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

**Suryajith Chillara**
CRI, University of Haifa, Israel
http://cmi.ac.in/~suryajith/
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 polynomial computed at every node has a bound on the individual degree of $r$ (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 and 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 a 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 with increasing $r$. It is a natural question to ask if we can prove a bound that does not deteriorate with increasing $r$ or a bound that holds for a *larger* regime of $r$.

We here prove a lower bound which does not deteriorate with $r$, however for a specific instance of $d = d(n)$ but for a *wider* range of $r$. Formally, 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 $r < n^{0.2}$ 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).

**2012 ACM Subject Classification** Theory of computation → Circuit complexity; Theory of computation → Algebraic complexity theory

**Keywords and phrases** Lower Bounds, Multilinear, Multi-$r$-ic

**Digital Object Identifier** 10.4230/LIPIcs.STACS.2020.47

**Funding** 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.

**Acknowledgements** The author is grateful to Nutan Limaye and Srikanth Srinivasan for their invaluable feedback.

## 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 arithmetic 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.

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 ×. Every node computes a polynomial by operating on its inputs with the operand

37th International Symposium on Theoretical Aspects of Computer Science (STACS 2020).
Editors: Christophe Paul and Markus Bläser; Article No. 47; pp. 47:1–47:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

given by its label. The flow of computation flows from the leaf to the output node. We refer the readers to the standard resources [30, 29] for more information on arithmetic formulas and arithmetic circuits.

Valiant conjectured that the Permanent does not have polynomial sized arithmetic circuits [33]. 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 [31, 3], and, the best known formula size lower bound is $\Omega(n^2)$, due to Kalorkoti [13]. Due to the slow progress towards proving general circuit/formula lower bounds, it is natural to study some restricted class 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 [23, 25, 24, 27, 28, 12, 26, 2, 6, 4, 5], the complexity of multilinear formulas and circuits is better understood than that of general formulas and circuits.

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

Kayal and Saha [18] first studied multi-$r$-ic circuits of depth three and proved exponential lower bounds. Kayal, Saha and Tavenas [20] 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 any product node, any variable appears in the support of at most $r$ many operands, and the total of the individual degrees 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 [21] 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 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 [20] proved an exponential size lower bound against multi-$r$-ic depth four circuits computing a variant of the iterated matrix multiplication polynomial. They achieved this bound using a measure that is inspired by the method *Shifted Partial Derivatives* [14, 9] and the method of *Skew Partial Derivatives* [17]. They referred to this new technique as the method of *Shifted Skew Partial Derivatives.* Hegde and Saha [11] improved upon [20] and showed a *near-optimal* size lower bound. However, the *best known* lower bounds are for polynomials that are multi-$r$-ic.

## Motivation for this work

Raz and Yehudayoff [28] 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. [20, Footnote 9]). Kayal, Saha and Tavenas [20] 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

[1] Agrawal and Vinay [1], Koiran, and Tavenas [32] showed that any general circuit can be depth reduced to a depth four circuit of non-trivial size.

by a multi-$r$-ic depth four circuit. This lower bound degrades with the increasing value of $r$ and 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 with increasing values of $r$.

▶ **Theorem 1.** *Let $n$ and $r$ be integers such that $r < n^{0.2}$. There exists an explicit polynomial $Q_n$ of degree $\Theta(\log^2 n)$, over $n^{O(1)}$ variables such that any syntactically multi-$r$-ic depth four circuit computing it must have size $\exp\left(\Omega(\log^2 n)\right)$.*

The explicit polynomial that we consider can be expressed as a $p$-projection of an 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. By substituting for $d = \Theta(\log^2 n)$ into the bound from [20], 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 lower bound is quantitatively better in this regime of parameters. In particular, we show a lower bound in the regimes of parameters where [20] cannot.

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 [20, 11], Theorem 1 is a step towards that direction.

## Proof overview

Analogous to the work of Fournier et al. [8], and Kumar and Saraf [22], 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 obtained by adding the terms $T_1, T_2, \ldots, T_s$. Note that each of these $T_i$'s is a product of low support polynomials $Q_{i,j}$, that is, every monomial in these $Q_{i,j}$'s is supported on a small set of variables (say $\mu$ many). One major observation at this point is to see that there can be at most $N \cdot r$ many factors in any of the $T_i$'s.

Kayal et al. [20] observed that the measure of shifted partial derivates [19, 8] 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 the variables into two disjoint and unequal sets $Y$ and $Z$ such that $|Y| \gg |Z|$, then considered the *low* order partial derivatives with respect to only the $Y$ variables and subsequently set all the $Y$ variables to zero in the partial derivatives obtained. This effectively reduces the number of factors in a partial derivative to at most $|Z| \cdot r$. They then shift these polynomials by monomials in $Z$ variables and look at the dimension 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

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

terms obtained after the shifts of the skew partial derivatives. This is analogous to the method employed by Kayal et al. [16] to prove exponential 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 suitable set of parameters.

## 2 Preliminaries

### Notation

- For a polynomial $f$, 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 $\mathbf{z}^{\leq \ell}$ to refer to the set of all the monomials of degree equal to $\ell$ and at most $\ell$, respectively, in $Z$ variables.
- We use $\mathbf{z}_{\mathrm{ML}}^{\leq \ell}$ to refer to the set of all the multilinear monomials of degree at most $\ell$ in $Z$ variables.
- We use $\mathbf{z}_{\mathrm{NonML}}^{\leq \ell}$ to refer to the set of all the non-multilinear monomials of degree at most $\ell$ in $Z$ variables.
- For a monomial $m$ we use $|\mathrm{MonSupp}(m)|$ to refer to the size of the set of variables that appear in it.
- For a polynomial $f$, we use $|\mathrm{MonSupp}(f)|$ to refer to the maximum $|\mathrm{MonSupp}(m)|$ over all monomials in $f$.

### 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, $\displaystyle\sum_{i=1}^{s}\prod_{j=1}^{d_i} Q_{i,j}(x_1, \ldots, x_n)$ 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 each polynomial $Q_{i,j}$ are supported on at most $t$ variables.

### Multi-$\mathbf{r}$-ic arithmetic circuits

▶ **Definition 2** (multi-$\mathbf{r}$-ic circuits). *Let $\mathbf{r} = (r_1, r_2, \cdots, r_n)$. An arithmetic circuit $\Phi$ is said to be a syntactically multi-$\mathbf{r}$-ic circuit if for all product gates $u \in \Phi$ and $u = u_1 \times u_2 \times \cdots \times u_t$, each variable $x_i$ can appear in at most $r_i$ many of the $u_i$'s ($i \in [t]$). Further the total degree with respect to every variable over the polynomials computed at $u_1, u_2, \cdots, u_t$, is bounded by $r$, i.e. $\sum_{j\in[t]} \deg_{x_i}(f_{u_j}) \leq r$ 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 [20].

This is analogous to the work of [15] where they study a *shifted partials inspired* measure called *Shifted Projected Partial derivatives* and then [16] 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 an *order* larger than $|Z|$. Let $\sigma_Y : \mathbb{F}[Y \sqcup Z] \mapsto \mathbb{F}[Z]$ be a map such that for any polynomial $f(Y, Z)$, $\sigma_Y(f) \in \mathbb{F}[Z]$ is obtained by setting every variable in $Y$ to zero by it and leaving $Z$ variables untouched. Let $\mathrm{mult} : \mathbb{F}[Z] \mapsto \mathbb{F}[Z]$ be a map such that for any polynomial $f(Y, Z)$, $\mathrm{mult}(f) \in \mathbb{F}[Z]$ is obtained by setting the coeficients of all the non-multilinear monomials in $f$ to 0 and leaving the rest untouched. We use $\mathbf{z}^{\leq \ell} \cdot \sigma_Y(\partial_Y^{=k} f)$ to refer to the linear span of polynomials obtained by multiplying each polynomial in $\sigma_Y(\partial_Y^{=k} f)$ with monomials of degree at most $\ell$ in $Z$ variables. We will now define our complexity measure, Dimension of Projected Shifted Skew Partial Derivatives (denoted by $\Gamma_{k,\ell}$) as follows.

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

This is a natural generalization of Shifted Skew Partial Derivatives of [20]. The following proposition is easy to verify.

▶ **Proposition 3** (Sub-additivity). *Let $k$ and $\ell$ be integers. Let the polynomials $f, f_1, f_2$ be such that $f = f_1 + f_2$. Then, $\Gamma_{k,\ell}(f) \leq \Gamma_{k,\ell}(f_1) + \Gamma_{k,\ell}(f_2)$.*

### Monomial Distance

We recall the following definition of distance between monomials from [7].

▶ **Definition 4** (Definition 2.7, [7]). *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.

The following beautiful lemma (from [9]) is key to the asymptotic estimates required for the lower bound analyses.

▶ **Lemma 5** (Lemma 6, [9]). *Let $a(n), f(n), g(n) : \mathbb{Z}_{\geq 0} \to \mathbb{Z}_{\geq 0}$ 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 use the following strengthening of the Principle of Inclusion and Exclusion in our proof.

▶ **Lemma 6** (Strong Inclusion-Exclusion [22]). *Let $W_1, W_2, \cdots, W_t$ be subsets of a finite set $W$. For a parameter $\lambda \geq 1$, let $\sum_{\substack{i,j \in [t] \\ i \neq j}} |W_i \cap W_j| \leq \lambda \sum_{i \in [t]} |W_i|$ . Then, $\left|\cup_{i \in [t]} W_i\right| \geq \frac{1}{4\lambda} \sum_{i \in [t]} |W_i|$ .*

### Polynomial Families

Let $n, \alpha, k$ be positive integers. We define the polynomial family $\{P_{n,\alpha,k}\}_{n,\alpha,k \geq 0}$ as follows. Let the variable set $X = \{x_1, \ldots, x_{N_0}\}$ be partitioned into two fixed, and disjoint sets $Y$ and $Z$. We first define the polynomial family $f_{n,\alpha,k}(Y, Z)$ as follows (as it was defined in [20]).

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

It is easy to see that $|Y|$ is $n^2 k$ and $|Z|$ is $2\alpha nk$. We shall henceforth use $m$ to refer to $|Z|$. Thus, $N_0 = |X| = |Y| + |Z| = k(n^2 + 2\alpha n)$.

Let $c$ be a fixed constant in $(0, 1)$. We shall now define another polynomial family $P_{n,\alpha,k}$ based on the definition of $f_{n,\alpha,k}$. Let $p = N_0^{-c}$. 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 $t = \frac{N_0 \log N_0}{p}$. Let $\mathrm{Lin}_p : X \mapsto \hat{X}$ be a linear map such that $x_i \mapsto \sum_{j=1}^{t} \hat{x}_{i,j}$ for all $i \in [N_0]$. Then the polynomial $P_{n,\alpha,k}(\hat{X})$ is defined to be $f_{n,\alpha,k} \circ \mathrm{Lin}_p(\hat{X})$. That is,

$$P_{n,\alpha,k} = f_{n,\alpha,k}\left(\sum_{j=1}^{t} \hat{x}_{1,j}, \sum_{j=1}^{t} \hat{x}_{2,j}, \cdots, \sum_{j=1}^{t} \hat{x}_{N_0,j}\right) \quad \text{where } t = \frac{N_0 \log N_0}{p}.$$

Note that $P_{n,\alpha,k}$ is a polynomial on $N = N_0^{2+c} \log N_0$ many variables.

We will now recall the following lemma which in the mentioned form is due to Kumar and Saptharishi [29].

▶ **Lemma 7** (Analogous to Lemma 20.5, [29]). *Let $\rho$ be a random restriction on the variable set $\hat{X}$ that sets each variable to zero with a probability of at least $(1 - p)$ where $p = N_0^{-c}$ for some constant $c \in (0, 1)$. Then $f_{n,\alpha,k}(X)$ is a projection of $\rho(P_{n,\alpha,k}(\hat{X}))$ with a probability of at least $(1 - 2^{-N_0})$.*

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

For some carefully chosen parameters $k$ and $\ell$, we shall first show that if a multi-$r$-ic depth four circuit $C$ of bottom support $\mu$ is *small* then $\Gamma_{k,\ell}(C)$ is not too large. We shall then show that $\Gamma_{k,\ell}(f_{n,\alpha,k})$ is large and thus it cannot be computed by $C$.

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

Recall that $C$ is 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 $T_i$ and derive a bound on $\Gamma_{k,\ell}(C)$ by using sub-additivity of the measure (cf. Proposition 3).

Let $T$ be a syntactic multi-$r$-ic product of polynomials $\tilde{Q}_1(Y, Z) \cdot \tilde{Q}_2(Y, Z) \cdot \ldots \cdot \tilde{Q}_D(Y, Z) \cdot R(Y)$ such that $\left|\mathrm{MonSupp}(\tilde{Q}_i)\right| \leq \mu$. We will first argue that $D$ is not too large since $T$ is a syntactically multi-$r$-ic product. We shall first pre-process the product $T$ by doing the following procedure.

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

1. Pick two factors $\tilde{Q}_{i_1}$ and $\tilde{Q}_{i,2}$ such that they have the smallest $Z$-support amongst $Q_1, \cdots, Q_D$.
2. If both of them have support strictly less than $\frac{\mu}{2}$, merge these factors to obtain a new factor. Else, stop.

In the afore mentioned procedure, it is important to note that the monomial support post merging will still be at most $\mu$ since the factors being merged are of support strictly less than $\frac{\mu}{2}$. Henceforth, W.L.O.G we shall consider that every product gate at the top, in any multi-$r$-ic depth four circuit to be in a 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. Each of the $Q_i$ has a $Z$-support of at least $\frac{\mu}{2}$. The total $Z$-support is at most $|Z|\, r = mr$ since $T$ is a syntactically multi-$r$-ic product. Thus $t$ could at most be $\frac{2mr}{\mu}$.

▶ **Lemma 8.** *Let $n, k, r, \ell$ and $\mu$ 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 $|\mathrm{MonSupp}(Q_i)| \leq \mu$. Then, $\Gamma_{k,\ell}(T)$ is at most $\binom{t}{k} \cdot \binom{m}{\ell+k\mu} \cdot (\ell + k\mu)$.*

Before proving Lemma 8, 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 9.** *Let $n, k, r, \ell$ and $\mu$ 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 $\mu$ and size $s$. Then, $\Gamma_{k,\ell}(C)$ is at most $s \cdot \binom{\frac{2mr}{\mu}}{k} \cdot \binom{m}{\ell+k\mu} \cdot (\ell + k\mu)$.*

**Proof.** W.L.O.G we can assume that $C$ be expressed as $\sum_i^s T^{(i)}$ such that $T^{(i)}$ is a processed syntactically multi-$r$-ic product of bottom support polynomials at most $\mu$. From Proposition 3, we get that $\Gamma_{k,\ell}(C) \leq \sum_{i=1}^s \Gamma_{k,\ell}(T^{(i)})$. From the afore mentioned discussion we know that the number of factors in $T^{(i)}$ with a non-zero $Z$-support is at most $\frac{2mr}{\mu}$. From Lemma 8, we get that $\Gamma_{k,\ell}(T^{(i)})$ is at most $\binom{\frac{2mr}{\mu}}{k} \cdot \binom{m}{\ell+k\mu} \cdot (\ell + k\mu)$. By putting all of this together, we get that

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

**Proof of Lemma 8.** We will first show by induction on $k$, the following.

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

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

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

where $S$ is a set of size $t-(k-1)$, $h_1(Z)$ is a polynomial in $Z$ variables of degree at most $(k-1)r\mu$, and $h_2(Y)$ is some polynomial in $Y$ variables. In fact, $h_1(Z)$ 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 $(k-1)r\mu$.

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

$$\begin{aligned}
\frac{\partial h(Y,Z)}{\partial y_u} &= \left(\sum_{j\in S}\prod_{\substack{i\in S\\ i\neq j}} Q_i(Y,Z)\cdot\frac{\partial Q_j(Y,Z)}{\partial y_u}\cdot h_1(Z)\cdot h_2(Y)\right)\\
&\qquad + \frac{\prod_{i\in S} Q_i}{Q_{i_0}}\cdot Q_{i_0}(Y,Z)\cdot h_1(Z)\cdot\frac{\partial h_2(Y)}{\partial y_k}\\
&\in \mathbb{F}\text{-span}\left\{\prod_{\substack{i\in S\\ i\neq j}} Q_i(Y,Z)\cdot\frac{\partial Q_j(Y,Z)}{\partial y_u}\cdot h_1(Z)\cdot\mathbb{F}[Y] \mid j\in[S]\right\}\\
&\qquad \bigcup \mathbb{F}\text{-span}\left\{\frac{\prod_{i\in S} Q_i}{Q_{i_0}}\cdot Q_{i_0}(Y,Z)\cdot h_1(Z)\cdot\mathbb{F}[Y]\right\}\\
&\subseteq \mathbb{F}\text{-span}\left\{\bigcup_{T\in\binom{S}{|S|-1}}\left\{\prod_{i\in T} Q_i(Y,Z)\cdot\mathbf{z}_{\mathrm{ML}}^{\mu}\cdot h_1(Z)\cdot\mathbb{F}[Y]\right\}\right\}\\
&\qquad \bigcup\left\{\bigcup_{T\in\binom{S}{|S|-1}}\left\{\prod_{i\in T} Q_i(Y,Z)\cdot\mathbf{z}_{\mathrm{NonML}}^{\leq r\mu}\cdot h_1(Z)\cdot\mathbb{F}[Y]\right\}\right\}\\
&\subseteq \mathbb{F}\text{-span}\left\{\bigcup_{T\in\binom{t}{t-k}}\left\{\prod_{i\in T} Q_i(Y,Z)\cdot\mathbf{z}_{\mathrm{ML}}^{\leq k\mu}\cdot\mathbb{F}[Y]\right\}\right\}\\
&\qquad \bigcup \mathbb{F}\text{-span}\left\{\bigcup_{T\in\binom{t}{t-k}}\left\{\prod_{i\in T} Q_i(Y,Z)\cdot\mathbf{z}_{\mathrm{NonML}}^{\leq kr\mu}\cdot\mathbb{F}[Y]\right\}\right\}
\end{aligned}$$

The last inclusion follows from 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 $(k-1)r\mu$. 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

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

Thus, we get that $\partial^{=k}T$ is a subset of $W$. This completes the proof by induction.

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

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

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

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

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

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

◀

## 3.2 Lower Bound on $\Gamma_{k,\ell}(f_{n,\alpha,k})$

First we recall the generalized Hamming bound [10, Section 1.7].

▷ Claim 10. Let the vectors $\mathbf{a} = (a_1, a_2, \ldots, a_k)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_k)$ correspond to the indices of the $Y$-monomial $y^{(1)}_{a_1,b_1} y^{(2)}_{a_2,b_2} \cdots y^{(k)}_{a_k,b_k}$ that is used to derive $f_{n,\alpha,k}$ with. For every $\Delta_0 < k$, there is a subset $\mathcal{P}_{\Delta_0} \subset [n]^{2k}$ of size $\frac{n^{2k-\Delta_0}}{\Delta_0\binom{2k}{\Delta_0}}$ such that for all $(\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}') \in \mathcal{P}$, $\mathrm{dist}((\mathbf{a}, \mathbf{b}), (\mathbf{a}', \mathbf{b}')) \geq \Delta_0$.

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

▷ Claim 11. Let $(\mathbf{a},\mathbf{b})$ and $(\mathbf{a}',\mathbf{b}')$ be such that $\text{dist}((\mathbf{a},\mathbf{b}),(\mathbf{a}',\mathbf{b}')) \geq \Delta_0$. Then $\text{dist}\left(\partial^k_{(\mathbf{a},\mathbf{b})} f_{n,\alpha,k}, \partial^k_{(\mathbf{a}',\mathbf{b}')} f_{n,\alpha,k}\right) \geq \alpha\Delta_0$.

Let $m_1, m_2, \ldots, m_t$ be the monomials in the set $\mathcal{M}_0 (= \partial^{=k}_{\mathcal{P}_{\Delta_0}} f_{n,\alpha,k})$, over $Z$ variables such that $\text{dist}(m_i, m_j) \geq \Delta \geq \alpha\Delta_0$ for all $i \neq j$. Further, $\sigma_Y(\mathcal{M}_0) = \mathcal{M}_0$. Let $\mathcal{M}$ be the set of mutlilinear monomials of the form $m_i m'$ over $Z$-variables for some $1 \leq i \leq t$ where $m'$ is a monomial of length $\ell$. It is important to note that the set $\mathcal{M}$ now corresponds to the set $\mathbf{z}^{=\ell} \cdot \sigma_Y \left(\partial^{=k}_{\mathcal{P}_{\Delta_0}} f_{n,\alpha,k}\right)$ . We shall now show that the cardinality of the set $\mathcal{M}$ is large enough for a suitable setting of parameters $\alpha$, $\Delta_0$ and $k$.

► **Lemma 12.** *Let $m, k, d, r, \Delta_0, \Delta, \ell$ and $\mu$ be positive integers such that $\ell + k\mu < \frac{m}{2}$, $(d-k)^2 = o(m)$, $\Delta^2 = o(m)$, $\Delta_0 = \delta k$ and $\ell = \frac{m}{2}(1-\varepsilon)$ for some fixed constants $\delta$ and $\varepsilon$. Then, $|\mathcal{M}| \geq \frac{1}{2}\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k} \cdot \binom{m-(d-k)}{\ell}$.*

**Proof.** For all $i \in [t]$, Let $B_i$ be the set of multilinear monomials of the form $m_i m'$ where $m'$ is a multilinear monomial of degree $\ell$. From the previous discussion, it follows that $|\mathcal{M}| = |\cup_{i=1}^t B_i|$. Using the principle of Inclusion and Exclusion, we get that

$$\left|\cup_{i=1}^t B_i\right| \geq \sum_{i=1}^t |B_i| - \sum_{i \neq j \in [t]}^t |B_i \cap B_j| \ .$$

▷ Claim 13. For all $i \in [t]$, $|B_i| = \binom{m-(d-k)}{\ell}$ .

Proof. Since $\deg(m_i)$ is equal to $d-k$, the cardinality of $B_i$ is equal to $\binom{m-(d-k)}{\ell}$. ◁

▷ Claim 14. For all $i, j \in [t]$ such that $i \neq j$, $|B_i \cap B_j| \leq \binom{m-(d-k)-\Delta}{\ell-\Delta}$.

Proof. Consider any two monomials $\hat{m}_i$ and $\hat{m}_j$ from $B_i$ and $B_j$ respectively. For $\hat{m}_i$ and $\hat{m}_j$ to be identical, $\hat{m}_i$ should contain at least $\Delta$ variables from $\hat{m}_j \setminus \hat{m}_i$ and similarly $\hat{m}_j$ should contain at least $\Delta$ variables from $\hat{m}_i \setminus \hat{m}_j$. The rest of the at most $(\ell - \Delta)$ many variables should be the same both in $\hat{m}_i$ and $\hat{m}_j$. The number of such multilinear monomials over $Z$ variables is at most $\binom{m-(d-k)-\Delta}{\ell-\Delta}$. ◁

Putting Claim 13 and Claim 14 together, we get the following.

$$|\mathcal{M}| = \left|\cup_{i=1}^t B_i\right| \geq t\binom{m-(d-k)}{\ell} - \frac{t^2}{2}\binom{m-(d-k)-\Delta}{\ell-\Delta} .$$

Let $T_1 = t\binom{m-(d-k)}{\ell}$ and $T_2 = \frac{t^2}{2}\binom{m-(d-k)-\Delta}{\ell-\Delta}$. Let us consider the case where $T_2 = \lambda T_1$ where $\lambda \geq 1$ for some setting of the parameters $\Delta$, $\alpha$, $\ell$ and $k$.

$$
\begin{aligned}
\lambda = \frac{T_2}{T_1} &= \frac{\frac{t^2}{2}\binom{m-(d-k)-\Delta}{\ell-\Delta}}{t\binom{m-(d-k)}{\ell}} \\
&= \frac{t}{2} \cdot \frac{(m-(d-k)-\Delta)!}{(\ell-\Delta)!(m-\ell-(d-k))!} \cdot \frac{(m-\ell-(d-k))!\ell!}{(m-(d-k))!} \\
&= \frac{t}{2} \cdot \frac{(m-(d-k)-\Delta)!}{(m-(d-k))!} \cdot \frac{\ell!}{(\ell-\Delta)!} \\
&= \frac{t}{2} \cdot \frac{(m-(d-k)-\Delta)!}{m!} \cdot \frac{m!}{(m-(d-k))!} \cdot \frac{\ell!}{(\ell-\Delta)!} \\
&= \frac{t}{2} \cdot \frac{m^{(d-k)} \cdot \ell^{\Delta}}{m^{(d-k)+\Delta}} && \text{(Using Lemma 5)} \\
&= \frac{t}{2} \cdot \left(\frac{\ell}{m}\right)^{\Delta}.
\end{aligned}
$$

The math block above crucially uses the fact that $\Delta^2 = o(\ell)$ and $(d-k)^2 = o(m)$ while invoking Lemma 5. Since $\lambda \geq 1$, we get that $\frac{t}{2} \cdot \left(\frac{\ell}{m}\right)^{\Delta} \geq 1$. For some suitably fixed constants $\delta$ and $\varepsilon$, let $\Delta_0$ be set to $\delta k$ and $\ell$ be set to $\frac{m}{2}(1-\varepsilon)$. Recall that for a fixed $\Delta_0$, $t = \frac{n^{2k-\Delta_0}}{\Delta_0\binom{2k}{\Delta_0}}$ and $\Delta = \alpha\Delta_0 = \delta\alpha k$. Thus,

$$
\begin{aligned}
\frac{n^{2k-\Delta_0}}{2\binom{2k}{\Delta_0}} \cdot \left(\frac{\ell}{m}\right)^{\Delta} &\geq 1 \\
n^{2k-\Delta_0} &\geq 2\Delta_0 \left(\frac{m}{\ell}\right)^{\Delta} \binom{2k}{\Delta_0} \\
n^{2k-\Delta_0} &\geq \left(\frac{2}{1-\varepsilon}\right)^{\Delta} \left(\frac{2k}{\Delta_0}\right)^{\Delta_0} \\
n^{(2-\delta)k} &\geq \left(\frac{2}{1-\varepsilon}\right)^{\alpha\delta k} \left(\frac{2k}{\Delta_0}\right)^{\delta k}
\end{aligned}
$$

and hence,

$$
\alpha \leq \frac{(2-\delta)\log n - \delta \log\left(\frac{2}{\delta}\right)}{\delta \log\left(\frac{2}{1-\varepsilon}\right)}.
$$

Invoking Lemma 6 with the previous discussion, we get that

$$
|\mathcal{M}| \geq \frac{T_1}{4\lambda} = \frac{t\binom{m-(d-k)}{\ell}}{4\lambda} = \frac{1}{2}\left(\frac{m}{\ell}\right)^{\Delta} \cdot \binom{m-(d-k)}{\ell} = \frac{1}{2}\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k} \cdot \binom{m-(d-k)}{\ell}. \blacktriangleleft
$$

▶ **Lemma 15.** *Let $m, k, d, r, \ell$ and $\mu$ be positive integers such that $\ell + k\mu < \frac{m}{2}$, $\Delta_0 = \delta k$ and $\ell = \frac{m}{2}(1-\varepsilon)$ for some fixed constants $\delta$ and $\varepsilon$. Then, $\Gamma_{k,\ell}(f_{n,\alpha,k}) \geq |\mathcal{M}| \geq \frac{1}{2}\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k} \cdot \binom{m-(d-k)}{\ell}$.*

**Proof.** Recall that $\mathcal{M}$ corresponds to the set $\mathbf{z}^{=\ell} \cdot \sigma_Y\left(\partial^{=k}_{\mathcal{P}_{\Delta_0}} f_{n,\alpha,k}\right)$. Since $\mathcal{M}$ is a bag of multilinear monomials over just the $Z$ variables

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

Since $\partial^{=k}_{\mathcal{P}_{\Delta_0}} f_{n,\alpha,k} \subseteq \partial^{=k} f_{n,\alpha,k}$ and $\mathbf{z}^{=\ell} \subseteq \mathbf{z}^{\leq \ell}$, we get that

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

Thus, $\Gamma_{k,\ell}(f_{n,\alpha,k}) \geq |\mathcal{M}| \geq \frac{1}{2}\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k} \cdot \binom{m-(d-k)}{\ell}$. ◀

We shall now a size lower bound on the depth four multi-$r$-ic circuits of low bottom support that compute $f_{n,\alpha,k}$.

▶ **Theorem 16.** *Let $\delta = 0.25$ and $\varepsilon = 0.8$. Let $n, r, \alpha, k$ and $\mu$ be positive integers such that $r \leq n^{0.2}$, $\mu \leq \frac{\log n}{50}$ and $\alpha = \frac{(2-\delta)\log n}{0.9\delta \log\frac{2}{1-\varepsilon}}$. Let $C$ be a depth four multi-$r$-ic circuit of low bottom support $\mu$ and size $s$. If $C$ computes the polynomial $f_{n,\alpha,k}$ then $s$ must be* $\exp(\Omega(k\log n))$.

**Proof.** Recall that the polynomial $f_{n,\alpha,k}$ is defined on the variable sets $Y$ and $Z$ such that $|Z| = m = 2\alpha nk$. Let $\ell$ be an integer such that $\ell = \frac{m}{2}(1-\varepsilon)$ 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 low bottom support $\mu$ and size $s$. Then it must be the case that $\Gamma_{k,\ell}(f_{n,\alpha,k}) = \Gamma_{k,\ell}(C)$.

$$\begin{aligned}
s &\geq \frac{\frac{1}{2}\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k}\cdot\binom{m-(d-k)}{\ell}}{\binom{\frac{2mr}{\mu}}{k}\cdot\binom{m}{\ell+k\mu}\cdot(\ell+k\mu)}\\
&\geq \frac{1}{2(\ell+k\mu)}\cdot\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k}\cdot\left(\frac{k\mu}{2emr}\right)^{k}\cdot\frac{\binom{m-(d-k)}{\ell}}{\binom{m}{\ell+k\mu}}\\
&= \frac{1}{2(\ell+k\mu)}\cdot\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k}\cdot\left(\frac{k\mu}{2emr}\right)^{k}\cdot\frac{(m-(d-k))!}{m!}\cdot\frac{(m-\ell-k\mu)!}{(m-\ell-(d-k))!}\cdot\frac{(\ell+k\mu)!}{\ell!}\\
&\geq \frac{1}{2(\ell+k\mu)}\cdot\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k}\cdot\left(\frac{k\mu}{2emr}\right)^{k}\cdot\frac{\ell^{k\mu}}{m^{(d-k)}}\cdot(m-\ell)^{(d-k)-k\mu}\\
&= \frac{1}{2(\ell+k\mu)}\cdot\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k}\cdot\left(\frac{k\mu}{2emr}\right)^{k}\cdot\left(\frac{\ell}{m-\ell}\right)^{k\mu}\cdot\left(\frac{m-\ell}{m}\right)^{d-k}\\
&= \frac{1}{2(\ell+k\mu)}\cdot\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k}\cdot\left(\frac{\mu}{4e\alpha nr}\right)^{k}\cdot\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{k\mu}\cdot\left(\frac{1+\varepsilon}{2}\right)^{d-k}\\
&= \frac{1}{2(\ell+k\mu)}\cdot\left(\frac{2}{1-\varepsilon}\right)^{\delta\alpha k}\cdot\left(\frac{\mu}{4e\alpha nr}\right)^{k}\cdot\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{k\mu}\cdot\left(\frac{1+\varepsilon}{2}\right)^{2\alpha k}\\
&= \frac{1}{2(\ell+k\mu)}\cdot\left(\left(\frac{2}{1-\varepsilon}\right)^{\delta}\cdot\left(\frac{1+\varepsilon}{2}\right)^{2}\right)^{\alpha k}\cdot\left(\frac{\mu}{4e\alpha nr}\right)^{k}\cdot\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{k\mu}\\
&= \frac{\exp\left(\alpha k\log\left(\left(\frac{2}{1-\varepsilon}\right)^{\delta}\cdot\left(\frac{1+\varepsilon}{2}\right)^{2}\right) - k\log n - k\log r - k\log\frac{4e\alpha}{\mu} + k\mu\log\frac{1-\varepsilon}{1+\varepsilon}\right)}{2(\ell+k\mu)}\,.
\end{aligned}$$

In the above math block, we use Lemma 5 to simplify the terms along with the fact that $k^2\mu^2 = o(m-\ell)$, $(d-k)^2 = o(m)$ and $k^2\mu^2 = o(\ell)$. To get a meaningful lower bound, we need $\alpha\log\left(\left(\frac{2}{1-\varepsilon}\right)^{\delta}\cdot\left(\frac{1+\varepsilon}{2}\right)^{2}\right)$ to be strictly greater than $(\log n + \log r + \log\frac{4e\alpha}{\mu})$. Let us set $\alpha$ to $\frac{(2-\delta)\log n}{0.9\delta\log\frac{2}{1-\varepsilon}}$. This reduces to showing that there exist constants $\delta$, $\varepsilon$ and $\nu$ such that

$$\frac{(2-\delta)\cdot\log\left(\left(\frac{2}{1-\varepsilon}\right)^{\delta}\cdot\left(\frac{1+\varepsilon}{2}\right)^{2}\right)}{0.9\delta\log\frac{2}{1-\varepsilon}} - 1 \geq \nu. \tag{1}$$

Let us fix the constants as follows: $\varepsilon = 0.8$, $\delta = 0.25$ and $\nu = 0.23$. Through some calculations, it can be verified that Equation 1 gets satisfied. Thus,

$$s \geq \frac{\exp\left(k\left(0.23\log n - \log r - \log\frac{4e\alpha}{\mu} + \mu\log\frac{1-\varepsilon}{1+\varepsilon}\right)\right)}{2(\ell + k\mu)}$$

If $\mu \leq \frac{\log n}{50}$ and $r \leq n^{0.2}$, we get that $s \geq \exp\left(\Omega\left(k\log n\right)\right)$. ◀

## 4 Multi-$r$-ic Depth Four Circuits

To prove the main theorem, we also need the following lemma.

▶ **Lemma 17** (Analogous to Lemma 20.4, [29]). *Let $P$ be a multi-$r$-ic polynomial that is computed by a syntactically multi-$r$-ic depth 4 circuit $C$ of size $s \leq N^{\gamma\mu}$ for some $\gamma > 0$ . Let $\rho$ be a random restriction that sets each variable to zero independently with probability $(1 - N^{-2\gamma})$ . Then with probability at least $(1 - N^{-\gamma\mu})$ the polynomial $\rho(P)$ is computed by a multi-$r$-ic depth four circuit $C'$ of bottom support at most $\mu$ and size $s$.*

**Proof of Theorem 1.** Let $n$ be a large positive integer. Let us set some relevant parameters in terms of $n$ or otherwise as follows.

$$\mu = \frac{\log n}{50}, \qquad \varepsilon = 0.8,$$

$$\delta = 0.25, \qquad \alpha = \frac{(2-\delta)\log n}{0.9\delta\log\frac{2}{1-\varepsilon}},$$

$$k = \frac{c}{100}\log n \qquad N_0 = k(n^2 + 2\alpha n),$$

$$N = N_0^{2+c}\log N_0, \qquad \gamma \text{ is a small constant such that } N^{2\gamma} \approxeq N_0^c.$$

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. Let $\rho : \hat{X} \mapsto \{0, *\}$ be a random restriction such that a variable is set to zero with a probability of $(1 - N^{-2\gamma})$, and is left untouched otherwise. Let $C$ be a syntactically multi-$r$-ic depth four circuit of size $s \leq N^{\gamma\mu}$ that computes $P_{n,\alpha,k}$. Lemma 17 tells us that $C' = \rho(C)$ is a multi-$r$-ic depth four circuit of size $s$ and bottom support at most $\mu$ 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 $s$.

Lemma 7 tells us that $f_{n,\alpha,k}$ is a $p$-projection of $\rho(P_{n,\alpha,k})$ with a probability of $(1 - 2^{-N_0})$ and hence $f_{n,\alpha,k}$ also has a multi-$r$-ic $\Sigma\Pi\Sigma\Pi^{\{\mu\}}$ of size at most $s$ with a probability of at least $(1 - N^{-\gamma\mu} - 2^{-N_0})$. In other words, there is a random restriction $\sigma$ such that $f_{n,\alpha,k}$ is a $p$-projection of $\rho(P_{n,\alpha,k})$ and $C' = \rho(C)$ is a multi-$r$-ic $\Sigma\Pi\Sigma\Pi^{\{\mu\}}$ circuit.

On the other hand, from Theorem 16 we know that any multi-$r$-ic $\Sigma\Pi\Sigma\Pi^{\{\mu\}}$ circuit that computes $f_{n,\alpha,k}$ must be of size $\exp(\Omega(k\log n))$. Thus, it must be that $\exp(\Omega(k\log n)) \leq s \leq N^{\gamma\mu}$. We can choose $c$ to be a small enough constant such that the aforementioned expression is satisfied. Thus, $s$ must at least be $\exp\left(\Omega(\log^2 n)\right)$. The explicit polynomial $Q_n$ is $P_{n,\alpha,k}$ where $\alpha = \frac{(2-\delta)\log n}{0.9\delta\log\frac{2}{1-\varepsilon}}$ and $k = \frac{c}{100}\log n$. ◀

## References

1 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.

2 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.

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

4 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.

5 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.

6 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.

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

8 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.

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

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

11 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.

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

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

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

15 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.

16 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.

17 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. doi:10.4230/LIPIcs.STACS.2016.46.

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

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

20 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. doi:10.4086/toc.2018.v014a016.

21 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.

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

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

24 Ran Raz. Multilinear-NC$^2$ $\neq$ multilinear-NC$^1$. In *proceedings of Foundations of Computer Science (FOCS)*, pages 344–351, 2004. doi:10.1109/FOCS.2004.42.

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

26 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.

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

28 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.

29 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/.

30 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. doi:10.1561/0400000039.

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

32 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.

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

## A Missing Proofs

Proof of Claim 10. There are $n^{2k}$ elements in $\mathcal{P}$. It is easy to see that the volume of a Hamming Ball of radius $\Delta_0$ for vectors of length $2k$ is at most $\sum_{i=0}^{\Delta_0}\binom{2k}{i}\cdot n^i \leq \Delta_0\binom{2k}{\Delta_0}n^{\Delta_0}$ and thus there are at most $\binom{2k}{\Delta_0}n^{\Delta_0}$ many vectors $(\mathbf{a},\mathbf{b})$ in that Hamming ball. Thus, there exists a packing of these Hamming balls in $\mathcal{P}$ with at least $\frac{n^{2k-\Delta_0}}{\Delta_0\binom{2k}{\Delta_0}}$ many balls. $\triangleleft$

Proof of Claim 11. For a vector $(\mathbf{a},\mathbf{b}) \in [n]^{2k}$, $\frac{\partial^k f_{n,\alpha,k}}{y^{(1)}_{a_1,b_1} y^{(2)}_{a_2,b_2} \cdots y^{(k)}_{a_k,b_k}} = \prod_{i=1}^{k}\prod_{v\in[\alpha]} z^{i,1}_{a_i,v}\cdot z^{i,1}_{v+\alpha,b_i}$. For all $i \in [k]$, let $B_i(\mathbf{a},\mathbf{b}) = \prod_{v\in[\alpha]} z^{i,1}_{a_i,v}\cdot z^{i,1}_{v+\alpha,b_i}$. Note that for some $i \in [k]$, if $a_i \neq a'_i$, $\mathrm{dist}(B_i(\mathbf{a},\mathbf{b}),(\mathbf{a}',\mathbf{b}'))$ is at least $\alpha$. Similar is the case when $b_i \neq b'_i$. Thus, if $\mathrm{dist}((\mathbf{a},\mathbf{b}),(\mathbf{a}',\mathbf{b}')) \geq \Delta_0$), there are at least $\Delta_0$ many locations where either $a_i \neq a'_i$ or $b_i \neq b'_i$ and hence $\mathrm{dist}\left(\partial^k_{(\mathbf{a},\mathbf{b})} f_{n,\alpha,k}, \partial^k_{(\mathbf{a}',\mathbf{b}')} f_{n,\alpha,k}\right) \geq \alpha\Delta_0$. $\triangleleft$

The following proofs are a step by step adaptation, rather a replication of proofs Lemma 20.4 and Lemma 20.5 respectively in [29].

**Proof of Lemma 17.** Let $C$ be a multi-$r$-ic depth four circuit of size $s$. Let $m_1, m_2, \cdots, m_t$ be the set of monomials computed at the lower product gate of $C$, that have at least $\mu$ distinct variables in their support. Note that $t$ is at most $s$. For all $i \in [t]$, $\Pr[\rho(m_i) \neq 0] \leq N^{-2\gamma\mu}$. By taking an union bound, the probability that there exists in a monomial amongst

$m_1, m_2, \cdots, m_t$ which is not set to 0 by $\rho$ is at most $t \cdot N^{-2\gamma\mu} \leq s \cdot N^{-2\gamma\mu} \leq N^{-\gamma\mu}$. Thus with a probability of at least $(1 - N^{-\gamma\mu})$, all the monomials at the bottom product gate have at most $\mu$ distinct variables in their support. ◀

**Proof of Lemma 7.** For all $i \in [N_0]$,

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

By union bound, 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 $\frac{1}{2^{N0}}$. Thus, with a probability of at least $(1 - 2^{-N_0})$, 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. ◀