Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 32 (2020)

On Computing Multilinear Polynomials Using Multi-r-ic Depth Four Circuits

Suryajith Chillara
CRI, University of Haifa, Israel.
suryajith@cmi.ac.in

Abstract

In this paper, we are interested in understanding the complexity of computing multilinear polynomials using depth four circuits in which the polynomial computed at every node has a bound on the individual degree of $r \geqslant 1$ with respect to all its variables (referred to as multi-r-ic circuits). The goal of this study is to make progress towards proving superpolynomial lower bounds for general depth four circuits computing multilinear polynomials, by proving better bounds as the value of r increases.

Recently, Kayal, Saha and Tavenas (Theory of Computing, 2018) showed that any depth four arithmetic circuit of bounded individual degree r computing an explicit multilinear polynomial on $n^{O(1)}$ variables and degree $d = o(n)$, must have size at least $\left(\frac{n}{r^{1.1}}\right)^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$ when r is $o(d)$ and is strictly less than $n^{1.1}$. This bound however deteriorates as the value of r increases. It is a natural question to ask if we can prove a bound that does not deteriorate as the value of r increases, or a bound that holds for a *larger* regime of r.

In this paper, we prove a lower bound which does not deteriorate with increasing value of r, albeit for a specific instance of $d = d(n)$ but for a *wider* range of r. Formally, for all large enough integers n and a small constant η, we show that there exists an explicit polynomial on $n^{O(1)}$ variables and degree $\Theta(\log^2 n)$ such that any depth four circuit of bounded individual degree $r \leqslant n^{\eta}$ must have size at least $\exp\left(\Omega\left(\log^2 n\right)\right)$. This *improvement* is obtained by suitably adapting the complexity measure of Kayal et al. (Theory of Computing, 2018). This adaptation of the measure is inspired by the complexity measure used by Kayal et al. (SIAM J. Computing, 2017).

1 Introduction

One of the major focal points in the area of algebraic complexity theory is to show that certain polynomials are hard to compute syntactically. Here, the hardness of computation is quantified by the number of algebraic operations that are needed to compute the target polynomial. Instead of the standard Turing machine model, we consider arithmetic circuits and formulas as models of computation for polynomials.

Arithmetic circuits are directed acyclic graphs such that the leaf nodes are labeled by variables or constants from the underlying field, and every non-leaf node is labeled either by a $+$ or $\times$. Every node computes a polynomial by operating on its inputs with the operand given by its label. The flow of computation flows from the leaf to the output node. We refer the readers to the standard resources [SY10, Sap19] for more information on arithmetic formulas and arithmetic circuits.

Valiant conjectured that the permanent polynomial does not have polynomial sized arithmetic circuits [Val79]. Working towards that conjecture, we aim to prove superpolynomial circuit size lower bounds. However, the best known circuit size lower bound is $\Omega(n \log n)$, for a power symmetric polynomial, due to Baur and Strassen [Str73, BS83], and, the best known formula size lower bound is

ISSN 1433-8092

$\Omega(n^2)$, due to Kalorkoti [Kal85]. Owing to the slow progress towards proving general circuit/formula lower bounds, it is natural to study some restricted classes of arithmetic circuits and formulas.

Since most of the polynomials of interest such as determinant, permanent, etc., are multilinear polynomials, it is natural to consider the restriction where every intermediate computation is in fact multilinear. Due to the phenomenal work in the last two decades [NW97, Raz06, Raz04, RY08, RY09, HY11, RSY08, AKV18, CLS19, CELS18, CLS18], the complexity of multilinear formulas and circuits is better understood than that of general formulas and circuits.

Backed by this progress it is natural to try to extend these results to a circuit model where the individual degree with every variable in the polynomial computed at every node in the circuit is at most r. We refer to these circuits as multi-r-ic circuits. When $r = 1$, the circuit model is multilinear.

Kayal and Saha [KS17a] first studied multi-r-ic circuits of depth three and proved exponential lower bounds. Kayal, Saha and Tavenas [KST18] have extended this and proved exponential lower bounds at depth three and depth four. These circuits that were considered were syntactically multi-r-ic . That is, at every product node, every variable appears in the support of at most r many operands, and the sum total of the individual degrees over all the operands is also at most r. Henceforth, all the multi-r-ic depth four circuits that we talk about shall be syntactically multi-r-ic .

Recently, Kumar, Oliviera and Saptharishi [KdOS19] showed that there is a chasm[1] for multi-r-ic circuits too. Formally, they showed that any polynomial sized (say n^c for a fixed constant c) multi-r-ic circuit of arbitrary depth computing a polynomial on n variables can be depth reduced to a syntactical multi-r-ic depth four circuits of size $\exp(O(\sqrt{rn \log n}))$. This provides us a motivation to study multi-r-ic depth four circuits and prove strong lower bounds against them.

Kayal, Saha and Tavenas [KST18] proved an exponential size lower bound against multi-r-ic depth four circuits computing the iterated matrix multiplication polynomial. They achieved this bound using a measure that is inspired by the method *Shifted Partial Derivatives* [Kay12, GKKS14] and the method of *Skew Partial Derivatives* [KNS16]. They referred to this new technique as the method of *Shifted Skew Partial Derivatives*. Hegde and Saha [HS17] improved upon [KST18] and showed a *near-optimal* size lower bound. However, the *best known* lower bounds are for polynomials that are not multilinear but multi-r-ic.

Motivation for this work

Raz and Yehudayoff [RY09] showed a lower bound of $\exp(\Omega\left(\sqrt{d \log d}\right))$ against multilinear depth four circuits which compute a multilinear polynomial over n variables and degree $d \ll n$ (cf. [KST18, Footnote 9]). Kayal, Saha and Tavenas [KST18] have shown a lower bound of $\left(\frac{n}{r^{1.1}}\right)^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$ for a multilinear polynomial over $n^{O(1)}$ variables and degree d that is computed by a multi-r-ic depth four circuit. This lower bound deteriorates as the value of r increases. Further, it is superpolynomial only when r is $o(d)$ and is strictly less than $n^{1.1}$. This raises a natural question if the dependence on r could be improved upon.

In this work, we show that for a certain regime of d, we can prove a lower bound that does not deteriorate as the value of r increases.

[1] Agrawal and Vinay [AV08], Koiran [Koi12], and Tavenas [Tav15] showed that any general circuit can be depth reduced to a depth four circuit of non-trivial size.

Theorem 1 (Main Theorem)**.** *Let n be a large enough integer. There exist a constant $\eta \in (0,1)$ and an explicit $n^{O(1)}$-variate, degree $\Theta(\log^2 n)$ multilinear polynomial Q_n such that for all $r \leqslant n^\eta$, any syntactically multi-r-ic depth four circuit computing Q_n must have size* $\exp\left(\Omega(\log^2 n)\right)$.

By substituting for $d = \Theta(\log^2 n)$ into the bound from [KST18], we get that their bound evaluates to $n^{\Omega\left(\frac{\log n}{\sqrt{r}}\right)}$. Note that this bound is superpolynomial only when $r = o(\log^2 n)$. Thus our lower bound is quantitatively better in this regime of parameters. Further, we show a lower bound in the regime of parameters where $r \geqslant d$, for which Kayal, Saha and Tavenas [KST18] do not.

If we can show superpolynomial size lower bounds against multi-r-ic depth four circuits for $r = n^c$ for any constant c, then we indeed have superpolynomial circuit size lower bounds against depth four circuits. We believe that by building on the work of [KST18, HS17], Theorem 1 is a step towards that direction.

The explicit polynomial that we consider can be expressed as a p-projection of Iterated Matrix Multiplication polynomial $\mathsf{IMM}_{\tilde{n},\tilde{d}}$ (where $\tilde{n} = n^{O(1)}$ and $\tilde{d} = \Theta(\log^2 n)$) and thus Theorem 1 implies a lower bound of $n^{\Omega(\log n)}$ for Iterated Matrix Multiplication polynomial as well.

Corollary 2 (Informal)**.** *Let n and d be integers such that $d = \Theta(\log^2 n)$. There exists a constant $\eta \in (0,1)$ such that for all $r \leqslant n^\eta$, any syntactically multi-r-ic depth four circuit computing Iterated Matrix Multiplication polynomial ($\mathsf{IMM}_{n,d}$) must have size at least* $\exp\left(\Omega(\log^2 n)\right)$.

Since Iterated Matrix Multiplication polynomial can be expressed as a p-projection of determinant polynomial [Sap19, Theorem 3.6], we get a similar lower bound for the determinant polynomial too.

Corollary 3 (Informal)**.** *Let N be a large integer. There exists a constant $\eta \in (0,1)$ such that for all $r \leqslant N^\eta$ any syntactically multi-r-ic depth four circuit computing the determinant polynomial over $N \times N$ matrix must have size at least* $\exp\left(\tilde{\Omega}(\log^2 N)\right)$.

Proof overview:

A depth four circuit computes polynomials that can be expressed as sums of products of polynomials. Analogous to the work of Fournier et al. [FLMS15], and Kumar and Saraf [KS17b], we first consider multi-r-ic depth four circuits of low bottom support[2] and prove lower bounds against them.

Let $T_1, T_2, \ldots, T_s$ be the terms corresponding to the product gates feeding into the output sum gate. The output polynomial is the sum of terms $T_1, T_2, \ldots, T_s$. Note that each of these T_i's is a product polynomials $Q_{i,j}$ such that every monomial in these $Q_{i,j}$'s depends on a small set of variables (say μ many). One major observation at this point is to see that there can at most be $N \cdot r$ many factors in any of the T_i's.

Kayal et al. [KST18] observed that the measure of shifted partial derivates [KSS14, FLMS15] does not yield any non-trivial lower bound if the number of factors is much larger than the number of variables itself. They worked around this obstacle by defining a *hybrid* complexity measure (refered to as *Shifted Skew Partial Derivatives*) where they first split all the variables into two disjoint sets Y and Z such that $|Y| \gg |Z|$. They then considered some *low* order partial derivatives with respect to monomials in $\mathbb{F}[Y]$

[2]That is, all the product gates at the bottom are supported on small set of variables.

and subsequently set all the variables from Y to zero in the partial derivatives obtained. This effectively reduces the number of factors in any summand in a partial derivatives of T to at most $|Z| \cdot r$. They then *shift* these polynomials by monomials in variables from Z and look at the dimension of the F-linear span of the polynomials thus obtained.

This measure gave them a size lower bound of $\left(\frac{n}{r^{1.1}}\right)^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$ against multi-r-ic depth four circuits computing an explicit polynomial on $n^{O(1)}$ variables and degree $d = o(n)$ when $r = o(d)$. To improve the dependence on r in the lower bound, we consider a variant of *Shifted Skew Partial Derivatives* that we call *Projected Shifted Skew Partial Derivatives*. Here, we project down the space of Shifted Skew Partials and only look at the multilinear terms. Since the polynomial of interest is multilinear, it makes sense to only look at the multilinear terms obtained after the shifts of the skew partial derivatives. This is analogous to the method employed by Kayal et al. [KLSS17] to prove exponential size lower bounds for homogeneous depth four circuits, through the measure of *Projected Shifted Partial Derivatives*.

We first show that the dimension of Projected Shifted Skew Partial derivatives is not too large for small multi-r-ic depth four circuits of low bottom support. We then show that there exists an explicit polynomial whose dimension of Projected Shifted Skew Partial derivatives is large and thus cannot be computed by small multi-r-ic depth four circuits. We then lift this result to multi-r-ic depth four circuits for a suitable set of parameters.

2 Preliminaries

Notation:

- For a polynomial $f \in \mathbb{F}[Y \sqcup Z]$, we use $\partial_Y^{=k}(f)$ to refer to the space of partial derivatives of order k of f with respect to monomials of degree k in Y.
- We use $\mathbf{z}^{=\ell}$ and $\mathbf{z}^{\leqslant \ell}$ to refer to the set of all the monomials of degree equal to ℓ and at most ℓ, respectively, in variables from Z.
- We use $\mathbf{z}_{\text{ML}}^{\leqslant \ell}$ to refer to the set of all the multilinear monomials of degree at most ℓ in variables from Z.
- We use $\mathbf{z}_{\text{NonML}}^{\leqslant \ell}$ to refer to the set of all the non-multilinear monomials of degree at most ℓ in variables from Z.
- For sets A and B of polynomials, we define the product $A \cdot B$ to be the set $\{f \cdot g \mid f \in A \text{ and } g \in B\}$.
- For a polynomial f, $\text{vars}(f)$ is the set of variables that the polynomial f depends on.
- For a gate u in a circuit, we use f_u to denote the polynomial computed at gate u.
- For a polynomial f in $\mathbb{F}[Y \sqcup Z]$, we define Z-support of f to be equal to $\text{vars}(f) \cap Z$ and Z-support size of f to be equal to $|\text{vars}(f) \cap Z|$.

Definition 4 (Depth four circuits)**.** *A depth four circuit (denoted by $\Sigma\Pi\Sigma\Pi$) over a field $\mathbb{F}$ and variables $\{x_1, x_2, \ldots, x_n\}$ computes polynomials which can be expressed in the form of sums of products of polynomials.*

That is, $\sum_{i=1}^{s}\prod_{j=1}^{d_i} Q_{i,j}(x_1,\ldots,x_n)$ *for some* d_i*'s. A depth four circuit is said to have a bottom support of* t *(denoted by* $\Sigma\Pi\Sigma\Pi^{\{t\}}$*) if it is a depth four circuit and all the monomials in every polynomial* $Q_{i,j}$ *(*$j \in [d_i], i \in [s]$*) depend on at most* t *variables.*

Definition 5 (multi-**r**-ic circuits)**.** *Let* $\mathbf{r} = (r_1, r_2, \cdots, r_n)$*. An arithmetic circuit* C *is said to be a syntactically multi-***r***-ic circuit if*

- *for all gates* $v \in C$ *and* $i \in [n]$*,* $\deg_{x_i}(f_v) \leqslant r_i$*,*
- *for all product gates* $u \in C$ *such that* $u = u_1 \times u_2 \times \cdots \times u_t$*, each variable* x_i *can appear in at most* r_i *many of the* u_i*'s (*$i \in [t]$*) and the total formal degree with respect to every variable* x_i *(*$i \in [n]$*) over the polynomials computed at* $u_1, u_2, \cdots, u_t$*, is bounded by* r_i*, i.e.* $\sum_{j\in[t]} \deg_{x_i}(f_{u_j}) \leqslant r_i$ *for all* $i \in [n]$.

If $\mathbf{r} = (r, r, \cdots, r)$*, then we simply refer to them as multi-*r*-ic circuits.*

Complexity Measure: We shall now describe our complexity measure which we shall henceforth refer to as Dimension of Projected Shifted Skew Partial Derivatives. This is a natural extension of the Dimension of Shifted Skew Partial Derivatives as used by [KST18].

This formulation is analogous to the work of [KLSS14] where they study a *shifted partials inspired* measure called *Shifted Projected Partial derivatives* and then [KLSS17] where they study *Projected Shifted Partial derivatives*.

Since the polynomial of interest is multilinear, it does make sense for us to only look at those shifts of the partial derivatives that maintain multilinearity. At the same time, since the individual degree of the intermediate computations in the multi-r-ic depth four circuit is large and non-multilinear terms *cancel* out to generate the multilinear polynomial, we can focus on the multilinear terms generated after the shifts by projecting our linear space of polynomials down to them. We describe this process formally, below.

Let the variable set X be partitioned into two fixed, disjoint sets Y and Z such that $|Y|$ is much larger than $|Z|$, $|Y| \gg |Z|$. Let $\sigma_Y : \mathbb{F}[Y \sqcup Z] \mapsto \mathbb{F}[Z]$ be a linear map such that for any polynomial $f(Y, Z)$, $\sigma_Y(f) \in \mathbb{F}[Z]$ is obtained by setting every variable from Y to zero and leaving the variables from Z untouched. Let $\mathrm{mult} : \mathbb{F}[Z] \mapsto \mathbb{F}[Z]$ be a linear map such that for any polynomial $g(Z)$, $\mathrm{mult}(g) \in \mathbb{F}[Z]$ is obtained by setting the coeficients of all the non-multilinear monomials in g to 0 and leaving the rest untouched.

Recall that we use $\partial_Y^{=k} f$ to denote the set of all partial derivatives of f of order k with respect to degree k monomials over variables just from Y, and $\mathbf{z}^{\leqslant \ell} \cdot \sigma_Y(\partial_Y^{=k} f)$ to refer to the set of polynomials obtained by multiplying each polynomial in $\sigma_Y(\partial_Y^{=k} f)$ with monomials of degree at most ℓ in Z variables. We will now define our complexity measure, Dimension of Projected Shifted Skew Partial Derivatives with respect to parameters k and ℓ (denoted by $\Gamma_{k,\ell}$) as follows.

$$\Gamma_{k,\ell}(f(Y,Z)) = \dim\left(\mathbb{F}\text{-span}\left\{\mathrm{mult}\left(\mathbf{z}^{\leqslant \ell} \cdot \sigma_Y\left(\partial_Y^{=k} f\right)\right)\right\}\right)$$

This is a natural generalization of Shifted Skew Partial Derivatives measure defined by Kayal, Saha and Tavenas [KST18]. The following proposition is easy to verify.

Proposition 6 (Sub-additivity)**.** *Let k and ℓ be integers. Let the polynomials f, f_1, f_2 be such that $f = f_1 + f_2$. Then, $\Gamma_{k,\ell}(f) \leqslant \Gamma_{k,\ell}(f_1) + \Gamma_{k,\ell}(f_2)$.*

Monomial Distance: We recall the following definition of distance between monomials from [CM19].

Definition 7 (Definition 2.7, [CM19])**.** *Let M_1, M_2 be two monomials over a set of variables. Let S_1 and S_2 be the multisets of variables corresponding to the monomials M_1 and M_2 respectively. The distance $\mathrm{dist}(M_1, M_2)$ between the monomials M_1 and M_2 is the $\min\{|S_1| - |S_1 \cap S_2|, |S_2| - |S_1 \cap S_2|\}$ where the cardinalities are the order of the multisets.*

For example, let $M_1 = x_1^2 x_2 x_3^2 x_4$ and $M_2 = x_1 x_2^2 x_3 x_5 x_6$. Then $S_1 = \{x_1, x_1, x_2, x_3, x_3, x_4\}$, $S_2 = \{x_1, x_2, x_2, x_3, x_5, x_6\}$, $|S_1| = 6$, $|S_2| = 6$ and $\mathrm{dist}(M_1, M_2) = 3$. It is important to note that two distinct monomials could have distance 0 between them if one of them is a multiple of the other and hence the triangle inequality does not hold.

For two vectors $\mathbf{a}, \mathbf{b}$, we use $\mathrm{HammingDist}(\mathbf{a}, \mathbf{b})$ to refer to the Hamming distance between these vectors $\mathbf{a}$ and $\mathbf{b}$.

The following beautiful lemma (from [GKKS14]) is key to the asymptotic estimates required for the lower bound analyses.

Lemma 8 (Lemma 6, [GKKS14])**.** *Let $a(n), f(n), g(n) : \mathbb{Z}_{\geqslant 0} \to \mathbb{Z}_{\geqslant 0}$ be integer valued functions such that $(f + g) = o(a)$. Then,*

$$\ln \frac{(a+f)!}{(a-g)!} = (f+g) \ln a \pm O\left(\frac{(f+g)^2}{a}\right)$$

We need the following strengthening of the Principle of Inclusion and Exclusion, due to Kumar and Saraf [KS17b].

Lemma 9 (Strong Inclusion-Exclusion, Lemma 3.8 [KS17b])**.** *Let $W_1, W_2, \cdots, W_t$ be subsets of a finite set W. For a parameter $\lambda \geqslant 1$, let $\sum_{\substack{i,j \in [t] \\ i \neq j}} \left|W_i \cap W_j\right| \leqslant \lambda \sum_{i \in [t]} |W_i|$. Then, $\left|\cup_{i \in [t]} W_i\right| \geqslant \frac{1}{4\lambda} \sum_{i \in [t]} |W_i|$.*

3 Multi-r-ic Depth Four Circuits of Low Bottom Support

Let C be a multi-r-ic depth four circuit of size s and bottom support at most μ. For some parameters k and ℓ which we shall fix later, we shall show that $\Gamma_{k,\ell}(C)$ is not too large if multi-r-ic depth four circuit C is of small size and is of low bottom support.

3.1 Upper bound on $\Gamma_{k,\ell}(C)$

Recall that C can be expressed a sum of at most s many products of polynomials $T_1 + \cdots + T_s$ where each T_i is a syntactically multi-r-ic product of polynomials of low monomial support.

We shall first prove a bound on $\Gamma_{k,\ell}(T_i)$ for an arbitrary $i \in [s]$ and derive a bound on $\Gamma_{k,\ell}(C)$ by using sub-additivity of the measure (cf. Proposition 6).

Let T be a syntactic multi-r-ic product of polynomials $P_1(Y, Z) \cdot P_2(Y, Z) \cdot \ldots \cdot P_D(Y, Z) \cdot R(Y)$ such that all the monomials in every polynomial factor in T depend on at most μ many variables. We shall first pre-process the product T by doing the following procedure.

Preprocessing: Repeat this process until all but at most one of the factors in T (except R) have a Z-support size of at least $\frac{\mu}{2}$.

1. Pick two factors P_{i_1} and P_{i_2} from T such that $R \notin \{P_{i_1}, P_{i_2}\}$ and they have the smallest Z-support size amongst all the factors but R in T.
2. If both of them have Z-support size strictly less than $\frac{\mu}{2}$, merge these factors to obtain a new factor P'. Else, stop.
3. Update the term T by replacing the factors P_{i_1} and P_{i_2} with P'. Repeat.

In the procedure described above, it is important to note that post merging, the monomials in the product polynomial will depend on at most μ many variables from Z as the factors being merged had Z-support size strictly less than $\frac{\mu}{2}$ each. Henceforth, W.L.O.G we shall assume that every product gate at the top, in multi-r-ic depth four circuit of low bottom support, is in the processed form.

Let $T = Q_1(Y, Z) \cdot Q_2(Y, Z) \cdot \ldots \cdot Q_t(Y, Z) \cdot R(Y)$ be the product obtained after the preprocessing. All but at most one of the Q_i's have a Z-support size of at least $\frac{\mu}{2}$. The total Z-support size is at most $|Z|\, r = mr$ since T is a syntactically multi-r-ic product. Thus,

$$(t-1) \cdot \frac{\mu}{2} \leqslant mr \quad \implies \quad t \leqslant \frac{2mr}{\mu} + 1.$$

Lemma 10. *Let n, k, r, ℓ and μ be positive integers such that $\ell + k\mu < \frac{m}{2}$. Let T be a processed syntactic multi-r-ic product of polynomials $Q_1(Y, Z) \cdot Q_2(Y, Z) \cdot \ldots \cdot Q_t(Y, Z) \cdot R(Y)$ such that all monomials in each of the Q_i's ($i \in [t]$) depend on at most μ many variables from Z. Then, $\Gamma_{k,\ell}(T)$ is at most $\binom{t}{k} \cdot \binom{m}{\ell+k\mu} \cdot (\ell + k\mu)$.*

Before presenting the proof of Lemma 10, we shall first use it to show an upper bound on the dimension of the space of Projected Shifted Skew Partial derivatives of a depth four multi-r-ic circuit of low bottom support.

Lemma 11. *Let n, k, r, ℓ and μ be positive integers such that $\ell + k\mu < \frac{m}{2}$. Let C be a processed syntactic multi-r-ic depth four circuit of bottom support μ and size s. Then, $\Gamma_{k,\ell}(C)$ is at most $s \cdot \binom{\frac{2mr}{\mu}+1}{k} \cdot \binom{m}{\ell+k\mu} \cdot (\ell + k\mu)$.*

Proof. From the above discussion, we get that C can be expressed as $\sum_i^s T_i$ such that each T_i is a processed syntactically multi-r-ic product of polynomials, all of whose monomials depend on at most μ many variables from Z. From Proposition 6, we get that $\Gamma_{k,\ell}(C) \leqslant \sum_{i=1}^s \Gamma_{k,\ell}(T_i)$. From the afore mentioned discussion we know that the number of factors in T_i with non-zero Z-support size is at most $\left(\frac{2mr}{\mu} + 1\right)$. From Lemma 10, we get that $\Gamma_{k,\ell}(T_i)$ is at most $\binom{\frac{2mr}{\mu}+1}{k} \cdot \binom{m}{\ell+k\mu} \cdot (\ell + k\mu)$. By putting all of this together, we get that

$$\Gamma_{k,\ell}(C) \leqslant s \cdot \binom{\frac{2mr}{\mu}+1}{k} \cdot \binom{m}{\ell + k\mu} \cdot (\ell + k\mu)\,.$$

$\square$

We now present the proof of Lemma 10 to complete the picture.

Proof of Lemma 10. We will first show by induction on k, the following for the set of kth order partial derivatives of T with respect to degree k monomials over variables from Y.

$$\partial_Y^{=k} T \subseteq \mathbb{F}\text{-span}\left(\left\{\bigcup_{S\in\binom{[t]}{t-k}}\left\{\left(\prod_{i\in S} Q_i(Y,Z)\right)\cdot \mathbf{z}_{\mathrm{ML}}^{\leqslant k\mu}\cdot \mathbb{F}[Y]\right\}\right\}\bigcup\left\{\bigcup_{S\in\binom{[t]}{t-k}}\left\{\left(\prod_{i\in S} Q_i(Y,Z)\right)\cdot \mathbf{z}_{\mathrm{NonML}}^{\leqslant kr\mu}\cdot \mathbb{F}[Y]\right\}\right\}\right)$$

The base case of induction for $k = 0$ is trivial as T is already in the required form. Let us assume the induction hypothesis for all derivatives of order $< k$. That is, $\partial_Y^{=k-1} T$ can be expressed as a linear combination of terms of the form

$$h(Y,Z) = \left(\prod_{i\in S} Q_i(Y,Z)\right)\cdot h_1(Z)\cdot h_2(Y)$$

where S is a set of size $t-(k-1)$, $h_1(Z)$ is a *structured* polynomial in $\mathbb{F}[Z]$ of degree at most $(k-1)r\mu$, and $h_2(Y)$ is some polynomial in $\mathbb{F}[Y]$. That is, $h_1(Z)$ can be expressed as a linear combination of multilinear monomials of degree at most $(k-1)\mu$, and non-multilinear monomials of degree at most $(k-1)r\mu$ over $\mathbb{F}[Z]$.

For some $u \in [|Y|]$ and some fixed i_0 in S,

$$\frac{\partial h(Y,Z)}{\partial y_u} = \left(\sum_{j\in S}\left(\prod_{\substack{i\in S\\ i\neq j}} Q_i(Y,Z)\right)\cdot\frac{\partial Q_j(Y,Z)}{\partial y_u}\cdot h_1(Z)\cdot h_2(Y)\right) + \frac{\prod_{i\in S} Q_i}{Q_{i_0}}\cdot Q_{i_0}(Y,Z)\cdot h_1(Z)\cdot\frac{\partial h_2(Y)}{\partial y_u}$$

where the first summand on the right hand side of the above equation lies in the subspace $\mathbb{F}\text{-span}\left\{\left(\prod_{\substack{i\in S\\ i\neq j}} Q_i(Y,Z)\right)\cdot\frac{\partial Q_j(Y,Z)}{\partial y_u}\cdot h_1(Z)\cdot\mathbb{F}[Y]\ :\ j\in[S]\right\}$ and the second summand in the same equation, lies in the subspace $\mathbb{F}\text{-span}\left\{\frac{\prod_{i\in S} Q_i}{Q_{i_0}}\cdot Q_{i_0}(Y,Z)\cdot h_1(Z)\cdot\mathbb{F}[Y]\right\}$.

Note that $\frac{\partial Q_j(Y,Z)}{\partial y_u}$ and Q_{i_0} are polynomials such that every monomial in these depends on at most μ many variables from Z. These monomials can be split into two sets, those that are multilinear and those that are strictly non-multilinear, over the variables from Z.

$$\begin{aligned}\frac{\partial h(Y,Z)}{\partial y_u} \in \mathbb{F}\text{-span}&\left\{\bigcup_{T\in\binom{S}{|S|-1}}\left\{\left(\prod_{i\in T} Q_i(Y,Z)\right)\cdot\mathbf{z}_{\mathrm{ML}}^{\leqslant\mu}\cdot h_1(Z)\cdot\mathbb{F}[Y]\right\}\right\}\\ &\bigcup\mathbb{F}\text{-span}\left\{\bigcup_{T\in\binom{S}{|S|-1}}\left\{\left(\prod_{i\in T} Q_i(Y,Z)\right)\cdot\mathbf{z}_{\mathrm{NonML}}^{\leqslant r\mu}\cdot h_1(Z)\cdot\mathbb{F}[Y]\right\}\right\}\end{aligned}$$

In the above expression, the contribution from the variables in Y, to the monomials in $\frac{\partial Q_j(Y,Z)}{\partial y_u}$ and Q_{i_0} gets absorbed into $\mathbb{F}[Y]$.

Recall the fact that $h_1(Z)$ is a linear combination of multilinear monomials of degree at most $(k-1)\mu$, and non-multilinear monomials of degree at most $(k-1)r\mu$. Thus, we get that,

$$\frac{\partial h(Y,Z)}{\partial y_u} \in \mathbb{F}\text{-span}\left\{\bigcup_{T\in\binom{[t]}{t-k}}\left\{\left(\prod_{i\in T} Q_i(Y,Z)\right)\cdot \mathbf{z}_{\mathrm{ML}}^{\leqslant k\mu}\cdot \mathbb{F}[Y]\right\}\right\} \bigcup \mathbb{F}\text{-span}\left\{\bigcup_{T\in\binom{[t]}{t-k}}\left\{\left(\prod_{i\in T} Q_i(Y,Z)\right)\cdot \mathbf{z}_{\mathrm{NonML}}^{\leqslant kr\mu}\cdot \mathbb{F}[Y]\right\}\right\}.$$

From the discussion above we know that any polynomial in $\partial_Y^{=k}(T)$ can be expressed as a linear combination of polynomials of the form $\frac{\partial h}{\partial y_u}$. Further every polynomial of the form $\frac{\partial h}{\partial y_u}$ belongs to the set

$$W = \mathbb{F}\text{-span}\left\{\bigcup_{T\in\binom{[t]}{t-k}}\left\{\left(\prod_{i\in T} Q_i(Y,Z)\right)\cdot \mathbf{z}_{\mathrm{ML}}^{\leqslant k\mu}\cdot \mathbb{F}[Y]\right\}\right\} \bigcup \mathbb{F}\text{-span}\left\{\bigcup_{T\in\binom{[t]}{t-k}}\left\{\left(\prod_{i\in T} Q_i(Y,Z)\right)\cdot \mathbf{z}_{\mathrm{NonML}}^{\leqslant kr\mu}\cdot \mathbb{F}[Y]\right\}\right\}.$$

Thus, we get that $\partial_Y^{=k}T$ is a subset of W. This completes the inductive argument.

From the afore mentioned discussion, we can now derive the following expressions.

$$\sigma_Y\left(\partial_Y^{=k}T\right) \subseteq \mathbb{F}\text{-span}\left\{\bigcup_{S\in\binom{[t]}{t-k}}\left\{\left(\prod_{i\in S}\sigma_Y(Q_i)\right)\cdot \mathbf{z}_{\mathrm{ML}}^{\leqslant k\mu}\right\}\right\}\bigcup \mathbb{F}\text{-span}\left\{\bigcup_{S\in\binom{[t]}{t-k}}\left\{\left(\prod_{i\in S}\sigma_Y(Q_i)\right)\cdot \mathbf{z}_{\mathrm{NonML}}^{\leqslant kr\mu}\right\}\right\}.$$

It is easy to see that this inclusion holds under shift by monomials of degree at most ℓ over variables from Z.

$$\mathbf{z}^{\leqslant \ell}\cdot\sigma_Y\left(\partial_Y^{=k}T\right) \subseteq \mathbb{F}\text{-span}\left\{\bigcup_{S\in\binom{[t]}{t-k}}\left\{\left(\prod_{i\in S}\sigma_Y(Q_i)\right)\cdot \mathbf{z}_{\mathrm{ML}}^{\leqslant \ell+k\mu}\right\}\right\} \bigcup \mathbb{F}\text{-span}\left\{\bigcup_{S\in\binom{[t]}{t-k}}\left\{\left(\prod_{i\in S}\sigma_Y(Q_i)\right)\cdot \mathbf{z}_{\mathrm{NonML}}^{\leqslant \ell+kr\mu}\right\}\right\}.$$

By taking a multilinear projection of the elements on both sides, we get that

$$\mathbb{F}\text{-span}\left\{\mathrm{mult}\left(\mathbf{z}^{\leqslant \ell}\cdot\sigma_Y\left(\partial_Y^{=k}T\right)\right)\right\} \subseteq \mathbb{F}\text{-span}\left\{\bigcup_{S\in\binom{[t]}{t-k}}\left\{\mathrm{mult}\left(\left(\prod_{i\in S}\sigma_Y(Q_i)\right)\cdot \mathbf{z}_{\mathrm{ML}}^{\leqslant \ell+k\mu}\right)\right\}\right\}$$

$$\subseteq \mathbb{F}\text{-span}\left\{ \bigcup_{S\in\binom{[t]}{t-k}} \left\{ \left(\text{mult}\left(\prod_{i\in S}\sigma_Y(Q_i)\right)\right)\cdot \mathbf{z}_{\mathrm{ML}}^{\leqslant k\mu+\ell} \right\}\right\}.$$

Thus we get that $\dim\left(\mathbb{F}\text{-span}\left\{\text{mult}\left(\mathbf{z}^{\leqslant \ell}\cdot\sigma_Y(\partial_Y^{=k}T)\right)\right\}\right)$ is at most

$$\begin{aligned}
&\dim\left(\mathbb{F}\text{-span}\left\{ \bigcup_{S\in\binom{[t]}{t-k}} \left\{ \left(\text{mult}\left(\prod_{i\in S}\sigma_Y(Q_i)\right)\right)\cdot \mathbf{z}_{\mathrm{ML}}^{\leqslant k\mu+\ell} \right\}\right\}\right)\\
&\leqslant \dim\left(\mathbb{F}\text{-span}\left\{ \bigcup_{S\in\binom{[t]}{t-k}} \left\{ \text{mult}\left(\prod_{i\in S}\sigma_Y(Q_i)\right) \right\}\right\}\right)\cdot \dim\left(\mathbb{F}\text{-span}\left\{\mathbf{z}_{\mathrm{ML}}^{\leqslant k\mu+\ell}\right\}\right)\\
&\leqslant \binom{t}{t-k}\cdot\sum_{i=0}^{k\mu+\ell}\binom{m}{i}\\
&\leqslant \binom{t}{k}\cdot\binom{m}{\ell+k\mu}\cdot(\ell+k\mu) && \text{(Since } \ell+k\mu<m/2\text{).}
\end{aligned}$$

□

3.2 Polynomial family that is hard for multi-r-ic depth four circuits of low bottom support

Let n, α, k be positive integers and N_0 be equal to $k(n^2+2\alpha n)$. Let Y and Z be two disjoint sets of variables defined as follows. For all $i\in[k]$, let

$$\begin{aligned}
Y_i &= \left\{y_{a,b}^{(i)} \mid a,b\in[n]\right\}\\
Z_i &= \left\{z_{a,c}^{(i,1)} \mid a\in[n] \text{ and } c\in[\alpha]\right\}\bigcup\left\{z_{c+\alpha,b}^{(i,2)} \mid b\in[n] \text{ and } c\in[\alpha]\right\}.
\end{aligned}$$

Then,

$$Y=\bigcup_{i\in[k]}Y_i \qquad\text{and}\qquad Z=\bigcup_{i\in[k]}Z_i\,.$$

Let the variable set $X=\{x_1,\ldots,x_{N_0}\}$ be equal to $Y\sqcup Z$ under some suitable renaming. We define the polynomial family $f_{n,\alpha,k}(X)=f_{n,\alpha,k}(Y,Z)$ as follows (exactly as it was defined in [KST18]).

$$f_{n,\alpha,k}(Y,Z)=\prod_{i=1}^{k}g_i(Y_i,Z_i)\quad\text{where}\quad g_i(Y_i,Z_i)=\sum_{a,b\in[n]}y_{a,b}^{(i)}\prod_{c\in[\alpha]}z_{a,c}^{(i,1)}z_{c+\alpha,b}^{(i,2)}.$$

It is easy to see that $|Y|$ is n^2k and $|Z|$ is $2\alpha nk$. We shall henceforth use m to refer to $|Z|$. Thus, $N_0=|X|=|Y|+|Z|=k(n^2+2\alpha n)$. The degree of the polynomial $f_{n,\alpha,k}$ (denoted by d) is equal to $(2\alpha k+k)$.

The following lemma follows from the generalized Hamming bound [GRS19, Section 1.7].

Lemma 12. *For every $\Delta_0 < k$, there is a subset $\mathscr{P}_{\Delta_0} \subset [n]^{2k}$ of size $\frac{n^{2k-\Delta_0}}{\Delta_0\binom{2k}{\Delta_0}}$ such that for all $(\mathbf{a}, \mathbf{b}) \neq (\mathbf{a}', \mathbf{b}') \in \mathscr{P}_{\Delta_0}$,* $\text{HammingDist}((\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}')) \geqslant \Delta_0$.

Proof. There are n^{2k} elements in $[n]^{2k}$. Note that the volume of a Hamming ball of radius $\Delta_0 < k$ over vectors of length $2k$ is at most $\sum_{i=0}^{\Delta_0} \binom{2k}{i} \cdot n^i \leqslant \Delta_0\binom{2k}{\Delta_0} n^{\Delta_0}$. That is, there are at most $\Delta_0\binom{2k}{\Delta_0} n^{\Delta_0}$ many vectors $(\mathbf{a}, \mathbf{b})$ that are at most Δ_0-far from its center. Thus, there exists a packing of these Hamming balls in $[n]^{2k}$ with at least $\frac{n^{2k-\Delta_0}}{\Delta_0\binom{2k}{\Delta_0}}$ many balls. The centers of these balls are at least $2\Delta_0$ far away and thus at least Δ_0 far away, from each other. Set $\mathscr{P}_{\Delta_0}$ to be the collection of centers of these hamming balls. $\square$

Remark: Lemma 12 can be optimised in the above lemma to obtain a set $\mathscr{P}$ of size $\frac{2n^{2k-0.5\Delta_0}}{\Delta_0\binom{2k}{0.5\Delta_0}}$ by considering balls of radius 0.5Δ.

Let $\partial^k_{(\mathbf{a},\mathbf{b})} f_{n,\alpha,k} = \frac{\partial^k f_{n,\alpha,k}}{y^{(1)}_{a_1,b_1} y^{(2)}_{a_2,b_2} \cdots y^{(k)}_{a_k,b_k}}$. It is important to note that for any choice of $(\mathbf{a}, \mathbf{b}) \in [n]^{2k}$, we get that $\partial^k_{(\mathbf{a},\mathbf{b})} f_{n,\alpha,k}$ is a multilinear monomial of degree $d - k = 2\alpha k$, over just the variables from Z.

Lemma 13. *Let $(\mathbf{a}, \mathbf{b})$, $(\mathbf{a}', \mathbf{b}') \in [n]^{2k}$ be such that* $\text{HammingDist}((\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}')) \geqslant \Delta_0$. *Then* $\text{dist}\left(\partial^k_{(\mathbf{a},\mathbf{b})} f_{n,\alpha,k}, \partial^k_{(\mathbf{a}',\mathbf{b}')} f_{n,\alpha,k}\right) \geqslant \alpha\Delta_0$.

Proof. For a vector $(\mathbf{a}, \mathbf{b}) \in [n]^{2k}$, $\frac{\partial^k f_{n,\alpha,k}}{y^{(1)}_{a_1,b_1} y^{(2)}_{a_2,b_2} \cdots y^{(k)}_{a_k,b_k}} = \prod_{i=1}^k \prod_{c\in[\alpha]} z^{(i,1)}_{a_i,c} \cdot z^{(i,2)}_{c+\alpha,b_i}$. For all $i \in [k]$, let $h^{(i)}_{(\mathbf{a},\mathbf{b})} = \prod_{c\in[\alpha]} z^{(i,1)}_{a_i,c} \cdot z^{(i,2)}_{c+\alpha,b_i}$. Note that for some $i \in [k]$, if $a_i \neq a'_i$, $\text{dist}\left(h^{(i)}_{(\mathbf{a},\mathbf{b})}, h^{(i)}_{(\mathbf{a}',\mathbf{b}')}\right)$ is at least α. Similar is the case when $b_i \neq b'_i$. Thus, if $\text{HammingDist}((\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}')) \geqslant \Delta_0$, there are at least Δ_0 many locations such that either $a_i \neq a'_i$ or $b_i \neq b'_i$ and hence $\text{dist}\left(\partial^k_{(\mathbf{a},\mathbf{b})} f_{n,\alpha,k}, \partial^k_{(\mathbf{a}',\mathbf{b}')} f_{n,\alpha,k}\right) \geqslant \alpha\Delta_0$. $\square$

For any $\Delta_0 < k$, let $\mathscr{P}_{\Delta_0}$ be the set of vectors obtained from Lemma 12. Let $\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}$ be defined to be the set $\left\{\partial^k_{(\mathbf{a},\mathbf{b})} f_{n,\alpha,k} = \frac{\partial^k f_{n,\alpha,k}}{y^{(1)}_{a_1,b_1} y^{(2)}_{a_2,b_2} \cdots y^{(k)}_{a_k,b_k}} \mid (\mathbf{a}, \mathbf{b}) \in \mathscr{P}_{\Delta_0}\right\}$. By combining this with Lemma 13, we get that the pairwise distance between any two monomials in the set $\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}$ is at least $\alpha\Delta_0$. This can formally be summarized as follows.

Lemma 14. *Let Δ_0, n, α, k be integers. Let $\mathscr{P}_{\Delta_0}$ be a subset of $[n]^{2k}$ obtained from Lemma 12 such that for any $(\mathbf{a}, \mathbf{b}) \neq (\mathbf{a}', \mathbf{b}') \in \mathscr{P}_{\Delta_0}$,* $\text{HammingDist}((\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}')) \geqslant \Delta_0$. *Then $\partial^{=k}_{\mathscr{P}_{\Delta_0}}(f_{n,\alpha,k})$ is a set of monomials of degree $(d - k)$ such that for any pair of monomials $M_i \neq M_j$ in it,* $\text{dist}(M_i, M_j) \geqslant \alpha\Delta_0$.

We shall now show that the cardinality of the set $\text{mult}\left(\mathbf{z}^{=\ell} \cdot \sigma_Y\left(\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}\right)\right)$ is large enough for a suitable setting of parameters α, Δ_0 and k.

Lemma 15. *For ε and δ be some constants in $(0, 1)$. Let n be an asymptotically large integer. Let $m, k, d, \Delta_0, \alpha, \ell$ and μ be such that*

- $m = 2\alpha nk$,

- $d = 2\alpha k + k$,
- $\ell + k\mu < \frac{m}{2}$,
- $(d-k)^2 = o(m)$,
- $\alpha^2 \Delta_0^2 = o(m)$,
- $\Delta_0 = \delta k$ *and*
- $\ell = \frac{m}{2}(1-\varepsilon)$.

Then for all $\alpha \leqslant \frac{0.98 \cdot (2-\delta) \log n}{\delta \log\left(\frac{2}{1-\varepsilon}\right)}$, *we get that*

$$\left|\text{mult}\left(\mathbf{z}^{=\ell} \cdot \sigma_Y\left(\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}\right)\right)\right| \geqslant \frac{n^{(2-\delta)k}\binom{m-(d-k)}{\ell}}{4\delta k\binom{2k}{\delta k}}$$

where $\mathscr{P}_{\Delta_0}$ *is a set obtained from Lemma 12.*

Proof. Let $M_1, M_2, \ldots, M_t$ be the monomials in the set $\partial^{=k}_{\mathscr{P}_{\Delta_0}}(f_{n,\alpha,k})$, over variables from Z. From Lemma 14, we get that $\text{dist}(M_i, M_j) \geqslant \Delta = \alpha\Delta_0$ for all $i \neq j$. Further, $\sigma_Y(\partial^{=k}_{\mathscr{P}_{\Delta_0}}(f_{n,\alpha,k})) = \partial^{=k}_{\mathscr{P}_{\Delta_0}}(f_{n,\alpha,k})$.

Let $\mathscr{M}$ be the set of all mutlilinear monomials of the form $M_i \cdot M'$ over variables from Z where $i \in [t]$ and M' is a multilinear monomial of degree ℓ. It is important to note that the set $\mathscr{M}$ now corresponds to the set $\text{mult}\left(\mathbf{z}^{=\ell} \cdot \sigma_Y\left(\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}\right)\right)$.

For all $i \in [t]$, let B_i be the set of multilinear monomials of the form $M_i \cdot M'$ where M_i is a monomial from $\sigma_Y\left(\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}\right)$ and M' is a multilinear monomial of degree ℓ over variables from Z and is disjoint from M_i. From the aforementioned discussion, it follows that $|\mathscr{M}| = \left|\cup_{i=1}^{t} B_i\right|$.

For all $i \in [t]$, $\deg(M_i)$ is equal to $d-k$ (from Lemma 14). There are $\binom{m-(d-k)}{\ell}$ many monomials M' over variables from Z, that are disjoint from M_i. Thus the cardinality of the set B_i is equal to $\binom{m-(d-k)}{\ell}$.

For any $i, j \in [t]$ such that $i \neq j$, consider two monomials $\hat{M}_i = M_i \cdot M'$ and $\hat{M}_j = M_j \cdot M''$ from B_i and B_j respectively. For $\hat{M}_i$ and $\hat{M}_j$ to be identical, M' must contain variables from $M_j \setminus M_i$ and similarly M'' must contain variables from $M_i \setminus M_j$. The rest of the at most $(\ell - \Delta)$ many variables should be the same both in M' and M'' and thus in $\hat{M}_i$ and $\hat{M}_j$. The number of multilinear monomials $M \in B_i \cap B_j$, over variables from Z is at most $\binom{m-(d-k)-\Delta}{\ell-\Delta}$. Thus, for all $i, j \in [t]$ such that $i \neq j$, $\left|B_i \cap B_j\right| \leqslant \binom{m-(d-k)-\Delta}{\ell-\Delta}$. This inequality implicitly uses the fact that $(d-k)^2 = o(m)$, $\Delta^2 = o(m)$ and $\ell = \frac{m}{2}(1-\varepsilon)$.

Thus,

$$\sum_{i=1}^{t} |B_i| = t\binom{m-(d-k)}{\ell} \quad \text{and} \quad \sum_{i \neq j \in [t]} \left|B_i \cap B_j\right| \leqslant \frac{t^2}{2}\binom{m-(d-k)-\Delta}{\ell-\Delta}.$$

Let $T_1 = t\binom{m-(d-k)}{\ell}$ and $T_2 = \frac{t^2}{2}\binom{m-(d-k)-\Delta}{\ell-\Delta}$. Let $\lambda = \frac{T_2}{T_1}$. We get that $\sum_{i\neq j\in[t]} \left|B_i \cap B_j\right| \leqslant T_2 = \lambda T_1 = \lambda \sum_{i\in[t]} |B_i|$. We shall now show that $\lambda = \frac{T_2}{T_1} \geqslant 1$ for all $\alpha \leqslant \frac{0.98(2-\delta)\log n}{\delta \log\left(\frac{2}{1-\varepsilon}\right)}$. Once we prove that $\lambda \geqslant 1$, we can then invoke Lemma 9 and show that $\left|\cup_{i\in[t]} B_i\right| \geqslant T_1/4$.

By simplifying the expression for $\frac{T_2}{T_1}$, we get the following.

$$
\begin{aligned}
\frac{T_2}{T_1} &= \frac{\frac{t^2}{2}\binom{m-(d-k)-\Delta}{\ell-\Delta}}{t\binom{m-(d-k)}{\ell}} \\
&= \frac{t}{2} \cdot \frac{(m-(d-k)-\Delta)!}{(\ell-\Delta)!(m-\ell-(d-k))!} \cdot \frac{(m-\ell-(d-k))!\ell!}{(m-(d-k))!} \\
&= \frac{t}{2} \cdot \frac{(m-(d-k)-\Delta)!}{(m-(d-k))!} \cdot \frac{\ell!}{(\ell-\Delta)!} \\
&= \frac{t}{2} \cdot \frac{(m-(d-k)-\Delta)!}{m!} \cdot \frac{m!}{(m-(d-k))!} \cdot \frac{\ell!}{(\ell-\Delta)!} \\
&\approx O(1) \cdot t \cdot \frac{m^{(d-k)} \cdot \ell^{\Delta}}{m^{(d-k)+\Delta}} && \text{(Using Lemma 8)} \\
&= O(1) \cdot t \cdot \left(\frac{\ell}{m}\right)^{\Delta}.
\end{aligned}
$$

The math block above crucially uses the fact that $\Delta^2 = o(m) = o(\ell)$ and $(d-k)^2 = o(m)$ while invoking Lemma 8. The error term from invoking Lemma 8 has been absorbed by the constant 2 to give rise to $O(1)$ factor. For some suitably fixed constants δ and ε, let Δ_0 be set to δk and ℓ be set to $\frac{m}{2}(1-\varepsilon)$. Recall that for a fixed Δ_0, $t = \frac{n^{2k-\Delta_0}}{\Delta_0\binom{2k}{\Delta_0}}$ and $\Delta = \alpha\Delta_0 = \delta\alpha k$.

For the sake of contradiction, let us assume that $\frac{T_2}{T_1} < 1$. Then,

$$
\begin{aligned}
O(1) \cdot \frac{n^{2k-\Delta_0}}{\Delta_0\binom{2k}{\Delta_0}} \cdot \left(\frac{\ell}{m}\right)^{\Delta} &< 1 \\
n^{2k-\Delta_0} &< c_0 \cdot \Delta_0 \left(\frac{m}{\ell}\right)^{\Delta} \binom{2k}{\Delta_0} \\
n^{2k-\Delta_0} &< c_0 \cdot \Delta_0 \left(\frac{2}{1-\varepsilon}\right)^{\Delta} \left(\frac{2ek}{\Delta_0}\right)^{\Delta_0} \\
n^{(2-\delta)k} &< c_0 \cdot \Delta_0 \left(\frac{2}{1-\varepsilon}\right)^{\alpha\delta k} \left(\frac{2ek}{\delta k}\right)^{\delta k} \\
(2-\delta)k\log n &< \log(c_0\Delta_0) + \alpha\delta k \log\left(\frac{2}{1-\varepsilon}\right) + \delta k \log\left(\frac{2e}{\delta}\right)
\end{aligned}
$$

where c_0^{-1} is a constant hidden under the $O(1)$ in the first line of the math block. Hence,

$$
\alpha > \frac{(2-\delta)\log n - \delta \log\left(\frac{2e}{\delta}\right) - \frac{1}{k}\log(c_0 \cdot \Delta_0)}{\delta \log\left(\frac{2}{1-\varepsilon}\right)}.
$$

This contradicts our assumption on α for all asymptotically large n. Thus, we get that for all $\alpha \leqslant 0.98 \cdot \frac{(2-\delta)\log n}{\delta \log\left(\frac{2}{1-\varepsilon}\right)}$,

$$\left|\mathrm{mult}\left(\mathbf{z}^{=\ell} \cdot \sigma_Y\left(\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}\right)\right)\right| = |\mathscr{M}| \geqslant \frac{T_1}{4} = \frac{t\binom{m-(d-k)}{\ell}}{4} = \frac{n^{2k-\Delta_0}\binom{m-(d-k)}{\ell}}{4\Delta_0\binom{2k}{\Delta_0}} = \frac{n^{(2-\delta)k}\binom{m-(d-k)}{\ell}}{4\delta k\binom{2k}{\delta k}}.$$

□

Lemma 16. *Let δ and ε be any constants in $(0,1)$. Let n be an asymptotically large integer. Let m, k, d, α, ℓ and μ be such that*

- $m = 2\alpha n k$,
- $d = 2\alpha k + k$,
- $\ell + k\mu < \frac{m}{2}$,
- $(d-k)^2 = o(m)$,
- $\Delta_0 = \delta k$ *and*
- $\ell = \frac{m}{2}(1-\varepsilon)$.

Then for all $\alpha \leqslant 0.98 \cdot \frac{(2-\delta)\log n}{\delta \log\left(\frac{2}{1-\varepsilon}\right)}$ and $\varepsilon, \delta \in (0,1)$, we get

$$\Gamma_{k,\ell}(f_{n,\alpha,k}) \geqslant \frac{n^{(2-\delta)k}\binom{m-(d-k)}{\ell}}{4\delta k\binom{2k}{\delta k}}.$$

Proof. Recall that $\mathrm{mult}\left(\mathbf{z}^{=\ell} \cdot \sigma_Y\left(\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}\right)\right)$ is a set of multilinear monomials over just the variables from Z and thus,

$$\left|\mathrm{mult}\left(\mathbf{z}^{=\ell} \cdot \sigma_Y\left(\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}\right)\right)\right| \leqslant \dim\left(\mathbb{F}\text{-span}\left\{\mathrm{mult}\left(\mathbf{z}^{=\ell} \cdot \sigma_Y\left(\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}\right)\right)\right\}\right).$$

Since $\partial^{=k}_{\mathscr{P}_{\Delta_0}}(f_{n,\alpha,k}) \subseteq \partial^{=k}_{Y}(f_{n,\alpha,k})$ and $\mathbf{z}^{=\ell} \subseteq \mathbf{z}^{\leqslant \ell}$, we get that

$$\begin{aligned}\dim\left(\mathbb{F}\text{-span}\left\{\mathrm{mult}\left(\mathbf{z}^{=\ell} \cdot \sigma_Y\left(\partial^{=k}_{\mathscr{P}_{\Delta_0}} f_{n,\alpha,k}\right)\right)\right\}\right) &\leqslant \dim\left(\mathbb{F}\text{-span}\left\{\mathrm{mult}\left(\mathbf{z}^{\leqslant \ell} \cdot \sigma_Y\left(\partial^{=k}_{Y} f_{n,\alpha,k}\right)\right)\right\}\right) \\ &= \Gamma_{k,\ell}(f_{n,\alpha,k}).\end{aligned}$$

Putting this together with Lemma 15 we get that $\Gamma_{k,\ell}(f_{n,\alpha,k}) \geqslant \frac{n^{(2-\delta)k}\binom{m-(d-k)}{\ell}}{4\delta k\binom{2k}{\delta k}}$. □

3.3 Putting it all together

We shall now prove a size lower bound against depth four multi-r-ic circuits of low bottom support that compute $f_{n,\alpha,k}$ by instantiating α to a suitable value that is smaller than $\frac{0.98\cdot(2-\delta)\log n}{\delta\log\left(\frac{2}{1-\varepsilon}\right)}$ for some fixed constants δ and ε.

Lemma 17. *There exist constants δ, ε and ν in $(0,1)$ such that $\frac{(1-\delta-\nu)}{2\log\left(\frac{2}{1+\varepsilon}\right)} \leqslant \frac{0.98(2-\delta)}{\delta\log\left(\frac{2}{1-\varepsilon}\right)}$.*

Proof. Proof by instantiation. Let $\varepsilon = 0.1$, $\delta = 0.1$ and $\nu = 0.85$.

- $(1-\delta-\nu) = 0.05$,
- $\frac{2}{1+\varepsilon} \approx 1.818$,
- $2\log\left(\frac{2}{1+\varepsilon}\right) \approx 1.725$,
- $\frac{(1-\delta-\nu)}{2\log\left(\frac{2}{1+\varepsilon}\right)} \approx 0.029$,
- $0.98(2-\delta) = 1.862$,
- $\frac{2}{1-\varepsilon} \approx 2.22$,
- $\delta\log\left(\frac{2}{1-\varepsilon}\right) \approx 0.1152$ and
- $\frac{0.98(2-\delta)}{\delta\log\left(\frac{2}{1-\varepsilon}\right)} \approx 16.163$.

$\square$

Remark: There exist a lot of constants that satisfy the condition in Lemma 17. We can choose the constants such that μ and α are integers.

Theorem 18. *Let δ, ε and ν be some constants as obtained from Lemma 17. Let n be an asymptotically large integer. Let r, α and μ be such that*

- $r \leqslant n^{0.5\nu}$,
- $\mu = \frac{0.4\nu\log n}{\log\left(\frac{1+\varepsilon}{1-\varepsilon}\right)}$ *and*
- $\alpha = \frac{(1-\delta-\nu)\log n}{2\log\left(\frac{2}{1+\varepsilon}\right)}$.

Let C be a depth four multi-r-ic circuit of bottom support at most μ and size s. If C computes the polynomial $f_{n,\alpha,k}$ then s must at least be $n^{0.09\nu k}$.

Proof. Let δ, ε and ν be the constants obtained from Lemma 17. For a fixed value of $\alpha = \frac{(1-\delta-\nu)\log n}{2\log\left(\frac{2}{1+\varepsilon}\right)}$, the polynomial $f_{n,\alpha,k}$ is defined on the variable sets Y and Z such that $|Z| = m = 2\alpha nk$. Let ℓ, k, μ be such that $\ell = \frac{m}{2}(1-\varepsilon)$, $k^2\mu^2 = o(m)$ and $\ell + k\mu < \frac{m}{2}$. Let $\Delta_0 = \delta k$. Let us assume that the polynomial

$f_{n,\alpha,k}$ is computed by a depth four multi-r-ic circuit C of bottom support at most μ and size s. Then it must be the case that $\Gamma_{k,\ell}(f_{n,\alpha,k}) = \Gamma_{k,\ell}(C)$.

Invoking Lemma 16 with $\alpha = \frac{(1-\delta-\nu)\log n}{2\log\left(\frac{2}{1+\varepsilon}\right)} \leqslant \frac{0.98(2-\delta)\log n}{\delta\log\left(\frac{2}{1-\varepsilon}\right)}$, and the values of ε, δ and ν obtained from Lemma 17, we get that

$$\Gamma_{k,\ell}(f_{n,\alpha,k}) \geqslant \frac{n^{(2-\delta)k}\binom{m-(d-k)}{\ell}}{4\delta k\binom{2k}{\delta k}}.$$

Invoking Lemma 11 with $\ell + k\mu < \frac{m}{2}$, we get that

$$\Gamma_{k,\ell}(C) \leqslant s \cdot \binom{\frac{2mr}{\mu}+1}{k} \cdot \binom{m}{\ell+k\mu} \cdot (\ell+k\mu).$$

Putting these two together with the fact that $\Gamma_{k,\ell}(f_{n,\alpha,k}) = \Gamma_{k,\ell}(C)$, we get the following.

$$\begin{aligned}
s &\geqslant \frac{O(1)\cdot n^{(2-\delta)k}\cdot\binom{m-(d-k)}{\ell}}{\delta k\cdot\binom{2k}{\delta k}\cdot\binom{\frac{2mr}{\mu}+1}{k}\cdot\binom{m}{\ell+k\mu}\cdot(\ell+k\mu)} \\
&\geqslant \frac{O(1)}{\delta k\cdot(\ell+k\mu)}\cdot n^{(2-\delta)k}\cdot\left(\frac{\delta}{2e}\right)^{\delta k}\cdot\left(\frac{k\mu}{2emr+e\mu}\right)^{k}\cdot\frac{\binom{m-(d-k)}{\ell}}{\binom{m}{\ell+k\mu}} \\
&\geqslant \frac{O(1)}{\delta k\cdot(\ell+k\mu)}\cdot n^{(2-\delta)k}\cdot\left(\frac{\delta}{2e}\right)^{\delta k}\cdot\left(\frac{k\mu}{3emr}\right)^{k}\cdot\frac{(m-(d-k))!}{m!}\cdot\frac{(m-\ell-k\mu)!}{(m-\ell-(d-k))!}\cdot\frac{(\ell+k\mu)!}{\ell!} \\
&\approx \frac{O(1)}{\delta k\cdot(\ell+k\mu)}\cdot n^{(2-\delta)k}\cdot\left(\frac{\delta}{2e}\right)^{\delta k}\cdot\left(\frac{k\mu}{3emr}\right)^{k}\cdot\frac{\ell^{k\mu}}{m^{(d-k)}}\cdot(m-\ell)^{(d-k)-k\mu} \\
&= \frac{O(1)}{\delta k\cdot(\ell+k\mu)}\cdot n^{(2-\delta)k}\cdot\left(\frac{\delta}{2e}\right)^{\delta k}\cdot\left(\frac{k\mu}{3emr}\right)^{k}\cdot\left(\frac{\ell}{m-\ell}\right)^{k\mu}\cdot\left(\frac{m-\ell}{m}\right)^{d-k} \\
&= \frac{O(1)}{\delta k\cdot(\ell+k\mu)}\cdot n^{(1-\delta)k}\cdot\left(\frac{\delta}{2e}\right)^{\delta k}\cdot\left(\frac{\mu}{6e\alpha r}\right)^{k}\cdot\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{k\mu}\cdot\left(\frac{1+\varepsilon}{2}\right)^{2\alpha k} \\
&= \frac{O(1)}{\delta k\cdot(\ell+k\mu)}\cdot\exp\left[k\left((1-\delta)\log n - 2\alpha\log\left(\frac{2}{1+\varepsilon}\right) - \log r - \mu\log\left(\frac{1+\varepsilon}{1-\varepsilon}\right)\right)\right]\cdot\left(\frac{\delta}{2e}\right)^{\delta k}\cdot\left(\frac{\mu}{6e\alpha}\right)^{k} \\
&= \frac{O(1)}{\delta k\cdot(\ell+k\mu)}\cdot\exp\left[k\left(\nu\log n - \log r + \log\left(\frac{\mu}{6e\alpha}\right) + \delta\log\left(\frac{\delta}{2e}\right) - \mu\log\left(\frac{1+\varepsilon}{1-\varepsilon}\right)\right)\right].
\end{aligned}$$

In line 2 of the above math block, we use the inequality $\binom{n}{k} \leqslant \left(\frac{en}{k}\right)^k$. In line 4, we use Lemma 8 to simplify the terms along with the fact that $k^2\mu^2 = o(m-\ell)$, $(d-k)^2 = o(m)$ and $k^2\mu^2 = o(\ell)$. In line 6, we substitute $2\alpha nk$ for m and simplify the terms.

Recall that ε, δ and ν are constants in $(0,1)$ given by Lemma 17 and $\frac{\mu}{\alpha} = O(1)$. If $\mu\log\left(\frac{1+\varepsilon}{1-\varepsilon}\right) + \log r \leqslant 0.9\nu\log n$, we get that

$$s \geqslant \frac{O(1)\cdot n^{0.1\nu k}}{\delta k\cdot(\ell+k\mu)}\cdot\left(\frac{\delta}{2e}\right)^{\delta k}\cdot\left(\frac{\mu}{6e\alpha}\right)^{k} \geqslant \frac{n^{0.1\nu k}}{2^{O(k)}} \geqslant n^{0.09\nu k}$$

for all asymptotically large enough n.

□

4 Multi-r-ic Depth Four Circuits

We shall now define another polynomial family $P_{n,\alpha,k}$ based on the definition of $f_{n,\alpha,k}$ and then prove a lower bound for the polynomial family $P_{n,\alpha k}$ against multi-r-ic depth four circuits by lifting the lower bound for $f_{n,\alpha,k}$ against multi-r-ic depth four circuits of low bottom support.

Let c be a fixed constant in $(0,1)$. Let $\hat{X} = \{\hat{x}_{1,1}, \hat{x}_{1,2}, \ldots, \hat{x}_{1,t}, \ldots, \hat{x}_{N_0,1}, \hat{x}_{N_0,2}, \ldots, \hat{x}_{N_0,t}\}$ be a variable set distinct from X such that $t = N_0^{1+c} + N_0^c \ln N_0$. Then the polynomial $P_{n,\alpha,k}(\hat{X})$ is defined as follows.

$$P_{n,\alpha,k}(\hat{X}) = f_{n,\alpha,k}\left(\sum_{j=1}^{t} \hat{x}_{1,j}, \sum_{j=1}^{t} \hat{x}_{2,j}, \cdots, \sum_{j=1}^{t} \hat{x}_{N_0,j}\right).$$

Note that $P_{n,\alpha,k}$ is a polynomial on $N = N_0^{2+c} + N_0^{1+c} \ln N_0$ many variables and $\deg(P_{n,\alpha,k}) = \deg(f_{n,\alpha,k})$.

Definition 19 (p-projections). *A polynomial $g(y_1, \ldots, y_m)$ is said to be a p-projection of the polynomial $h(x_1, \ldots, x_n)$ if there exists a suitable substitution $\phi : X \mapsto Y \cup \mathbb{F}$, of $X = \{x_1, \ldots, x_n\}$ by either variables in $Y = \{y_1, \ldots, y_m\}$ or constants from the base field such that*

$$g(y_1, \ldots, y_m) = h(\phi(x_1), \ldots, \phi(x_n)).$$

It is easy to see that if $h(x_1, \ldots, x_n)$ has a circuit of size s then so does $g(y_1, \ldots, y_m)$.

Let us now recall the following lemmas from [Sap19]. Proofs of these lemmas are a step by step adaptation, rather a replication of proofs of Lemma 20.5 and Lemma 20.4 respectively in [Sap19].

We shall first show that the polynomial $P_{n,\alpha,k}$ *reduces* to the polynomial $f_{n,\alpha,k}$ upon taking random restrictions and p-projections, with a high probability.

Lemma 20 (Analogous to Lemma 20.5[3], [Sap19]). *Let c be a constant as fixed above. Let ρ be a random restriction on the variable set $\hat{X}$ that sets each variable to zero independently, with a probability of $(1 - N_0^{-c})$. Then $f_{n,\alpha,k}(X)$ is a p-projection of $\rho(P_{n,\alpha,k}(\hat{X}))$ with a probability of at least $(1 - e^{-N_0})$.*

Proof. For all $i \in [N_0]$, probability that all the variables $\hat{x}_{i,j}$ ($j \in [t]$) are set to zero by ρ is as follows.

$$\Pr[\rho(\hat{x}_{i,1}) = \rho(\hat{x}_{i,2}) = \cdots \rho(\hat{x}_{i,t}) = 0] = (1 - N_0^{-c})^t \approx e^{-\frac{t}{N_0^c}} = e^{-\frac{N_0^{1+c} + N_0^c \ln N_0}{N_0^c}} = \frac{1}{N_0 e^{N_0}}.$$

By union bound, the probability that there exists an $i \in [N_0]$ such that all the variables of the form $\hat{x}_{i,j}$ for $j \in [t]$ are set to zero is at most $\frac{1}{e^{N_0}}$. Thus, with a probability of at least $(1 - e^{-N_0})$, for each i, there exists at least one j such that $\rho(\hat{x}_{i,j}) \neq 0$. It is easy to see that the polynomial $f_{n,\alpha,k}$ can be written as a p-projection of $\rho(P_{n,\alpha,k})$ in such a case. For each $i \in [N_0]$, the substitution maps one of the non-zero $\rho(\hat{x}_{i,j})$'s to x_i and sets the rest to 0. $\square$

We shall now show that, under random restrictions any syntactically multi-r-ic depth four circuit *reduces* to a syntactically multi-r-ic depth four circuit of low bottom support with a high probability and without any blow up in size.

[3] The form of this lemma as mentioned in [Sap19] is due to Kumar and Saptharishi.

Lemma 21 (Analogous to Lemma 20.4, [Sap19]). *Let $\gamma > 0$ be a parameter. Let N and μ be integers. Let P be a N-variate polynomial that is computed by a syntactically multi-r-ic depth 4 circuit C of size $s \leqslant N^{\gamma\mu}$. Let ρ be a random restriction that sets each variable to zero independently with probability $(1 - N^{-2\gamma})$. Then with a probability of at least $(1 - N^{-\gamma\mu})$, polynomial $\rho(P)$ is computed by a multi-r-ic depth four circuit C' of bottom support at most μ, and size s.*

Proof. Let C be a multi-r-ic depth four circuit of size s computing P. Let $\{M_1, M_2, \ldots, M_t\}$ be the set of monomials computed at the lower product gate of C which have at least $\mu + 1$ distinct variables in their support. Note that t is at most s. For all $i \in [t]$,

$$\Pr[\rho(M_i) \neq 0] < \left(N^{-2\gamma}\right)^{\mu}.$$

By taking a union bound, the probability that there exists in a monomial amongst $\{M_1, M_2, \ldots, M_t\}$ that is not set to 0 by ρ is strictly less than $t \cdot N^{-2\gamma\mu} \leqslant s \cdot N^{-2\gamma\mu} \leqslant N^{-\gamma\mu}$. Thus with a probability of at least $(1 - N^{-\gamma\mu})$, all the monomials at the bottom product gate depend on at most μ distinct variables. □

With this background, we are now ready to present the proof of Theorem 1.

Proof of Theorem 1. Let ε, δ and ν be the constants obtained from Lemma 17 and c be a small constant in $(0, 1)$ as fixed above. Let n be a large positive integer. Let the parameters N, N_0, r, μ, α and k be set in terms of n or otherwise as follows.

- $r \leqslant n^{0.5\nu}$,
- $\mu = \frac{0.4\nu \log n}{\log\left(\frac{1+\varepsilon}{1-\varepsilon}\right)}$,
- $\alpha = \frac{(1-\delta-\nu)\log n}{2\log\left(\frac{2}{1+\varepsilon}\right)}$,
- $N_0 = k(n^2 + 2\alpha n k)$,
- $N = N_0^{2+c} + N_0^{1+c} \ln N_0$,
- γ be a parameter given by the equation $N^{2\gamma} = N_0^c$ and
- $k = \frac{10\gamma\mu \log N}{\nu \log n}$.

The above setting of parameters also satisfies the condtions that $k^2\mu^2 = o(m)$ and $(d-k)^2 = O(\alpha^2 k^2) = o(m)$.

Let $\hat{X} = \{\hat{x}_{1,1}, \hat{x}_{1,2}, \ldots, \hat{x}_{1,t}, \ldots, \hat{x}_{N_0,1}, \hat{x}_{N_0,2}, \ldots, \hat{x}_{N_0,t}\}$ be a set of variables over which the polynomial $P_{n,\alpha,k}$ is defined where $t = N_0^{1+c} + N_0^c \ln N_0$. Let ρ be a random restriction such that a variable is set to zero with a probability of $(1 - N_0^{-c}) = (1 - N^{-2\gamma})$, and is left untouched otherwise. Let C be a syntactically multi-r-ic depth four circuit of size $s \leqslant N^{\gamma\mu}$ that computes $P_{n,\alpha,k}$.

Lemma 21 tells us that $C' = \rho(C)$ is a multi-r-ic depth four circuit of size s and bottom support at most μ with a probability of at least $(1 - N^{-\gamma\mu})$. Conditioned on this probability, $\rho(P_{n,\alpha,k})$ has a multi-r-ic $\Sigma\Pi\Sigma\Pi^{\{\mu\}}$ size at most s.

By invoking Lemma 20, we get that $f_{n,\alpha,k}$ is a p-projection of $\rho(P_{n,\alpha,k})$ with a probability of at least $(1-e^{-N_0})$. Since $\rho(P_{n,\alpha,k})$ has a multi-r-ic $\Sigma\Pi\Sigma\Pi^{\{\mu\}}$ circuit of size at most s with a probability of at least $1-N^{-\gamma\mu}$, with a probability of at least $(1-N^{-\gamma\mu}-e^{-N_0})$, $f_{n,\alpha,k}$ is computed by a multi-r-ic $\Sigma\Pi\Sigma\Pi^{\{\mu\}}$ circuit of size at most s. In other words, there exists a multi-r-ic depth four circuit of bottom support at most μ and size at most s, that computes $f_{n,\alpha,k}$.

On the other hand, by invoking Theorem 18 with the set of parameters as defined above, we get that any multi-r-ic $\Sigma\Pi\Sigma\Pi^{\{\mu\}}$ circuit that computes $f_{n,\alpha,k}$ must be of size $\exp((0.09\nu k \log n)$. Upon putting both of these facts together, it must be the case that

$$n^{0.09\nu k} = N^{0.9\gamma\mu} \leqslant s \leqslant N^{\gamma\mu}.$$

Since ε, δ and ν are constants, and $N = n^{O(1)}$, we get that s must at least be $\exp\left(\Omega(\log^2 n)\right)$. The explicit polynomial Q_n is $P_{n,\alpha,k}$ where α and k are set to values described above. $\square$

Acknowledgements

The author was supported by the Institite Post Doctoral Fellowship at IIT Bombay where a part of this work was done. The author is currently supported by a CHE-PBC (VATAT) fellowship at University of Haifa. The author is also affiliated to Caesarea Rothschild Institute at University of Haifa. The author is grateful to Nutan Limaye and Srikanth Srinivasan for listening to a presentation of a preliminary version of this paper. We thank the anonymous reviewers for pointing out an error in instantiation in Theorem 18 and for helping the paper take the current form.

References

[AKV18] Noga Alon, Mrinal Kumar, and Ben Lee Volk. Unbalancing sets and an almost quadratic lower bound for syntactically multilinear arithmetic circuits. In *CCC*, volume 102 of *LIPIcs*, pages 11:1–11:16. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018.

[AV08] Manindra Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *proceedings of Foundations of Computer Science (FOCS)*, pages 67–75, 2008. doi:10.1109/FOCS.2008.32.

[BS83] Walter Baur and Volker Strassen. The complexity of partial derivatives. *Theor. Comput. Sci.*, 22:317–330, 1983.

[CELS18] Suryajith Chillara, Christian Engels, Nutan Limaye, and Srikanth Srinivasan. A near-optimal depth-hierarchy theorem for small-depth multilinear circuits. In *FOCS*, pages 934–945. IEEE Computer Society, 2018.

[CLS18] Suryajith Chillara, Nutan Limaye, and Srikanth Srinivasan. A quadratic size-hierarchy theorem for small-depth multilinear formulas. In *ICALP*, volume 107 of *LIPIcs*, pages 36:1–36:13. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018.

[CLS19] Suryajith Chillara, Nutan Limaye, and Srikanth Srinivasan. Small-depth multilinear formula lower bounds for iterated matrix multiplication with applications. *SIAM J. Comput.*, 48(1):70–92, 2019.

[CM19] Suryajith Chillara and Partha Mukhopadhyay. Depth-4 lower bounds, determinantal complexity: A unified approach. *computational complexity*, May 2019. URL: https://doi.org/10.1007/s00037-019-00185-4, doi:10.1007/s00037-019-00185-4.

[FLMS15] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth-4 formulas computing iterated matrix multiplication. *SIAM J. Comput.*, 44(5):1173–1201, 2015.

[GKKS14] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the chasm at depth four. *Journal of the ACM (JACM)*, 61(6):33, 2014.

[GRS19] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. *Essential coding theory*. 2019. URL: https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/.

[HS17] Sumant Hegde and Chandan Saha. Improved lower bound for multi-r-ic depth four circuits as a function of the number of input variables. *Proceedings of the Indian National Science Academy*, 83(4):907–922, 2017.

[HY11] Pavel Hrubeš and Amir Yehudayoff. Homogeneous formulas and symmetric polynomials. *Computational Complexity*, 20(3):559–578, 2011.

[Kal85] K. Kalorkoti. A lower bound for the formula size of rational functions. *SIAM J. Comput.*, 14(3):678–687, 1985.

[Kay12] Neeraj Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:81, 2012.

[KdOS19] Mrinal Kumar, Rafael Mendes de Oliveira, and Ramprasad Saptharishi. Towards optimal depth reductions for syntactically multilinear circuits. In *ICALP*, volume 132 of *LIPIcs*, pages 78:1–78:15. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2019.

[KLSS14] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. Super-polynomial lower bounds for depth-4 homogeneous arithmetic formulas. In *STOC*, pages 119–127. ACM, 2014.

[KLSS17] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An exponential lower bound for homogeneous depth four arithmetic formulas. *SIAM J. Comput.*, 46(1):307–335, 2017.

[KNS16] Neeraj Kayal, Vineet Nair, and Chandan Saha. Separation between read-once oblivious algebraic branching programs (ROABPs) and multilinear depth three circuits. In *proceedings of Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 46:1–46:15, 2016. URL: https://doi.org/10.4230/LIPIcs.STACS.2016.46, doi:10.4230/LIPIcs.STACS.2016.46.

[Koi12] Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theor. Comput. Sci.*, 448:56–65, 2012. URL: https://doi.org/10.1016/j.tcs.2012.03.041, doi:10.1016/j.tcs.2012.03.041.

[KS17a] Neeraj Kayal and Chandan Saha. Multi-k-ic depth three circuit lower bound. *Theory Comput. Syst.*, 61(4):1237–1251, 2017.

[KS17b] Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. *SIAM J. Comput.*, 46(1):336–387, 2017.

[KSS14] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *STOC*, pages 146–153. ACM, 2014.

[KST18] Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. On the size of homogeneous and of depth-four formulas with low individual degree. *Theory of Computing*, 14(16):1–46, 2018. URL: http://www.theoryofcomputing.org/articles/v014a016, doi:10.4086/toc.2018.v014a016.

[NW97] Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997. doi:10.1007/BF01294256.

[Raz04] Ran Raz. Multilinear-$\mathsf{NC}^2 \neq$ multilinear-NC^1. In *proceedings of Foundations of Computer Science (FOCS)*, pages 344–351, 2004. URL: https://doi.org/10.1109/FOCS.2004.42, doi:10.1109/FOCS.2004.42.

[Raz06] Ran Raz. Separation of multilinear circuit and formula size. *Theory of Computing*, 2(1):121–135, 2006. doi:10.4086/toc.2006.v002a006.

[RSY08] Ran Raz, Amir Shpilka, and Amir Yehudayoff. A lower bound for the size of syntactically multilinear arithmetic circuits. *SIAM Journal of Computing*, 38(4):1624–1647, 2008. doi:10.1137/070707932.

[RY08] Ran Raz and Amir Yehudayoff. Balancing syntactically multilinear arithmetic circuits. *Computational Complexity*, 17(4):515–535, 2008. doi:10.1007/s00037-008-0254-0.

[RY09] Ran Raz and Amir Yehudayoff. Lower bounds and separations for constant depth multilinear circuits. *Computational Complexity*, 18(2):171–207, 2009. doi:10.1007/s00037-009-0270-8.

[Sap19] Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity version 8.0.4. Github survey, 2019. URL: https://github.com/dasarpmar/lowerbounds-survey/releases/.

[Str73] Volker Strassen. Berechnungen in partiellen algebren endlichen typs. *Computing*, 11(3):181–196, 1973.

[SY10] Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5:207–388, March 2010. URL: http://dx.doi.org/10.1561/0400000039.

[Tav15] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. *Information and Computation*, 240:2–11, 2015. doi:10.1016/j.ic.2014.09.004.

[Val79] Leslie G. Valiant. Completeness classes in algebra. In *STOC*, pages 249–261. ACM, 1979.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il