$. For every $k \leq l \leq \sqrt{nk}$, denote $V_l = \{v_1, ..., v_l\}$. Then, for every such l,

$$P_G[I_G, O_G, V_l] \leq \sum_{i=l+1}^{m} (deg(v_i))^2 \ .$$

Let $c_i = \frac{deg(v_i)}{10\epsilon n}$. Then,

$$\sum_{i=1}^{m} c_i = \frac{1}{10\epsilon n} \sum_{i=1}^{m} deg(v_i) = \frac{1}{5\epsilon n}|E_G| \leq \frac{1}{5}\left\lceil \log\left(\frac{n}{k}\right)\right\rceil < \frac{1}{2}\log\left(\frac{\sqrt{n}}{\sqrt{k}}\right) = \frac{1}{2}\log\left(\frac{\sqrt{nk}}{k}\right) \ .$$

Therefore, by Lemma 3.2, for some $k \leq l \leq \sqrt{nk} = o(n)$,

$$\sum_{i=l}^{m} {c_i}^2 < \frac{1}{l} \ .$$

Hence,

$$P_G[I_G, O_G, V_l] \leq \sum_{i=l}^{m} (deg(v_i))^2 = (100\epsilon^2 n^2) \sum_{i=l}^{m} c_i^2 \leq \frac{100\epsilon^2 n^2}{l} = \frac{100\epsilon^2 n^2}{|V_l|} \ .$$

□

As a corollary, we obtain our main lemma for $d = 2$.

Corollary 3.3 *Let G be a leveled graph of depth 2 with at most $\epsilon n \lambda_2(n)$ edges, for some $0 < \epsilon < 1/400$. Assume that $|L_1| \geq \sqrt{n}$. Then there exists a set $V \subset L_1$ of size $\sqrt{n} \leq |V| = o(n)$ such that*

$$P_G[I_G, O_G, V] \leq \frac{\epsilon n^2}{|V|}.$$

Proof:
Note that

$$\epsilon n \lambda_2(n) \leq 2\epsilon n \lambda_2\left(\frac{n}{\sqrt{n}}\right) \ .$$

By Lemma 3.1, there is a set $V \subset L_1$, of size $\sqrt{n} \leq |V| = o(n)$ such that,

$$P_G[I_G, O_G, V] \leq \frac{100(2\epsilon)^2 n^2}{|V|} \leq \frac{\epsilon n^2}{|V|} \ .$$

□

3.2 Reducing the depth by 2

Roughly speaking, the main lemma shows that if G has a small number of edges, then one can remove from G a small set J of inputs and outputs, and a small set V of other vertices, such that the total number of paths of length d from $I_G \setminus J$ to $O_G \setminus J$, that do not pass through vertices in V, is small (i.e., $P_G[\ I_G \setminus J\ ,\ O_G \setminus J\ ,\ V\ ]$ is small).

The idea of the proof for depth larger than 2 is as follows. We start with a graph G of depth d and we reduce it's depth by steps. In each step we will reduce the depth of G from d to $d-2$ by eliminating the levels L_1 and L_{d-1}. This is done in the following way. We partition $L_1 \cup L_{d-1}$ into 3 sets (A, B, C). We will then eliminate each one of these sets of vertices. However, we will have a different method to eliminate each one of these sets. The vertices in A are just removed from the graph. More formally, we just add the vertices of A to the set of vertices V (that are later on ignored). The set B is eliminated by removing all the inputs and outputs that are connected to it. More formally, we just add all the inputs and outputs that are connected to vertices in B to the set J of inputs and outputs (that are later on ignored). The set C is eliminated by connecting directly every pair of vertices that are connected by a path of length 2 through C. More formally, we replace each path of length 2 (v, c, w), such that $v \in I_G$, $c \in C$ and $w \in L_2$, by the one edge (v, w) (the same with L_{d-2} and O_G). The set C can hence be ignored. We continue to reduce the depth of G until we have a graph of depth 2 or 3 (w.l.o.g. 3). We then finish by proving the main lemma for depth 3.

To summarize, the proof is by induction on the depth. The proof uses a specific way to reduce the depth by 2. Given a partition (A, B, C) of $L_1 \cup L_{d-1}$, we will eliminate $L_1 \cup L_{d-1}$ by:

1. Removing A.

2. Removing all neighbors of B among $I_G \cup O_G$.

3. Adding an edge between vertices from I_G and L_2 that are connected through C (and the same with L_{d-2} and O_G) and then removing C.

This leads us to the following definition.

Definition 3.4 *Let G be a leveled graph of depth $d \geq 3$. Let (A, B, C) be a partition of $L_1 \cup L_{d-1}$. The graph $\hat{G}$ of depth $d-2$ is defined in the following way:*
The inputs of $\hat{G}$ are

$$I_{\hat{G}} = I_G \setminus \Gamma(B).$$

The outputs of $\hat{G}$ are

$$O_{\hat{G}} = O_G \setminus \Gamma(B).$$

The $d-1$ levels of $\hat{G}$ are

$$I_{\hat{G}}, L_2, ..., L_{d-2}, O_{\hat{G}}$$

(note that for $d = 3$ the levels of $\hat{G}$ are just $I_{\hat{G}}, O_{\hat{G}}$). The edges between the levels $L_2, ..., L_{d-2}$ are the same as they are in G. The other edges are defined in the following way:

- *For $d > 3$, we have to define the edges between $I_{\hat{G}}$ and L_2 and between L_{d-2} and $O_{\hat{G}}$. For every v, c, w such that $v \in I_{\hat{G}}$, $c \in C$, $w \in L_2$, and there are edges $(v, c), (c, w)$ in G,* *we add the edge (v, w) to $\hat{G}$ (i.e., we replace every such path of length 2 with an edge). In the same way, for every v, c, w such that $v \in L_{d-2}$, $c \in C$, $w \in O_{\hat{G}}$, and there are edges $(v, c), (c, w)$ in G, we add the edge (v, w) to $\hat{G}$.*

- *For $d = 3$, we have to define the edges between $I_{\hat{G}}$ and $O_{\hat{G}}$. This case is slightly different because we don't have a level between L_1 and L_2 to absorb the new edges. So instead, for every path of length 3 (v, c_1, c_2, w), such that, $v \in I_{\hat{G}}$, $w \in O_{\hat{G}}$ and $c_1, c_2 \in C$, we put an edge (v, w) in $\hat{G}$.*

Clearly $\hat{G}$ is a function of G, A, B, C. For convenience we will use the notation $\hat{G}$ instead of $\hat{G}(G, A, B, C)$.

Obviously, $\hat{G}$ is a multi-graph of depth $d-2$. In the construction of $\hat{G}$ we replaced each path of length 2 through C (or a path of length 3 in the case $d = 3$) with an edge. Therefore, we have the following easy corollary. The corollary shows that the number of relevant paths in the graph $\hat{G}$ is the same as it is in the graph G after removing the sets A and $\Gamma(B)$. Hence, in order to count paths in G it is enough to count the corresponding paths in $\hat{G}$. This easy observation makes the use of induction possible.

Proposition 3.5 *Let $G, A, B, C, \hat{G}$ be as in Definition 3.4. Then,*

1. $P_G[\, I_G \setminus \Gamma(B)\, ,\ O_G \setminus \Gamma(B)\, ,\ A\,] = P_{\hat{G}}[\, I_{\hat{G}}\, ,\ O_{\hat{G}}\, ,\ \emptyset\,]$.

2. *More generally: for any set of vertices $\hat{V} \subset L_2 \cup \cdots \cup L_{d-2}$, and any set of inputs and outputs $\hat{J} \subset I_{\hat{G}} \cup O_{\hat{G}}$,*

$$P_G[\, I_G \setminus (\hat{J} \cup \Gamma(B))\, ,\ O_G \setminus (\hat{J} \cup \Gamma(B))\, ,\ \hat{V} \cup A\,] = P_{\hat{G}}[\, I_{\hat{G}} \setminus \hat{J}\, ,\ O_{\hat{G}} \setminus \hat{J}\ ,\hat{V}\,]\ .$$

As mentioned above we will prove the main lemma by induction. In each step we reduce the depth by 2 by removing a set A of intermediate vertices and a set $\Gamma(B)$ of inputs and outputs. The final set V will be the union of the sets A from all steps of the induction, and the final set J will be the union of the sets $\Gamma(B)$ from all these steps.

In each step of the induction, we assume that the graph G has a relatively small number of edges. We would like to make sure that $\hat{G}$ also has a small number of edges. The next lemma shows that given certain bounds for $|A|$, $maxdeg(C)$, and for $|E_G|$, one can bound the number of edges in $\hat{G}$. The idea of the proof is that the only edges that we add to $\hat{G}$ are related to paths through C. Therefore (roughly speaking) we can bound $|E_{\hat{G}}|$ by $|E_G| \cdot maxdeg(C)$. Note that we assume here the bound

$$|E_G| \le \epsilon n \lambda_d \left(\frac{n}{k}\right)$$

for some $k \ge 1$. This is more general than the original assumption $|E_G| \le \epsilon n \lambda_d(n)$. The reason that we need a more general assumption is that the graph G may be a graph that was obtained after several steps of reduction (rather than the original graph). Roughly speaking, the parameter k corresponds to the number of intermediate vertices that were already removed from the original graph (the set V in the statement of the lemma). Since, we think of the set A as being removed from the graph as well (and being added to the set V) the bound that we want for $|E_{\hat{G}}|$ is

$$|E_{\hat{G}}| \leq \epsilon n \lambda_{d-2}\left(\frac{n}{k+|A|}\right).$$

Lemma 3.6 *Let $G, A, B, C, \hat{G}$ be as in Definition 3.4 and assume that for some $1 \leq k = o(n)$ and for some $0 < \epsilon < 1/3$,*

$$|E_G| \leq \epsilon n \lambda_d\left(\frac{n}{k}\right).$$

Assume that for some integer $1 \leq i \leq \frac{\lambda_d(\frac{n}{k})}{2} - 3$ we have:

1. *$|A| \leq \frac{2\epsilon n}{\lambda_{d-2}^{(i)}(\frac{n}{k})}$.*

2. *$maxdeg(C) \leq \lambda_{d-2}^{(i+3)}(\frac{n}{k})$.*

Then,

$$|E_{\hat{G}}| \leq \epsilon n \lambda_{d-2}\left(\frac{n}{k+|A|}\right).$$

The proof is by a straight forward calculation using Claim 2.2, and Claim 2.4.

Proof:
Since $k = o(n)$ we have:

$$k\lambda_{d-2}^{(i)}\left(\frac{n}{k}\right) \leq k\lambda_1\left(\frac{n}{k}\right) \leq k\sqrt{\frac{n}{k}} < \epsilon n \quad ,$$

therefore

$$k + |A| \leq k + \frac{2\epsilon n}{\lambda_{d-2}^{(i)}(\frac{n}{k})} = \frac{k\lambda_{d-2}^{(i)}(\frac{n}{k}) + 2\epsilon n}{\lambda_{d-2}^{(i)}(\frac{n}{k})} < \frac{3\epsilon n}{\lambda_{d-2}^{(i)}(\frac{n}{k})} < \frac{n}{\lambda_{d-2}^{(i)}(\frac{n}{k})}.$$

We now have two different cases that follow from the construction in Definition 3.4:

1. $d > 3$: ¿From the construction of $\hat{G}$ we get that the degree of each vertex in $L_2 \cup L_{d-2}$ has increased by a factor of $maxdeg(C)$ at most. Therefore we have:

$$|E_{\hat{G}}| \leq |E_G| \cdot maxdeg(C) \leq \epsilon n \lambda_d\left(\frac{n}{k}\right)\lambda_{d-2}^{(i+3)}\left(\frac{n}{k}\right) \leq$$

$$\leq 2\epsilon n\left(\lambda_{d-2}^{(i+3)}\left(\frac{n}{k}\right)\right)^2 \leq \epsilon n \lambda_{d-2}^{(i+2)}\left(\frac{n}{k}\right) \leq \epsilon n \lambda_{d-2}^{(2)}\left(\frac{n}{k+|A|}\right) < \epsilon n \lambda_{d-2}\left(\frac{n}{k+|A|}\right)$$

(where all inequalities are due to Claim 2.2, Claim 2.4 and the bound that we proved on $k + |A|$).

2. $d = 3$: Since we dropped $L_1 \cup L_2$, the set of inputs and outputs now absorbs both the edges of $C \cap L_1$ and the edges of $C \cap L_2$, so we have:

$$|E_{\hat{G}}| \leq |E_G| \cdot maxdeg(C)^2 \leq \epsilon n \lambda_d \left(\frac{n}{k}\right) \left(\lambda_{d-2}^{(i+3)}\left(\frac{n}{k}\right)\right)^2 \leq$$

$$\leq \frac{1}{2}\epsilon n \lambda_d \left(\frac{n}{k}\right) \lambda_{d-2}^{(i+2)}\left(\frac{n}{k}\right) \leq \frac{1}{2}\epsilon n \lambda_{d-2}^{(i+1)}\left(\frac{n}{k}\right) \leq \frac{1}{2}\epsilon n \lambda_{d-2}\left(\frac{n}{k+|A|}\right)$$

(where, as before, all inequalities are due to Claim 2.2, Claim 2.4 and the bound that we proved on $k + |A|$).

□

So far, we have presented the construction of $\hat{G}$ from G, given an arbitrary partition (A, B, C). In order to maintain the bound on the number of edges of $\hat{G}$, we need A, C to satisfy the conditions of Lemma 3.6. Also, we need $\Gamma(B)$ to be not too large, in order to make sure that the total number of inputs and outputs removed in the process is small. The next lemma shows how to partition $L_1 \cup L_{d-1}$ into suitable sets (A, B, C). The way it is done is by first ordering the vertices in $L_1 \cup L_{d-1}$ according to their degrees (from highest to lowest), and then finding appropriate numbers $r_1 > r_2$, such that A will be the set of vertices with degree larger than r_1, B will be the set of vertices with degree larger than r_2 and at most r_1, and C will be the set of vertices with degree at most r_2.

Lemma 3.7 *Let G be a leveled graph of depth $d \geq 3$, such that, $|E_G| \leq \epsilon n \lambda_d(r)$ for large enough r (more accurately, we need $\lambda_d(r) > 72$). Then, there exists a partition (A, B, C) of $L_1 \cup L_{d-1}$, and $1 \leq i \leq \lambda_d(r)/2 - 3$ with the following properties:*

1. $|A| \leq \frac{2\epsilon n}{\lambda_{d-2}^{(i)}(r)}$.
2. $|\Gamma(B)| \leq 9\epsilon n$.
3. $maxdeg(C) \leq \lambda_{d-2}^{(i+3)}(r)$.

Proof:
Denote,

$$W_0 = \{v \in L_1 \cup L_{d-1} \mid deg(v) > \lambda_{d-2}(r)\},$$

and for $i \geq 1$,

$$W_i = \{v \in L_1 \cup L_{d-1} \mid \lambda_{d-2}^{(i)}(r) \geq deg(v) > \lambda_{d-2}^{(i+1)}(r)\}.$$

Claim 3.8 *For every $1 \leq i \leq \lambda_d(r)/2 - 3$,*

$$|W_0 \cup W_1 \cup \cdots \cup W_{i-1}| \leq \frac{2\epsilon n}{\lambda_{d-2}^{(i+1)}(r)}.$$

Proof: The proof follows from the fact that the degree of each vertex in $W_0 \cup W_1 \cup \cdots \cup W_{i-1}$ is at least $\lambda_{d-2}^{(i)}(r)$. If the claim wasn't true we would have had

$$|E_G| \geq |E(W_0 \cup W_1 \cup \cdots \cup W_{i-1})| \lambda_{d-2}^{(i)}(r) \geq \frac{2\epsilon n}{\lambda_{d-2}^{(i+1)}(r)} \lambda_{d-2}^{(i)}(r) \geq$$

$$\geq 2\epsilon n \lambda_{d-2}^{(i+1)}(r) \geq \epsilon n \lambda_d(r) \ ,$$

in contradiction. □

Claim 3.9 *For some* $0 \leq i \leq \lambda_d(r)/2 - 4$,

$$|E(W_i \cup W_{i+1} \cup W_{i+2} \cup W_{i+3})| \leq 9\epsilon n.$$

Proof: The proof follows from the bound on $|E_G|$. Since $|E_G| \leq \epsilon n \lambda_d(r)$ we must have

$$\sum_{i=0}^{\frac{\lambda_d(r)/2-4}{4}} |E(W_{4i} \cup W_{4i+1} \cup W_{4i+2} \cup W_{4i+3})| \leq \epsilon n \lambda_d(r) \ .$$

If each of the sets $E(W_{4i} \cup W_{4i+1} \cup W_{4i+2} \cup W_{4i+3})$ was of size larger than $9\epsilon n$ then we would have had

$$|E_G| \geq 9\epsilon n \frac{\lambda_d(r)/2 - 4}{4} > \epsilon n \lambda_d(r)$$

(for large enough r), in contradiction. □

Fix i' to be such that Claim 3.9 is satisfied for i'. The proof of the lemma now follows for $i = i' + 1$, by taking

1. $A = W_0 \cup W_1 \cup \cdots \cup W_{i'-1}$.
2. $B = W_{i'} \cup W_{i'+1} \cup W_{i'+2} \cup W_{i'+3}$.
3. $C = (L_1 \cup L_{d-1}) \setminus (A \cup B)$.

□

3.3 Depth 3

The induction that we are about to perform on the depth will end with a graph of depth 2 or 3 (w.l.o.g., 3). We will now give the proof of the main lemma for the special case $d = 3$. As mentioned before, we need a more general lemma that assumes a more general bound for the number of edges in the graph.

The proof for $d = 3$ already gives the main idea of the proof for the general case. First we partition $L_1 \cup L_2$ into 3 sets (A, B, C) as described before. The partition (A, B, C) will satisfy the following: The first set A is small, the second set B is not connected to many inputs and outputs, and each vertex in the third set C has small degree. We then reduce the depth of the graph to 1, using Definition 3.4. Thus we get a graph, $\hat{G}$, with the same number of input-output paths, between $I_{\hat{G}}$ and $O_{\hat{G}}$. Since $\hat{G}$ is a graph of depth 1, the number of such paths is simply the number of edges. The last step would be to calculate the number of edges in $\hat{G}$, which is done using Lemma 3.6.

Lemma 3.10 *Let G be a leveled graph of depth 3 with at most $\epsilon n \lambda_3(\frac{n}{k})$ edges, for some $1 \leq k = o(n)$ and for some $0 < \epsilon < 1/3$. Then, there exists a partition (A, B, C) of $L_1 \cup L_2$ such that:*

1. $|A| = o(n)$,
2. $|\Gamma(B)| \leq 9\epsilon n$,
3. $P_G[\ I_G \setminus \Gamma(B)\ ,\ O_G \setminus \Gamma(B)\ ,\ A\ ] \leq \epsilon n \lambda_1\left(\frac{n}{k+|A|}\right)$.

Proof:
By Lemma 3.7, we can partition $L_1 \cup L_2$ into 3 sets (A, B, C), such that for some $1 \leq i \leq \lambda_3(\frac{n}{k})/2 - 3$, we have:

1. $|A| \leq \frac{2\epsilon n}{\lambda_1^{(i)}(\frac{n}{k})} = o(n)$.
2. $|\Gamma(B)| \leq 9\epsilon n$.
3. $maxdeg(C) \leq \lambda_1^{(i+3)}(\frac{n}{k})$.

Let $\hat{G}$ be as in Definition 3.4 (with respect to (A, B, C)). By Proposition 3.5,

$$P_G[\ I_G \setminus \Gamma(B)\ ,\ O_G \setminus \Gamma(B)\ ,\ A\ ] = P_{\hat{G}}[\ I_{\hat{G}}\ ,\ O_{\hat{G}}\ ,\ \emptyset\ ].$$

Since $\hat{G}$ is of depth 1, the right hand-side equals $|E_{\hat{G}}|$. By Lemma 3.6,

$$|E_{\hat{G}}| \leq \epsilon n \lambda_1\left(\frac{n}{k + |A|}\right) .$$

Hence,

$$P_G[\ I_G \setminus \Gamma(B)\ ,\ O_G \setminus \Gamma(B)\ ,\ A\ ] \leq \epsilon n \lambda_1\left(\frac{n}{k + |A|}\right) .$$

□

3.4 Higher depth

We are now ready to state and prove our main lemma. As mentioned above, we actually prove a stronger lemma that will be needed for the induction. The main lemma will then follow as an easy corollary. First note that the functions $\lambda_d(n)$ satisfy $\lambda_{2i}(n) = \theta(\lambda_{2i+1}(n))$. Hence, we can assume w.l.o.g. that the depth d is odd, otherwise we can just increase the depth by 1. (We couldn't prove better results for the even levels because of the constructions given in [DDPW83]).

Lemma 3.11 *Let G be a leveled graph of constant depth $d \geq 3$, (and assume that d is odd), with at most $\epsilon n \lambda_d(\frac{n}{k})$ edges, for some $1 \leq k = o(n)$ and for some $0 < \epsilon < 1/3$. Then there exist a set V of vertices and a set J of inputs and outputs, such that:*

1. $|V| = o(n)$,
2. $|J| \leq 5\epsilon dn$,
3. $P_G[\, I_G \setminus J \ , \ O_G \setminus J \ , \ V \,] \leq \epsilon n \lambda_1 \left(\frac{n}{k+|V|}\right)$.

Proof:
The proof is by induction on d. The base case $d = 3$ was proved in Lemma 3.10. For $d \geq 5$, let (A, B, C) be the partition of $L_1 \cup L_{d-1}$ from Lemma 3.7. Let $\hat{G}$ be the depth $d-2$ graph defined in Definition 3.4 (with respect to (A, B, C)). Then by Lemma 3.6,

$$|E_{\hat{G}}| \leq \epsilon n \lambda_{d-2} \left(\frac{n}{k+|A|}\right) .$$

Since $|A| = o(n)$, we have that $k + |A| = o(n)$ and hence the inductive assumption holds for $\hat{G}$. Hence for $\hat{G}$, there exist a set $\hat{V}$ of vertices and a set $\hat{J}$ of inputs and outputs, such that:

1. $|\hat{V}| = o(n)$,
2. $|\hat{J}| \leq 5\epsilon(d-2)n$,
3. $P_{\hat{G}}[\, I_{\hat{G}} \setminus \hat{J} \ , \ O_{\hat{G}} \setminus \hat{J} \ , \ \hat{V} \,] \leq \epsilon n \lambda_1 \left(\frac{n}{k+|A|+|\hat{V}|}\right)$.

Define $V = \hat{V} \cup A$ and $J = \hat{J} \cup \Gamma(B)$. Then since $|\hat{V}|, |A| = o(n)$ we have $|V| = o(n)$ (note that since the original depth is constant, the total number of induction steps is constant). Since $|\hat{J}| \leq 5\epsilon(d-2)n$ and $|\Gamma(B)| \leq 9\epsilon n$ we have that $|J| \leq 5\epsilon dn$.

Finally, by Proposition 3.5,

$$P_G[\, I_G \setminus J \ , \ O_G \setminus J \ , \ V \,] = P_{\hat{G}}[\, I_{\hat{G}} \setminus \hat{J} \ , \ O_{\hat{G}} \setminus \hat{J} \ , \ \hat{V} \,] \leq$$

$$\leq \epsilon n \lambda_1 \left(\frac{n}{k+|A|+|\hat{V}|}\right) = \epsilon n \lambda_1 \left(\frac{n}{k+|V|}\right) .$$

□

Our main lemma is now stated as the following corollary. Note that the requirement $\epsilon < 1/400$ is only needed for the case $d = 2$ (a weaker requirement is needed for $d > 2$).

Corollary 3.12 *Let G be a leveled graph of constant depth $d \geq 2$, with more than n vertices and less than $\epsilon n \lambda_d(n)$ edges, for some $0 < \epsilon < 1/400$. Then there exist a set V of vertices and a set J of inputs and outputs, such that:*

1. $\sqrt{n} \leq |V| = o(n)$,

2. $|J| \leq 5\epsilon dn$,

3. $P_G[\, I_G \setminus J \ , \ O_G \setminus J \ , \ V \,] \leq \epsilon \frac{n^2}{|V|}$.

Proof:
First note that (as mentioned above) Lemma 3.11 is correct also for even depth if we just require ϵ to be slightly smaller (a factor of 2 or 3 is enough). Since here we require $\epsilon < 1/400$ we can apply Lemma 3.11 for any depth larger than 2.

We apply Lemma 3.11 with $k = 1$. If $|V| \geq \sqrt{n}$ we are done. Otherwise, just add arbitrary vertices to V. Call the resulting set V'. We have

$$P_G[\, I_G \setminus J \ , \ O_G \setminus J \ , \ V' \,] \leq P_G[\, I_G \setminus J \ , \ O_G \setminus J \ , \ V \,] \leq$$

$$\leq \epsilon n \lambda_1 \left(\frac{n}{1+|V|} \right) \leq \epsilon n \sqrt{\frac{n}{1+|V|}} = \epsilon \frac{n^2}{\sqrt{n(1+|V|)}} \leq \epsilon \frac{n^2}{|V'|} \ .$$

This completes the proof for depth higher than 2. For $d = 2$, the corollary was already stated as Corollary 3.3. □

3.5 Graphs for matrix product

As mentioned in the introduction, the main function that we concentrate on in this work is matrix product. A circuit for matrix product has $2m^2$ inputs and m^2 outputs, where m is the dimension of each matrix. Therefore, we will assume here that our graph has $2m^2$ inputs and m^2 outputs and that the outputs are ordered as a matrix. We will refer to such a graph as a graph for matrix product. For convenience, we will prove here a lemma that will be specific for such graphs. The proof will follow easily by Corollary 3.12.

Denote by O_i the outputs in the i^{th} column of the output matrix. Denote by $[m]$ the set $\{1, ..., m\}$. We think of $[m]$ as the set of all output columns. For a subset $D \subset [m]$, denote by O_D the outputs in all the columns in D. That is, $O_D = \cup_{i \in D} O_i$.

Roughly speaking, our lemma will state that after removing from the graph a small set of inputs and outputs (I and O) and a set V of size k of intermediate nodes, one can find a set D of $10k/m$ output columns, such that there are no paths between the inputs and the outputs in O_D. For simplicity, we will not state the lemma for a general constant ϵ and just fix some constant that will be good enough.

Lemma 3.13 *Let G be a leveled graph for matrix product, of constant depth $d \geq 2$, with less than $\epsilon m^2 \lambda_d(m^2)$ edges, for $\epsilon = 1/(1000 \cdot d)$. Then there exist sets $V \subset V_G$, $D \subset [m]$, $O \subset O_D$, $I \subset I_G$, such that:*

1. $m \leq |V| = o(m^2)$.

2. $|D| \geq \frac{10|V|}{m}$.

3. For every $i \in D$, $|O_i \cap O| < \frac{1}{10}m$.

4. $|I| < \frac{1}{10}m^2$.

5. $P_G[\ I_G \setminus I\ ,\ O_D \setminus O\ ,\ V\ ] = 0$.

Proof:
Let V, J be the sets guaranteed by Corollary 3.12 with $n = m^2$. Define $\tilde{I} = J \cap I_G$ and $\tilde{O} = J \cap O_G$, and let $k = |V|$. Then

1. $m \le |V| = k = o(n)$. (hence, requirement (1) is satisfied).
2. $|\tilde{I}|, |\tilde{O}| \le (1/200) \cdot m^2$.
3. $P_G[\ I_G \setminus \tilde{I}\ ,\ O_G \setminus \tilde{O}\ ,\ V\ ] \le \epsilon \frac{m^4}{k}$.

Denote by $\tilde{D}$ the set of all $i \in [m]$ such that $|O_i \cap \tilde{O}| < (1/100) \cdot m$. Then by the bound we have on $|\tilde{O}|$ we know that

$$|\tilde{D}| \ge (1/2) \cdot m.$$

For every $i \in \tilde{D}$, let $P(i)$ be the total number of paths between outputs in $O_i \setminus \tilde{O}$ and inputs in $I_G \setminus \tilde{I}$ that do not pass through V. Denote by D the set of $\lceil 10k/m \rceil$ indices i with the smallest $P(i)$. (then by the definition of D, requirement (2) is satisfied). Denote $O = \tilde{O} \cap O_D$. (then by the fact that $D \subset \tilde{D}$, requirement (3) is satisfied). Since

$$\sum_{i \in \tilde{D}} P(i) \le P_G[\ I_G \setminus \tilde{I}\ ,\ O_G \setminus \tilde{O}\ ,\ V\ ] \le \epsilon \frac{m^4}{k},$$

we have,

$$\sum_{i \in D} P(i) \le \frac{|D|}{|\tilde{D}|} \sum_{i \in \tilde{D}} P(i) \le \frac{\lceil 10k/m \rceil}{m/2} \cdot \epsilon \frac{m^4}{k} < (1/50) \cdot m^2.$$

Hence, if we denote by $\hat{I}$ the set of all inputs that are connected by a path (that do not pass through V) to some output in $O_D \setminus O$ we have

$$|\hat{I}| < (1/50) \cdot m^2.$$

Denote $I = \tilde{I} \cup \hat{I}$. (then by the bounds we have on $|\tilde{I}|, |\hat{I}|$, requirement (4) is satisfied). By the definition of $\hat{I}$, all paths between $O_D \setminus O$ and $I_G \setminus I$ pass through V. (hence requirement (5) is satisfied). □

4 Arithmetic Model

In this section we present our results for bilinear arithmetic circuits. An arithmetic circuit 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 $v_1, ..., v_k$ 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, each node in the circuit computes a polynomial in the input variables.

In this section we prove lower bounds on the size of circuits computing the product of two $m \times m$ matrices. The input is of size $n = 2m^2$, and it consists of two $m \times m$ matrices X, Y. The output is the matrix $Z = X \cdot Y$, i.e., there are m^2 outputs, and the $(i, j)^{th}$ output is:

$$z_{i,j} = \sum_{k=1}^{m} x_{i,k} \cdot y_{k,j} \ .$$

Each output $z_{i,j}$ is hence a bilinear form in X and Y.

Since the product of two matrices 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 a product gate is only allowed to compute the product of two linear functions, one in the variables of X and the other 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 now define the *size* and *depth* of the circuit.

Definition 4.1 *For a bilinear circuit C, we denote by $s(C)$ (the size of C) the number of edges between the product gates and the outputs. We denote by $d(C)$ (the depth of C) the length of the longest directed path from a product gate to an output.*

Note that these definitions ignore all gates and edges below the product gates (i.e., between the inputs and the products). That is, we allow the circuit to get for free any number of linear functions in the variables of X, and any number of linear functions in the variables of Y. We only count the size and depth above the product gates.

The requirement that the circuit is bilinear seems restrictive. It is easy to show, however, that over fields of characteristic zero, the bilinearity assumption does not change (up to a constant factor) the size and the depth of the circuit. More accurately, by paying a constant factor in the size and in the depth, we can transform any arithmetic circuit computing a bilinear form into an equivalent bilinear circuit. Roughly speaking, this is done in the following way: Since the circuit computes polynomials of degree two, it doesn't really need to keep track of any monomial of higher degree. Therefore, we only need to keep track of monomials of degree one or two, and we can do that by replacing each gate by a constant number of new gates (at most 5 new gates) that satisfy the bilinearity assumption. Thus we have:

Proposition 4.2 *If a set of bilinear forms is computed by an arithmetic circuit of depth d and size s over a field of characteristic zero, then there is a bilinear circuit of depth $3d$ and size $5s$ over the same field computing the same set of bilinear forms.*

Over finite fields, the bilinearity assumption may be restrictive. We prove lower bounds for the general case of arithmetic circuits over finite fields in section 5. It is also worth noting that all known algorithms for matrix product are by bilinear arithmetic circuits.

Our main tool in proving our lower bounds is Lemma 3.13. Since that lemma is stated for leveled graphs, we would like our circuit to be leveled. We hence assume that our circuit is a leveled bilinear arithmetic circuit. Since we consider leveled circuits, the depth of the circuit is just the number of levels above the product gates. Our main result in this section is the following lower bound:

Theorem 4.3 *Any leveled bilinear arithmetic circuit C of depth d, for the product of two $m \times m$ matrices, is of size:*

$$s(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega\left(\frac{1}{d} m^2 \lambda_d(m^2)\right) & d > 1 \end{cases}$$

In order to prove lower bounds for a non-leveled bilinear circuit, we just level it. We can do that by increasing it's size by a factor of d. We can then use the lower bounds for leveled circuits. We hence have the following corollary:

Corollary 4.4 *Any bilinear arithmetic circuit C of depth d, for the product of two $m \times m$ matrices, is of size:*

$$s(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega\left(\frac{1}{d^2} m^2 \lambda_d(m^2)\right) & d > 1 \end{cases}$$

In particular, this gives the following size-depth tradeoff: there is no linear size and constant depth bilinear arithmetic circuit for the product of two matrices. Note that the theorem is valid for any field. It just needs the bilinearity assumption.

After proving the main theorem, we will use the result of Baur and Strassen [BS82] to prove a lower bound on the size of bounded depth arithmetic circuits for the trace of the product of 3 matrices:

$$\sum_{i,j,k=1}^{m} x_{i,j} \cdot y_{j,k} \cdot z_{k,i} \ ,$$

which is a function with a single output.

Let us start with bilinear circuits of depth 1. The structure of such circuits is very simple. First they compute linear forms. Then there is one level of product gates computing bilinear forms. Finally, there is one level of m^2 plus gates computing the outputs. For the proof, we will use the following notation.

$$O_j = \{ \ z_{i,j} \mid i \in \{1, ..., m\} \ \} \ ,$$

i.e., O_j denotes the set of outputs of the j^{th} column of the output matrix.

Theorem 4.5 *Any leveled bilinear circuit C of depth 1, for the product of two $m \times m$ matrices, is of size:*

$$s(C) = \Omega(m^3) .$$

Proof:
Since the circuit is of depth one, the outputs come right after the product gates. Each output is computed by a plus gate that adds the results of the product gates that are connected to it. We will show that there are at least m^2 edges connected to each output column O_j. Hence, since there are m output columns, the result follows.

Assume for a contradiction that an output column O_j is connected to $r < m^2$ product gates. Denote the functions computed by these gates by $M_1, ..., M_r$. For each M_k denote by $L_{k,1}(X), L_{k,2}(Y)$ the two linear functions that it multiplies. That is,

$$M_k(X, Y) = L_{k,1}(X) \cdot L_{k,2}(Y).$$

Since $r < m^2$, we can find a substitution to the matrix X such that:

1. $X \neq 0$.
2. For every $1 \leq k \leq r, \;\; L_{k,1}(X) = 0$.

Hence, for every $1 \leq k \leq r$,

$$M_k(X, Y) = 0.$$

Therefore, $O_j = 0$, no matter what Y is. At the other hand, $X \neq 0$, and hence we can find a substitution for the matrix Y such that no column of $X \cdot Y$ is all zero (a contradiction). □

The main idea of the proof for depth 1 was the following: if there is a small number of edges in the circuit then one can find a substitution for the matrix X, such that a large number of outputs are forced to be 0 (no matter what Y is). The main idea of the proof for larger depth is the following: First apply Lemma 3.13 to transform the circuit into a circuit of depth 1. This is done by removing from the circuit a certain number of inputs, outputs and intermediate gates. Then use the argument for depth 1. However, since we remove from the circuit a certain number of nodes, we will need a more general argument. Roughly speaking, we will need to generalize the proof for depth 1 to the case where X and Y are restricted to certain subspaces of matrices. We will show that even if X and Y are restricted to (not too small) subspaces, their product can still not be computed by a small circuit of depth 1. We will use the following notations.

Definition 4.6 *For a matrix X, denote by $(X)_j$ the j^{th} column of X. For a linear subspace of matrices $\mathcal{A}$, denote*

$$(\mathcal{A})_j = \{ \; (X)_j \mid X \in \mathcal{A} \; \} .$$

Since $\mathcal{A}$ is a linear subspace then so is $(\mathcal{A})_j$.

We clearly have:

Proposition 4.7 *For any linear subspace of matrices $\mathcal{A}$,*

$$dim(\mathcal{A}) \leq \sum_{j=1}^{m} dim((\mathcal{A})_j) \; .$$

We will also need the following lemma. Roughly speaking, the lemma claims that if X is a matrix of large rank and $\mathcal{B}$ is a subspace of matrices of high dimension, then for many columns j, $dim(\{\ (X \cdot Y)_j \mid Y \in \mathcal{B}\})$ is high.

Lemma 4.8 *Let X be an $m \times m$ matrix of rank $\geq \frac{2}{3}m$. Let $\mathcal{B}$ be a linear subspace of $m \times m$ matrices, such that $dim(\mathcal{B}) \geq m^2 - k$. Then for any subset of columns, D, of size $|D| \geq \frac{10k}{m}$, there exists a column $j \in D$ such that*

$$dim(\{\ (X \cdot Y)_j \mid Y \in \mathcal{B}\ \}) \geq \frac{m}{2} \; .$$

Proof:
Let D be a subset of columns such that $|D| \geq \frac{10k}{m}$. We first show that there is a column $j \in D$ such that $dim((\mathcal{B})_j) \geq \frac{9}{10}m$. If this wasn't the case then by Proposition 4.7 we would have had:

$$m^2 - k \leq dim(\mathcal{B}) \leq |D| \left(\frac{9}{10}m - 1\right) + (m - |D|)m = m^2 - \frac{m}{10}|D| - |D| < m^2 - k \; .$$

Let $j \in D$ be such that $dim((\mathcal{B})_j) \geq \frac{9}{10}m$. Since $rank(X) \geq \frac{2}{3}m$, we have

$$dim(\{\ (X \cdot Y)_j \mid Y \in \mathcal{B}\ \}) = dim(\{\ X \cdot v \mid v \in (\mathcal{B})_j\ \}) \geq rank(X) \; - \; (m - \; dim((\mathcal{B})_j)) \geq$$

$$\geq \frac{2}{3}m - \frac{1}{10}m > \frac{m}{2} \; .$$

□

In the proof for depth 1, we had $X \neq 0$. Since in the proof for higher depth Y will be restricted to a subspace of matrices, we will need X to satisfy a stronger condition. Namely, we need X to be of high rank. However, X itself will also be restricted to a subspace of matrices. Therefore, we want to show that in any subspace (of matrices) of high dimension, there is a matrix of high rank.

Lemma 4.9 *In any subspace of $m \times m$ matrices of dimension larger than $(2mr - r^2 + m)$, there is a matrix of rank at least r.*

Proof:
We have two different proofs. The first is for finite fields and the second is for fields of characteristic zero. We wish to compare the number of matrices with rank at most r to the number of matrices in our linear subspace. If we prove that the number of matrices in the linear subspace is larger, then it must contain a matrix of rank larger than r.

- Assume that F is a finite field. Denote $|F| = q$. The number of elements in a subspace of dimension larger than $(2mr - r^2 + m)$ is larger than q^{2mr-r^2+m}. We will now count the number of $m \times m$ matrices with rank at most r. Note that for every such matrix there are r rows, such that every row in the matrix is in their linear span. There are $\binom{m}{r}$ possible ways to choose these r rows. There are q^{mr} possible ways to choose the r vectors for these rows. Every other row is in the linear span of these r rows, so it can be one of q^r vectors. Therefore the number of matrices of rank at most r is bounded from above by:
$$\binom{m}{r} q^{mr} (q^r)^{m-r} < q^{2mr-r^2+m} \ .$$
- Assume that F is a field of characteristic zero. Instead of counting, we will consider the dimension of the variety of matrices with rank at most r. The same argument as above shows that this variety is included in the union of $\binom{m}{r}$ varieties, each of dimension at most $mr + r(m-r)$, (as before, mr is for the freedom of choice of the first r vectors, and $r(m-r)$ is for spanning the other $m-r$ rows). Therefore the dimension is $2mr - r^2 < 2mr - r^2 + m$.

□

Before giving the formal proof for Theorem 4.3, let us first give a sketch of this proof. Let C be a leveled bilinear arithmetic circuit of depth d for the product of two $m \times m$ matrices. Let G be the leveled graph of depth d, corresponding to the graph of the circuit between the product gates and the outputs (i.e., the product gates are the inputs of the graph, the outputs are the outputs of the graph and the levels of the circuit between the product gates and the outputs are the levels of the graph). We would like to prove that
$$s(C) \geq \Omega\left(\frac{1}{d} m^2 \lambda_d(m^2)\right) \ .$$
Assume for a contradiction that $s(C) < \frac{1}{1000d} m^2 \lambda_d(m^2)$, or in other words
$$|E_G| < \frac{1}{1000d} m^2 \lambda_d(m^2) \ .$$
By Lemma 3.13, we can find a set of columns, D, and 3 sets of vertices, V, I, O, in the graph G, such that:

- $m \leq |V| = k = o(m^2)$.
- I is a small set of inputs.
- D is a set of $\lceil \frac{10k}{m} \rceil$ output columns.
- For every $i \in D$, $|O_i \cap O|$ is small (where O_i is the i^{th} output column).
- All the paths from $O_D \setminus O$ to the inputs pass through V or reach I.

We will derive a contradiction in 4 steps:

1. Since I is a small set of product gates, we can find a subspace of matrices, $\mathcal{A}$, such that for every matrix X in $\mathcal{A}$ all the gates in I output 0. The matrix X will be restricted to the subspace $\mathcal{A}$.

2. Note that once a matrix X is fixed, the nodes of V just compute linear functions in the variables of Y. Therefore, for every matrix $X \in \mathcal{A}$, we can find a subspace of matrices $\mathcal{B}_X$ such that for every pair $(X \in \mathcal{A}\ ,\ Y \in \mathcal{B}_X)$ all the gates in V output zero. Since V is a small set, the dimension of $\mathcal{B}_X$ is high. The matrix Y will be restricted to the subspace $\mathcal{B}_X$.

3. The subspace $\mathcal{A}$ is of high dimension, and for every X the subspace $\mathcal{B}_X$ is of high dimension. Therefore, we can find $X \in \mathcal{A}$ such that $dim((X \cdot \mathcal{B}_X)_j)$ is large for some $j \in D$ (this will follow from Lemma 4.8, and Lemma 4.9).

4. Since we restrict $X \in \mathcal{A}$ and $Y \in \mathcal{B}_X$, all the gates in V and I output zero. Therefore $X \cdot Y$ is computed by a circuit with no paths between $O_D \setminus O$ and $I_G \setminus I$. Hence, all the outputs in $O_D \setminus O$ must give zero. Because of the third step, this is a contradiction.

Let us now give the formal proof.

Proof of Theorem 4.3:
We already gave the proof for $d = 1$, so assume $d > 1$. Assume for a contradiction that we have a leveled bilinear arithmetic circuit C of depth d for the product of two $m \times m$ matrices, and such that

$$s(C) < \frac{1}{1000d} m^2 \lambda_d(m^2) \ .$$

Let G be the leveled graph of depth d between the product gates and the outputs of C (as explained above).

As before, denote by O_j the j^{th} output column. As before, for a set $D \subset [m]$ we denote $O_D = \cup_{i \in D} O_i$. By Lemma 3.13 there exist sets $V \subset V_G$, $D \subset [m]$, $O \subset O_D$, $I \subset I_G$, such that:

1. $m \leq |V| = o(m^2)$.

2. $|D| \geq \frac{10|V|}{m}$.

3. For every $i \in D$, $|O_i \cap O| < \frac{1}{10}m$.

4. $|I| < \frac{1}{10}m^2$.

5. $P_G[\ I_G \setminus I\ ,\ O_D \setminus O\ ,\ V\ ] = 0$.

Denote $k = |V|$. Hence, we have a set of outputs O_D consisting of at least $\frac{10k}{m}$ output columns, such that each of the output columns in O_D has a small intersection with O, and there are no paths between $O_D \setminus O$ and $I_G \setminus I$ that do not pass through V.

In the circuit C, the set I is just a set of product gates. Since

$$|I| < \frac{1}{10}m^2 ,$$

we can find a subspace of matrices $\mathcal{A}$ of dimension

$$dim(\mathcal{A}) \geq m^2 - |I| > \frac{9}{10}m^2 ,$$

such that for every $X \in \mathcal{A}$, all the product gates in I give zero.

Assume that a matrix X is fixed. All the functions computed by the vertices of V are now linear functions in the variables of Y. Therefore, for every matrix X, there is a subspace of matrices, $\mathcal{B}_X$, such that for every $Y \in \mathcal{B}_X$, all the gates in V give zero, and such that

$$dim(\mathcal{B}_X) \geq m^2 - |V| = m^2 - k .$$

Since

$$dim(\mathcal{A}) \geq \frac{9}{10}m^2,$$

by Lemma 4.9 we can find a matrix $X \in \mathcal{A}$, such that $rank(X) \geq \frac{2}{3}m$. We fix this X. Since $|D| \geq \frac{10k}{m}$ and $dim(\mathcal{B}_X) \geq m^2 - k$, we can apply Lemma 4.8 to get a column $j \in D$, such that

$$dim((X \cdot \mathcal{B}_X)_j) \geq \frac{m}{2} .$$

At the other hand, for $X \in \mathcal{A}$ and $Y \in \mathcal{B}_X$, all the product gates in I and all the gates in V output zero. Since all the paths to $O_D \setminus O$ pass through V or I, we get that for every $X \in \mathcal{A}$ and $Y \in \mathcal{B}_X$, all the outputs in $O_D \setminus O$ must give zero. Therefore, for every $j \in D$,

$$dim((X \cdot \mathcal{B}_X)_j) \leq |O_j \cap O| < \frac{m}{10} ,$$

(a contradiction). □

Theorem 4.3 gives a superlinear lower bound for a multi-output function. The following theorem of [BS82] (it is an immediate corollary of the results presented there) shows that we can also obtain a superlinear lower bound for a single-output function.

Theorem 4.10 *[BS82] Suppose that $f(x_1, ..., x_n)$ is computed by an arithmetic circuit of size s and depth d over a field of characteristic zero, then there is an arithmetic circuit of size $3s$ and depth $2d$ that computes $f, \frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n}$.*

(note that in this theorem the size and depth of a circuit are just the usual size and depth, i.e., the size is the number of edges and the depth is the length of the longest directed path).

Consider the following function:

$$f(X, Y, Z) = \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} x_{i,j}y_{j,k}z_{k,i} = trace(X \cdot Y \cdot Z) ,$$

where X, Y, Z are $m \times m$ matrices. Notice that

$$\frac{\partial f}{\partial z_{k,i}} = \sum_{j=1}^{m} x_{i,j} y_{j,k} = (X \cdot Y)_{i,k} \ .$$

Therefore the partial derivatives of f with respect to the $Z_{k,i}$'s are the outputs of the product of two matrices. If we take into account the price that we have to pay when transforming a circuit to a bilinear leveled one we get:

Theorem 4.11 *Every arithmetic circuit C of depth d, that computes the trace of the product of three $m \times m$ matrices, over a field of characteristic zero, is of size:*

$$\Omega\left(\frac{1}{d^2} m^2 \lambda_{6d}(m^2)\right) \ .$$

We can also generalize this result for circuits over finite fields, but we must make another assumption: the circuit computes

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} x_{i,j} \cdot y_{j,k} \cdot z_{k,i} \ ,$$

as a polynomial and not as a function. As mentioned in the introduction, over finite fields there are many polynomials that represent the same function. For example, $x^p - x = 0$ over a field with p elements. Therefore, we demand that the circuit computes this exact polynomial. In this case we can apply the theorem of Baur and Strassen as before and get the same lower bound as in the case of characteristic zero.

5 Circuits with Arbitrary Gates

In this section, we prove lower bounds for circuits with arbitrary gates over finite fields. Since the proofs are similar for all finite fields, we will only give in details the proofs for the field $GF(2)$, that is, the Boolean case. The proofs for other fields are only sketched.

A Boolean circuit with arbitrary gates 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. All nodes other than the inputs are labeled with arbitrary Boolean functions. That is, if v is a node of in-degree k, then v is labeled with some function

$$g_v : \{0,1\}^k \to \{0,1\} \ .$$[1]

The inputs to the circuit are Boolean variables and each node in the circuit computes in a natural way a Boolean function in the original input variables. The size of a circuit C is denoted by $size(C)$ and is defined to be the number of edges in it. The depth of a circuit is

[1] Since g_v is not necessarily a symmetric function, we need to order the inputs to v, so we know which input is the first variable of g_v etc.

defined to be the length of the longest directed path from an input to an output in the circuit. Note that the standard definition of a Boolean circuit requires $g_v \in \{\vee, \wedge, \neg\}$. Our definition is more general and allows g_v to be any function.

The model of Boolean circuits with arbitrary gates includes all other models of Boolean circuits, e.g.: standard Boolean circuits, Boolean circuits with threshold gates, Boolean circuits with $MODP$ gates etc'. For some of these models almost nothing is known. For example, for constant depth threshold circuits only slightly super-linear lower bounds are known [IPS97] (exponential lower bounds are known for depth 2 [HMPST87]).

We will mainly concentrate on matrix product over $GF(2)$. Our main result in this section is the following lower bound:

Theorem 5.1 *Any leveled Boolean circuit with arbitrary gates, C, of depth d, for the product of two $m \times m$ Boolean matrices over $GF(2)$, is of size:*

$$size(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega(\frac{1}{d} m^2 \lambda_d(m^2)) & d > 1 \end{cases}$$

As before, in order to prove lower bounds for a non-leveled circuit, we just level it. We can do that by increasing it's size by a factor of d. We can then use the lower bounds for leveled circuits. We hence have the following corollary.

Corollary 5.2 *Any Boolean circuit with arbitrary gates, C, of depth d, for the product of two $m \times m$ Boolean matrices over $GF(2)$, is of size:*

$$size(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega(\frac{1}{d^2} m^2 \lambda_d(m^2)) & d > 1 \end{cases}$$

Our proof for these lower bounds is quite general and can be applied for many other functions. In particular, the lower bound applies also for the following functions:

1. The product of two $m \times m$ matrices over $GF(p)$, where each element of the field is represented by $\lceil \log p \rceil$ bits.
2. All the functions considered in [Pud94], e.g., the parallel prefix linear transformation over $GF(2)$.

As mention above, we define a similar model of circuits with arbitrary gates over any finite field. We prove similar lower bounds for this model for any finite field. In particular, for the field $GF(p)$, we prove a lower bound for the product of two $m \times m$ matrices, where the inputs take values in the field.

Let us start with a short sketch of the proof of Theorem 5.1. As before, the proof is based on Lemma 3.13. In all that comes below we use the notations of Subsection 3.5. We will apply Lemma 3.13 on the circuit C. By Lemma 3.13, there is a set of columns, $D \subset [m]$, and small sets of vertices I, O, V, such that, I is a set of inputs, O is a set of outputs, V is a set of intermediate gates, and such that,

$$P_G[\ I_G \setminus I\ ,\ O_D \setminus O\ ,\ V\ ] = 0$$

(where O_D is the set of outputs corresponding to the set of columns D). The values of the outputs in $O_D \setminus O$ are hence determined by the outputs of the gates in V and by the values of the inputs in I. Therefore, for any fixed assignment for the inputs in I, the total number of possible values that the outputs in $O_D \setminus O$ can take is at most $2^{|V|}$ (which is the total number of values that the gates in V can output). Since V is a small set, $2^{|V|}$ is a relatively small number and we conclude that for any fixed assignment for the inputs in I the outputs in $O_D \setminus O$ can only get a small number of values. We will derive a contradiction by finding a fixed assignment for the inputs in I, such that the outputs in $O_D \setminus O$ can get many values.

Let us describe our assignment for I. First, we fix Y to be a matrix in which any minor of size $\frac{m}{2} \times |D|$ is of high rank. Therefore, there will be many vectors in the image of any such minor. After fixing Y, we set to zero all the inputs in I that come from the matrix X. That is, we allow X to be any matrix in which this certain set of entries is zero (the entries that appear in I). Since I is a relatively small set, there are many such matrices. More accurately, the set of all such matrices form a linear subspace of dimension $\geq m^2 - |I|$. The last step will be to show that after fixing Y as above, and after fixing to 0 all the inputs in I that come from X, there are still many possibilities for the product $X \cdot Y$. In particular, we will get that the number of possible values for the outputs in $O_D \setminus O$ is larger than $2^{|V|}$.

Thus, the first step is to show that there is a matrix Y, in which any minor of size $\frac{m}{2} \times |D|$ is of high rank. Note that it is not hard to prove that there exists a matrix Y, in which any such minor is of maximal rank. For our proof, however, it will be enough to have the weaker requirement that any such minor is of high rank.

Definition 5.3 *Let $Y = (y_{i,j})$ be an $m \times m$ matrix. For sets $\alpha, \beta \subset [m]$, denote*

$$(Y)_{\alpha,\beta} = (y_{i,j}) \textit{ such that } i \in \alpha,\ j \in \beta\ ,$$

i.e., $(Y)_{\alpha,\beta}$ is the minor of Y with the set of rows α and the set of columns β.

Claim 5.4 *For any $l = o(m)$, there exists an $m \times m$ matrix Y (over $GF(2)$), such that for any $\alpha, \beta \subset [m]$, with $|\alpha| = \lceil \frac{m}{2} \rceil$ and $|\beta| = l$,*

$$rank((Y)_{\alpha,\beta}) \geq \frac{l}{2}\ .$$

Proof:
We assume for convenience that m, l are even numbers. We will show that a random matrix Y satisfies the requirement of the lemma (with high probability). Let us first calculate the probability that a certain minor of size $\frac{m}{2} \times l$ is of rank $\leq \frac{l}{2}$. As in the proof of Lemma 4.9, the number of $\frac{m}{2} \times l$ matrices of rank $\leq \frac{l}{2}$ is at most

$$\binom{l}{\frac{l}{2}} \cdot 2^{\frac{m}{2} \cdot \frac{l}{2}} \cdot 2^{\frac{l}{2} \cdot (l - \frac{l}{2})} < 2^{\frac{1}{4}ml + \frac{1}{4}l^2 + l}\ .$$

Therefore, the probability that a certain minor of size $\frac{m}{2} \times l$ is of rank $\leq \frac{l}{2}$ is at most

$$2^{\frac{1}{4}ml + \frac{1}{4}l^2 + l} \cdot 2^{-\frac{m}{2}l} = 2^{-(\frac{1}{4}ml - \frac{1}{4}l^2 - l)}\ .$$

Hence, if Y is a random matrix the probability that some $\frac{m}{2} \times l$ minor is of rank $\leq \frac{l}{2}$ is at most

$$\binom{m}{\frac{m}{2}} \cdot \binom{m}{l} \cdot 2^{-(\frac{1}{4}ml - \frac{1}{4}l^2 - l)} \leq 2^{-(\frac{1}{4}ml - \frac{1}{4}l^2 - l - 2m)} < 1 .$$

Consequently, there is an $m \times m$ matrix Y, in which every $\frac{m}{2} \times l$ minor is of rank $> \frac{l}{2}$. □

For a set of coordinates $\beta \subset [m]$, and an m-vector v, denote by v_β the restriction of v to β, i.e.,

$$v_\beta = (v_i)_{i \in \beta} .$$

For a set of coordinates $\alpha \subset [m]$, denote by $\mathcal{V}_\alpha$ the subspace of all vectors that have the value 0 in all the coordinates outside α, i.e.,

$$\mathcal{V}_\alpha = \{ \ v \in \{0,1\}^m \mid \forall i \in [m] \setminus \alpha \ , \ v_i = 0 \ \} .$$

We will now prove that if Y is the matrix guaranteed by Claim 5.4, for $l = |\beta|$, then for every vector space $\mathcal{V}_\alpha$ of high dimension, there are many different vectors in the set

$$\{ \ (v \cdot Y)_\beta \mid v \in \mathcal{V}_\alpha \ \}.$$

Claim 5.5 *Let $\beta \subset [m]$ be such that $|\beta| = o(m)$. Let $\alpha \subset [m]$ be such that $|\alpha| = \lceil \frac{m}{2} \rceil$. Let Y be an $m \times m$ matrix that satisfies the requirement of Claim 5.4 for $l = |\beta|$. Then, the number of different vectors of the form $(v \cdot Y)_\beta$, where $v \in \mathcal{V}_\alpha$, is at least $2^{\frac{|\beta|}{2}}$.*

Proof:
Since Y satisfies the requirement of Claim 5.4 for $l = |\beta|$, the minor $(Y)_{\alpha,\beta}$ is of rank at least $|\beta|/2$. Hence, the image of this minor is of dimension at least $|\beta|/2$. The image of this minor is just the set of all vectors of the form $(v \cdot Y)_\beta$, where $v \in \mathcal{V}_\alpha$. Hence, this set is of size at least $2^{\frac{|\beta|}{2}}$. □

We are now ready to give the proof of Theorem 5.1.

Proof of Theorem 5.1:
For $d = 1$ the proof is trivial by the observation that every output column depends on all the variables in X. For larger depth, assume for a contradiction that $size(C) < (1/1000d) \cdot m^2 \lambda_d(m^2)$. Let G be the graph of the circuit. By Lemma 3.13 there exist sets $V \subset V_G$, $D \subset [m]$, $O \subset O_D$, $I \subset I_G$, such that:

1. $m \leq |V| = k = o(m^2)$.
2. $|D| = \left\lceil \frac{10k}{m} \right\rceil = o(m)$.
3. For every $i \in D$, $|O_i \cap O| < \frac{1}{10}m$.
4. $|I| < \frac{1}{10}m^2$.
5. $P_G[\ I_G \setminus I \ , \ O_D \setminus O \ , \ V \ ] = 0$.

Since $P_G[\ I_G \setminus I\ ,\ O_D \setminus O\ ,\ V\ ] = 0$, the values of the outputs in $O_D \setminus O$ are determined by the outputs of the gates in V and by the inputs in I.

Fix Y to be a matrix in which every $\lceil \frac{m}{2} \rceil \times |D|$ minor is of rank higher than $\frac{|D|}{2}$ (by Claim 5.4, there exists such a matrix). Denote by I_X the set of inputs in I that are variables of X. Obviously,

$$|I_X| \leq |I| < \frac{1}{10}m^2 \ .$$

Fix all the input variables in I_X to be 0. Since all the inputs in I are now fixed, the values of the outputs in $O_D \setminus O$ are determined by the outputs of the gates in V. Since there are at most 2^k possible values for the outputs of the gates in V, we conclude that after fixing Y and I_X as above, the outputs in $O_D \setminus O$ can get at most 2^k different values.

On the other hand, we only fixed less than $\frac{1}{10}m^2$ of the entries of X. Therefore, the number of rows of X, in which we fixed at most $\frac{m}{2}$ entries, is at least $\frac{8}{10}m$. For each one of these rows, we can apply Claim 5.5 (with $\beta = D$) and conclude that the outputs in the corresponding row in O_D can get at least $2^{\frac{|D|}{2}}$ possible values. Since the value of each of these $\frac{8}{10}m$ rows of X is independent of the values of the other rows, we conclude that the total number of different values that the outputs in O_D can get is at least

$$2^{\frac{|D|}{2} \cdot \frac{8}{10}m} = 2^{\frac{4}{10} \cdot |D| \cdot m} \ .$$

Since for every $i \in D$, $|O_i \cap O| < \frac{1}{10}m$, we get that

$$|O| < \frac{1}{10} \cdot |D| \cdot m \ .$$

Hence, the outputs in O can get at most

$$2^{\frac{1}{10} \cdot |D| \cdot m}$$

different values. Therefore, the outputs in $O_D \setminus O$ can get at least

$$2^{\frac{4}{10} \cdot |D| \cdot m} \ / \ 2^{\frac{1}{10} \cdot |D| \cdot m} = 2^{\frac{3}{10} \cdot |D| \cdot m} \geq 2^{3k}$$

different values (a contradiction). □

As mentioned above, we can also obtain similar lower bounds for circuits with arbitrary gates over any finite field. A circuit with arbitrary gates over a finite field $GF(p)$ is defined similarly to a circuit with arbitrary gates over $GF(2)$. The only difference is that the inputs take values in $GF(p)$ and every gate of in-degree k is labeled with an arbitrary function from $GF(p)^k$ to $GF(p)$. Note that in particular this model includes the model of arithmetic circuits over $GF(p)$.

Theorem 5.6 *Any circuit with arbitrary gates over $GF(p)$, C, of depth d, for the product of two $m \times m$ matrices over $GF(p)$, is of size:*

$$size(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega(\frac{1}{d^2}m^2\lambda_d(m^2)) & d > 1 \end{cases}$$

The proof is similar to the proof of Theorem 5.1 with some minor modifications: We prove a version of Claim 5.4 to get a matrix with the same properties over $GF(p)$. Then we prove a version of Claim 5.5 for $GF(p)$. (Both proofs are similar to the original proofs). We then repeat the proof of Theorem 5.1 to get Theorem 5.6 (we have to make some minor changes, e.g., the outputs of the gates in V can get $p^{|V|}$ different values (rather than $2^{|V|}$) etc').

We can also prove similar lower bounds for Boolean circuits with arbitrary gates for the product of two $m \times m$ matrices over the field $GF(p)$, where each element of the field is represented by $\lceil \log p \rceil$ bits (note that the input to the circuit is of size $2m^2 \lceil \log p \rceil$). One way of proving this lower bound is by a reduction to Theorem 5.6. Just observe that $\{0,1\} \subset GF(p)$, and hence any Boolean circuit (with arbitrary gates) can be viewed as a circuit with arbitrary gates over $GF(p)$. Another way of proving this lower bound is by proving a version of Lemma 3.13 with different parameters (because of the $\lceil \log p \rceil$ factor), and then we can repeat the proof of Theorem 5.1 with minor modifications.

Theorem 5.7 *Any Boolean circuit with arbitrary gates, C, of depth d, for the product of two $m \times m$ matrices over $GF(p)$, is of size:*

$$size(C) = \begin{cases} \Omega(m^3) & d = 1 \\ \Omega(\frac{1}{d^2} m^2 \lambda_d(m^2)) & d > 1 \end{cases}$$

Acknowledgment

We would like to thank Toni Pitassi and Avi Wigderson for helpful conversation.

References

[Ajt83] Miklós Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[Bla99] Markus Blaser. A - Lower Bound for the Rank of - Matrix Multiplication over Arbitrary Fields. FOCS 1999: 45–50.

[Bsh89] N.H. Bshouty. A lower bound for matrix multiplication. *SIAM Journal of Computing*, 18:759–765, 1989.

[BS82] W. Baur, V. Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22:317–330, 1982.

[DDPW83] D. Dolev, C. Dwork, N. J. Pippenger, A. Wigderson. Superconcentrators, generalizer and generalized connectors. STOC 1983: 42–51.

[FSS81] M.L. Furst, J.B. Saxe, M. Sipser. Parity, circuits, and the polynomial-time hierarchy. FOCS 1981: 260–270.

[Gat88] J. von zur Gathen. Algebraic Complexity Theory. *Ann. Rev. Computer Science*, 3:317–347, 1988.

[GK98] Dima Grigoriev, Marek Karpinski. An Exponential Lower Bound for Depth 3 Arithmetic Circuits. STOC 1998: 577-582.

[GR98] Dima Grigoriev, Alexander A. Razborov. Exponential Complexity Lower Bounds for Depth 3 Arithmetic Circuits in Algebras of Functions Over Finite Fields. FOCS 1998: 269–278.

[Hås86] Johan Håstad. Almost optimal lower bounds for small depth circuits. STOC 1986: 6–20.

[HMPST87] A.Hajnal, W.Maass, P.Pudlak, M.Szegedy, G.Turan. Threshold circuits of bounded depth. FOCS 1987: 99–110.

[IPS97] R.Impagliazzo, R.Paturi, M.Saks. Size-depth trade-offs for threshold circuits. *SIAM Journal of Computing*, 26:693–707, 1997.

[Pud94] P. Pudlak. Communication in bounded depth circuits. *Combinatorica* 14(2):203–216, 1994.

[Razb87] A. A. Razborov. Lower bounds for the size of circuits with bounded depth with basis $\{\wedge, \oplus\}$. *Mat. Zametki*, 1987.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. STOC 1987: 77–82.

[Str69] V. Strassen. Gaussian elimination is not optimal. *Numer. Math*, 13:354–356, 1969.

[Str73] V. Strassen. Die berechnungskomplexitat vo elementarsymmetrischen funktionen und von interpolationskoefizienten. *Numer. Math*, 20:238–251, 1973.

[SW99] Amir Shpilka, Avi Wigderson. Depth-3 Arithmetic Formulae over Fields of Characteristic Zero. Electronic Colloquium on Computational Complexity (ECCC) 6(023), 1999.

[Val77] Leslie G. Valiant. Graph-theoretic arguments in low-level complexity. *Lecture notes in Computer Science*, 53:162–176.

[Yao85] A. C. Yao. Separating the polynomial hierarchy by oracles. FOCS 1985: 1–10.