An almost Cubic Lower Bound for Depth Three Arithmetic Circuits

Neeraj Kayal
Microsoft Research India
neeraka@microsoft.com

Chandan Saha
Indian Institute of Science
chandan@csa.iisc.ernet.in

Sébastien Tavenas
Microsoft Research India
sebastien.tavenas@free.fr

Abstract

We show an $\Omega\left(\frac{n^3}{(\ln n)^3}\right)$ lower bound on the number of gates of any depth three ($\Sigma\Pi\Sigma$) arithmetic circuit computing an explicit multilinear polynomial in n variables over any field. This improves upon the previously known quadratic lower bound by Shpilka and Wigderson [SW99, SW01].

1 Introduction

An arithmetic circuit is a directed acyclic graph with leaves (nodes with in-degree zero) labeled by formal variables and other nodes labeled by addition (+) or multiplication ($\times$) operations. Nodes with out-degree zero are the output nodes; for simplicity and without losing generality we will assume that there is only one output node in a circuit. Non-leaf nodes are also referred to as addition or multiplication *gates*. Such a circuit naturally represents a multivariate polynomial; we say this polynomial is *computed* at the output node of the circuit (or simply computed by the circuit). Two parameters that determine the complexity of a circuit are its *size* and *depth*, which are respectively the number of edges and the length of the longest path from any input node to the output node of the underlying directed acyclic graph. Computations involving arithmetic operations can be naturally modeled by arithmetic circuits and hence study of these objects forms a fundamental aspect of complexity theory.

Research on arithmetic circuits received a great impetus from the seminal paper by Valiant [Val79] who defined two non-uniform complexity classes that are algebraic analogues of classes P and NP. These algebraic complexity classes are known as VP and VNP in the literature. Class VP consists of families of polynomials $\{g_n\}_{n\geq 1}$ such that the number of variables and the degree of g_n are $n^{O(1)}$, and there is an arithmetic circuit of size $n^{O(1)}$ computing g_n. A family of polynomials $\{f_n\}_{n\geq 1}$ is in VNP if there is another family of polynomials $\{g_n(\mathbf{x},\mathbf{y})\}_{n\geq 1}$ in VP such that $f_n = \sum_{\mathbf{y}\in\{0,1\}^{|\mathbf{y}|}} g_n(\mathbf{x},\mathbf{y})$. Valiant defined a notion of completeness for the classes VNP and VP, and showed that the family of permanent polynomials is VNP-complete whereas the family of determinant polynomials is *almost* complete for VP. This gave rise to the famous 'determinantal complexity of the permanent' problem, a suitable resolution of which would imply $\mathsf{VP} \neq \mathsf{VNP}$ or equivalently a super-polynomial size lower bound for arithmetic circuits. We refer the reader to the surveys [Mah13, SY10], the book [Bü00] and the paper [MP08] for more on these and other related algebraic complexity classes, their inter-relationships and their associations with Boolean complexity classes. Throughout this article, whenever we use the term 'circuit(s)' we will mean 'arithmetic circuit(s)'.

Starting with Valiant's work there has been significant progress in proving lower bounds for several restricted models of arithmetic circuits. Multilinear [Raz09, Raz06, RY08], noncommutative [Nis91, LMS15], monotone [JS82] and special low-depth circuits [NW96, SS97, GK98, SW01, RY09, Raz10a, ASSS12, KLSS14, KS15a, KS15c, KST15] are examples of such interesting circuit classes. But still, our knowledge of general circuit lower bound is rather limited. The best known size lower bound for general circuits is Baur and Strassen's $\Omega(n \log d)$ bound [Str73, BS83] for circuits computing the simple polynomial $\sum_{i\in[n]} x_i^d$. A recent line of work on depth reduction, starting with [AV08, VSBR83] and culminating with [Koi12, GKKS13a, Tav13], has shown that a moderately strong lower bound for circuits of *depth three*[1] implies a super-polynomial lower bound for general circuits. Also, Raz [Raz10b] showed that a strong enough lower bound for a special kind of (namely, set-multilinear) depth three circuits implies a super-polynomial lower bound for general arithmetic formulas[2]. These depth reduction results have opened up the possibility of proving a super-polynomial lower bound for general circuits/formulas by first proving strong lower bounds for low-depth, in particular depth three, circuits. The hope is depth three circuits, which have an apparent simple structure, might be more amenable to lower bound proofs. But, unfortunately, even at depth three we do not know of any super-polynomial lower bound over fields of characteristic zero!

Depth three circuits. In this paper, whenever we mention a depth three circuit we will mean a $\Sigma\Pi\Sigma$-circuit that has an addition gate at the top, followed by a layer of multiplication gates and finally a bottom layer of sum gates. Such a circuit is a "sum of product of affine polynomials" representation of the computed polynomial. The fan-in of the top addition gate is called the top fan-in, and that of the bottom layer of addition gates the bottom fan-in of the circuit. Observe that bottom fan-in can be at most $n + 1$ where n is the number of variables. The *multiplicative complexity* of a depth three circuit $\mathtt{C}$ is the sum of the fan-ins of the multiplication gates of the circuit, i.e. if $\mathtt{C} = \sum_{i=1}^{s} l_{i1} \cdots l_{id_i}$ where l_{ij}'s are affine polynomials then multiplicative complexity of $\mathtt{C}$ is $\sum_i^s d_i$. It is easy to see that multiplicative complexity is less than the number of edges of a depth three circuit. Circuit $\mathtt{C}$ is *homogeneous* if l_{ij}'s are homogeneous linear polynomials (a.k.a. linear forms).

Previous works on depth three circuit lower bound. In [SW99, SW01], Shpilka and Wigderson proved an $\Omega(n^2)$ lower bound on the multiplicative complexity of depth three circuits computing the elementary symmetric polynomial $\mathrm{ESym}_n^d(x_1, \ldots, x_n) \stackrel{\text{def}}{=} \sum_{S\subseteq[n],|S|=d} \prod_{i\in S} x_i$ on n-variables and degree $d = \Theta(n)$. This bound is essentially optimal for fields of size more than n, as n-variate elementary symmetric polynomials can be computed by depth three circuits with multiplicative complexity $O(n^2)$ [3]. A similar tight quadratic lower bound on the multiplicative complexity but for the power symmetric polynomial $\sum_{i\in[n]} x_i^n$ was shown in [JR07]. Also, a near quadratic lower bound is known for the determinant polynomial [SW99, SW01]. The situation is a lot better over small fields or under the restriction of homogeneity. An exponential size lower bound was shown by [GK98] (and by [GR98]) for depth three circuits over any fixed finite field computing the determinant polynomial (even if the circuit and the determinant are treated in the algebra of functions over the finite field). It was shown in [NW96] that any homogeneous depth three circuit computing ESym_n^{2d} has size $\Omega((n/4d)^d)$. Recently, [KS15a] showed a lower bound of $n^{\Omega(\sqrt{d})}$ for depth three

[1]over fields of characteristic zero

[2]a *formula* is a circuit whose underlying directed acyclic graph is a tree

[3]this follows from an interpolation trick attributed to Michael Ben-Or in [NW96]

circuits, with bottom fan-in bounded by n^ε for any fixed $\varepsilon < 1$, computing an explicit n-variate polynomial of degree d.

1.1 Our results

Theorem 1. (Depth three circuit lower bound on the multiplicative complexity) *There is a family of homogeneous multilinear polynomials* $\{f_n\}_{n\geq 1}$ *in* VNP*, where* f_n *is a* $\Theta(n)$*-variate polynomial of degree* $\Theta(n)$ *such that any depth three circuit computing* f_n *has multiplicative complexity (and hence size)* $\Omega\left(\frac{n^3}{(\ln n)^2}\right)$.

Theorem 1 can be seen as an improvement in the state of the art of the long-standing quadratic lower bound for depth three circuits [SW99, SW01], although our target multilinear polynomial family is harder – it is in VNP and not known to be in VP. Also, from our analysis, we arrive at a near quadratic lower bound for the *symmetric model* defined in [Shp01] thereby improving upon the linear bound therein (Theorem 2).

Let ESym_m^d be an elementary symmetric polynomial in m variables and of degree d. Borrowing terminologies from [Shp01], a *symmetric circuit* has a bottom layer of addition gates computing affine polynomials, and a top gate that computes some elementary symmetric polynomial on the affine polynomials computed at the bottom level gates. In other words, a symmetric circuit is an affine projection of some elementary symmetric polynomial. Such a circuit with m bottom level gates outputs a polynomial of the form $\mathrm{ESym}_m^d(l_1, \ldots, l_m)$ for some d, where $l_1, \ldots, l_m$ are affine polynomials computed by the m bottom level gates. The parameter m is defined as the *size* of the symmetric circuit. This model was shown to be complete or universal in [Shp01] (i.e. every polynomial can be computed in this model), and linear lower bounds were shown on the size of the smallest symmetric circuit computing the determinant polynomial and the polynomial $\prod_{i=1}^{n/2} x_i + \prod_{i=n/2+1}^{n} x_i$. The following theorem improves this lower bound but once again the target polynomial family is likely harder than the ones studied in [Shp01].

Theorem 2. (Symmetric circuit lower bound) *Let* $\{f_n\}_{n\geq 1}$ *be the polynomial family of Theorem 1. The size of the smallest symmetric circuit computing* f_n *is* $\Omega\left(\frac{n^2}{(\ln n)^2}\right)$ *over any infinite field.*

The multiplicative complexity is a lower bound for the number of edges. However, in a $\Sigma\Pi\Sigma$-circuit an addition gate from the bottom layer can be reused many times, and so the number of gates can be much smaller than the multiplicative complexity. For instance, the polynomial $g = \sum_{i\in[n]} x_i^n$ can be computed by a $\Sigma\Pi\Sigma$-circuit with $\Theta(n)$ gates, but any $\Sigma\Pi\Sigma$-circuit computing g has multiplicative complexity $\Omega(n^2)$ [JR07] and any general circuit doing the same has $\Omega(n \log n)$ edges [Str73, BS83]. To the knowledge of the authors, no super-linear lower bound was shown for the number of gates of a general $\Sigma\Pi\Sigma$-circuit over a field of characteristic zero. We prove that the almost cubic lower bound of Theorem 1 can also be obtained for the number of gates of $\Sigma\Pi\Sigma$-circuits by paying an extra logarithmic factor.

Theorem 3. (Depth three circuit lower bound on the number of gates) *There is a family of homogeneous multilinear polynomials* $\{f_n\}_{n\geq 1}$ *in* VNP*, where* f_n *is a* $\Theta(n)$*-variate polynomial of degree* $\Theta(n)$ *such that any depth three circuit computing* f_n *has* $\Omega\left(\frac{n^3}{(\ln n)^3}\right)$ *gates.*

In an attempt to make progress in understanding lower bounds for circuit models where formal degree of the circuit is much higher than the number of variables (as might be the case for a depth three circuit), [KS15b] posed the problem of proving lower bounds for homogeneous depth three circuits with formal degree much larger than the number of variables. The following theorem gives a solution to this problem.

Theorem 4. (Homogeneous depth three circuits with high degree) *For any positive integer $d = d(n) \geq n$, there exists an explicit family $\{f_{n,d}\}$ of n-variate polynomials of degree d such that any homogeneous depth three circuit computing $f_{n,d}$ must have size at least $2^{\Omega(n)}$. Moreover, one can even choose such a family $f_{n,d}$ so that it can in fact be computed by a $(nd)^{O(1)}$-sized algebraic branching program[4].*

The above theorem can be viewed as a generalization of the lower bound by [NW96] for homogeneous depth three circuits. Since elementary symmetric polynomials in n-variables have degree at most n, the lower bound in [NW96] holds for homogeneous depth three circuits with degree less than the number of variables. To the best of our knowledge, a lower bound of $(nd)^{\omega(1)}$ for homogeneous depth three circuits with degree d much greater than the number of variables n was not known. Theorem 4 fills in this gap in our understanding as long as $d = 2^{o(n)}$. However, note that the lower bound in the above theorem is independent of d, ideally one should get $d^{\Omega(n)}$ instead of $2^{\Omega(n)}$.

1.2 Proof ideas

Like in many of the previous works, we use a measure $\mu : \mathbb{F}[\mathbf{x}] \to \mathbb{N}$ to capture some 'weakness' of a circuit family as opposed to a 'hard' family of polynomials which leads to a lower bound for the circuit family. In both Theorem 1 and 4, the improvements are achieved by applying the *dimension of the shifted partials* measure, introduced in [Kay12], and used subsequently (at times with certain crucial alterations) in many other recent lower bound results [GKKS13b, KSS14, FLMS14, KLSS14, KS14, KS15a, KS15e, KS15c, KST15, KS15d]. The shifted partials measure is a generalization of the dimension of the partial derivatives measure used previously in [NW96, SW01]. It is quite effective in proving lower bounds for the model of depth four ($\Sigma\Pi\Sigma\Pi$) circuits with formal degree close to the actual degree of the computed polynomial, and somewhat low bottom fan-in [GKKS13b, KSS14]. In fact, all the recent lower bounds (for restricted depth 3 and 4 circuits) obtained using shifted partials 'reduce' to this case of depth four circuits one way or the other. We take a similar route here, but make the crucial observation that a simple "grouping" step in the analysis with shifted partials gives some *leeway to the formal degree* of the circuit and allows it to grow over the actual degree of the computed polynomial. This observation and a careful construction of the target family of polynomials to take advantage of this leeway are the primary sources of improvement of the depth three lower bound.

An immediate hurdle in proving lower bounds for depth three circuits is that the formal degree of the circuit can be much larger than the degree and number of variables of the computed polynomial. The existing proof techniques and measures have had limited success in handling high formal degree circuits [KS15a, KS15d]. To get around this first hurdle, we begin by following the same approach as in [SW01] of pruning the circuit of high degree product gates by going modulo some affine polynomials picked from among the factors of such 'heavy' product gates. This step is exactly

[4] definition of an algebraic branching program can be found in Section 8

(borrowing terminologies from [SW01]) satisfying some affine linear constraints and restricting the circuit to an affine subspace. However, the degree threshold used to define 'heavy' product gates can now be chosen higher than that in [SW01] because of the 'leeway to formal degree' provided by shifted partials. In the pruned or restricted circuit, a simple "grouping" of affine polynomials in every product term of a depth three circuit turns out to be surprisingly effective in handling the remaining product gates. The grouping step transforms a depth three circuit to a depth four circuit with bottom fan-in more than 1, but at the same time brings down the number of factors in every product term. The tradeoff between the bottom fan-in and the number of factors per product term is then analyzed to obtain a suitable upper bound on the shifted partials dimension of a depth three circuit.

If a degree-δ multiplication gate is just a power of an affine form then by pruning this gate we decrease by δ the multiplicative complexity of the circuit, but only by two its number of gates. That is why, to achieve a lower bound on the number of gates instead of the multiplicative complexity (in Theorem 3), we will define in Section 7 a 'heavy' gate as a multiplication gate which is divisible by many distinct affine forms. Then, if the total number of affine forms is sufficiently small then either the number of heavy gates is small, or some affine forms divide a lot of heavy gates. In all cases, we show that we can prune all the heavy gates. Finally, the pruned circuit is a sum of products of few powers of distinct affine forms and we need to bound the shifted partials measure for these circuits.

Finally, in order to maximize the gain and obtain a near cubic bound we need an explicit multilinear polynomial with *degree linear in the number of variables*, and that has close to the maximum possible shifted partials dimension even when restricted to an affine subspace. The polynomial family $\{f_n\}_{n\geq 1}$ in Theorem 1 is a variant of the family of Nisan-Wigderson polynomials used in [KSS14, KLSS14]. A notable difference between the Nisan-Wigderson families used in earlier works and the one used here is that the degree of f_n is linearly related to its number of variables, unlike $d = n^{o(1)}$ in previous works. Although, a greedy construction of a Nisan-Wigderson family can make degree $\Theta(n)$, it is not clear if such a family is in VNP. To ensure both – a VNP family and linear degree – we construct a family by 'composing' two smaller families of Nisan-Wigderson polynomials, one is obtained by a greedy algorithm and the other explicitly defined in [KSS14, KLSS14]. A detailed description of the polynomial family is given in Section 6.

Few more details on the polynomial families. Polynomial f_n in Theorem 1 is homogeneous with three sets of variables $\mathbf{u}, \mathbf{y}, \mathbf{x}$ such that $|\mathbf{u}| = |\mathbf{y}| = |\mathbf{x}| = \frac{10n}{9}$. (To avoid a few ceil and floor notations in the analysis, we shall assume without any loss of generality that n is divisible by $1872 = 9 \cdot 13 \cdot 16$.) Let $\mathbf{u} = \{u_1, \ldots, u_{\frac{10n}{9}}\}, \mathbf{y} = \{y_1, \ldots, y_{\frac{10n}{9}}\}$ and $\mathbf{x} = \{x_1, \ldots, x_{\frac{10n}{9}}\}$. Every monomial of f_n is a product of a $\mathbf{u}$-monomial of degree $d_{\mathbf{u}} = n$, a $\mathbf{y}$-monomial of degree $d_{\mathbf{y}} = \lfloor \ln n \rfloor$, and an $\mathbf{x}$-monomial of degree $d_{\mathbf{x}} \in \left[\frac{2n}{13}, \frac{n}{3}\right]$. Thus the number of variables and the degree of f_n are both $\Theta(n)$. The $\mathbf{x}$ and the $\mathbf{y}$ variables are the primary variables; derivatives of f_n of order $\lfloor \ln n \rfloor$ with respect to the $\mathbf{y}$-variables give rise to $\mathbf{x}$-monomials with large 'pairwise distance' that help estimate the shifted partials dimension of the target polynomial. The $\mathbf{u}$-variables are auxiliary variables which ensure that the measure remains high for the target polynomial even when restricted to an affine subspace.

The polynomial family $\{f_{n,d}\}$ used in Theorem 4 is a simple variant of the multi-r-ic iterated matrix multiplication polynomial family used in [KST15].

1.3 Organization

Sections 3 to 6 are devoted to the proofs of Theorem 1 and 2. We prove Theorem 3 in Section 7 and Theorem 4 in Section 8.

2 Preliminaries

2.1 Basic notations

For any $m \in \mathbb{N}$, the set of natural numbers, the set $\{1, \ldots, m\}$ will be denoted by $[m]$. We will use upper-case letters (like A or S) to denote sets of numbers, calligraphic upper-case letters (like $\mathcal{B}, \mathcal{D}$ or $\mathcal{L}$) to denote sets of polynomials, and bold lower-case letters (like $\mathbf{x}$ or $\mathbf{y}$) to denote sets of variables. When the base ring of polynomials is clear from the context, the ideal generated by a set of polynomials of the ring, say $\mathcal{L}$, will be denoted by $\langle \mathcal{L} \rangle$. A circuit will be denoted using typewriter font, as in $\mathtt{C}$ or $\mathtt{D}$. For a set of numbers $S \subseteq [m]$, $\bar{S}$ will denote the complement of S. Sometimes, we will use the notation $\mathsf{poly}(n)$ to mean $n^{O(1)}$.

2.2 The measure

Although, the results in this paper can be derived using the shifted partials measure as it is in [Kay12], we choose to work with a variant of this measure for better clarity in the analysis. This variant is similar in outlook to the *shifted skewed partials* measure used recently in [KST15], although for our application there is no difference (or skew) between the number of $\mathbf{x}$ and $\mathbf{y}$ variables. Such a skew between $|\mathbf{y}|$ and $|\mathbf{x}|$ was important for the results in [KST15].

Let $A \subset \left[\frac{10n}{9}\right]$ of size $|A| = n$. Let $\mathbf{x}_A = \{x_i : i \in A\}$ and $g(\mathbf{y}, \mathbf{x}_A) \in \mathbb{F}[\mathbf{y}, \mathbf{x}_A]$. For $k, \ell \in \mathbb{N}$, define the measure $\mathrm{SP}_{k,\ell,A} : \mathbb{F}[\mathbf{y}, \mathbf{x}_A] \to \mathbb{N}$ as follows.

$$\mathrm{SP}_{k,\ell,A}(g) \stackrel{\text{def}}{=} \dim(\mathbf{x}_A^{\leq \ell} \cdot \sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k} g)),$$

where $\boldsymbol{\partial}_{\mathbf{y}}^{=k} g$ is the set of all k-th order partial derivatives of g with respect to the $\mathbf{y}$-variables, and $\sigma_{\mathbf{y}} : \mathbb{F}[\mathbf{y}, \mathbf{x}_A] \to \mathbb{F}[\mathbf{x}_A]$ is a map that sets all the $\mathbf{y}$-variables to zero. Naturally, $\sigma_{\mathbf{y}}$ is a homomorphism from $\mathbb{F}[\mathbf{y}, \mathbf{x}_A]$ to $\mathbb{F}[\mathbf{x}_A]$, and $\sigma_{\mathbf{y}}(\mathcal{D})$ is defined by $\{\sigma_{\mathbf{y}}(h) : h \in \mathcal{D}\}$ for any set of polynomials $\mathcal{D} \subseteq \mathbb{F}[\mathbf{y}, \mathbf{x}_A]$. $\mathbf{x}_A^{\leq \ell}$ is the set of all monomials in the $\mathbf{x}_A$-variables of degree ℓ or less. For two sets of polynomials $\mathcal{B}$ and $\mathcal{D}$, $\mathcal{B}.\mathcal{D} \stackrel{\text{def}}{=} \{h_1.h_2 : h_1 \in \mathcal{B} \text{ and } h_2 \in \mathcal{D}\}$, and the dimension of a set of polynomials $\mathcal{D}$ (denoted by $\dim(\mathcal{D})$) is the dimension of the vector space spanned by the polynomials in $\mathcal{D}$ over the field $\mathbb{F}$.

It is worth noting that the above measure (as in [KST15]) can be thought of as a hybrid of the *rank of the partial derivatives matrix* measure of [Nis91] and the shifted partials measure of [Kay12]. The former measure has been refined and used in several other subsequent work, most notably in [Raz09, RY09], and is also identified with the *evaluation dimension* measure in [FS13] over fields of characteristic zero. The following proposition is easy to verify.

Proposition 5. (Sub-additivity) *For any* $k, \ell \in \mathbb{N}$, $\mathbf{x}_A \subseteq \mathbf{x}$ *and* $g_1, g_2 \in \mathbb{F}[\mathbf{y}, \mathbf{x}_A]$,

$$\mathrm{SP}_{k,\ell,A}(g_1 + g_2) \leq \mathrm{SP}_{k,\ell,A}(g_1) + \mathrm{SP}_{k,\ell,A}(g_2).$$

3 Lower bounding the measure for the target polynomial family

We will show that the measure SP (from Section 2.2) is considerably large when applied suitably to the polynomial family $\{f_n\}_{n\geq 1}$. The precise statement is given in the theorem below.

Polynomials restricted to an affine subspace. Let $S \subseteq \left[\frac{10n}{9}\right]$ be a set of size $|S| = \frac{n}{9}$. Let

$$\mathcal{L}_S = \{x_i - h_i\}_{i\in S} \tag{1}$$

be a set of $|\mathcal{L}_S| = |S| = \frac{n}{9}$ affine polynomials in $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]$ such that $h_i \in \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}]$ for every $i \in S$, where $\bar{S} = \left[\frac{10n}{9}\right] \setminus S$.

Denote the ideal of $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]$ generated by the affine polynomials of $\mathcal{L}_S$ by $\langle \mathcal{L}_S \rangle$. For any polynomial $f \in \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]$, let

$$f_{\langle \mathcal{L}_S\rangle} \stackrel{\text{def}}{=} f \mod \langle \mathcal{L}_S \rangle$$

be the image of f in the ring $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]/\langle \mathcal{L}_S \rangle$. Since $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]/\langle \mathcal{L}_S \rangle$ is isomorphic to $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}]$, $f_{\langle \mathcal{L}_S\rangle}$ can be represented by a polynomial in the ring $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}]$; this polynomial is obtained from f by replacing x_i by h_i for every $i \in S$. Hence, we will treat $f_{\langle \mathcal{L}_S\rangle}$ as an element of $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}]$.

Finally, let $f_{\langle \mathcal{L}_S\rangle, \mathbf{u}_S=0}$ be the polynomial obtained from $f_{\langle \mathcal{L}_S\rangle} \in \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}]$ by setting the $\mathbf{u}$-variables to 0/1-values as follows: $u_i = 0$ if $i \in S$, else $u_i = 1$. We will describe the family $\{f_n\}_{n\geq 1}$ and prove the following theorem in Section 6.

Theorem 6. *Let* n *be the parameter that defines the polynomial family* $\{f_n\}_{n\geq 1}$. *Let* $k = \lfloor \ln n \rfloor$, q *be the smallest prime greater or equal to* $\left\lceil \frac{n}{1000 \cdot \ln n} \right\rceil$, $\ell = \left\lfloor \frac{n^2}{32 \cdot k \cdot \ln q} \right\rfloor$. *Then for every set* $S \subseteq \left[\frac{10n}{9}\right]$ *of size* $|S| = \frac{n}{9}$, *and every set of affine polynomials* $\mathcal{L}_S$ *as in Equation 1, and* $f = f_n$,

$$\mathrm{SP}_{k,\ell,\bar{S}}(f_{\langle \mathcal{L}_S\rangle, \mathbf{u}_S=0}) \geq \frac{1}{2} \cdot q^k \cdot \binom{n+\ell}{n}.$$

Let us next show an upper bound of the measure for a depth three circuit and prove Theorem 1.

4 Upper bounding the measure for a depth three circuit

Pruning 'heavy' product gates from a depth three circuit. Let $\mathtt{C} = \sum_{i=1}^{s} T_i$ be a depth three circuit computing $f = f_n$, where T_i is a product term[5] of $\mathtt{C}$. Let c_0 be a constant to be fixed later in the analysis. Then either of the following two cases is obviously true.

- Case 1: The number of product terms of $\mathtt{C}$, with $\mathbf{x}$-degree greater or equal to $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$, is greater than $\frac{n}{9}$.

[5] a product term corresponds to a multiplication gate of $\mathtt{C}$

- Case 2: The number of product terms of $\mathtt{C}$, with $\mathbf{x}$-degree greater or equal to $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$, is less than or equal to $\frac{n}{9}$.

If Case 1 is true then the multiplicative complexity of $\mathtt{C}$ is at least $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor \cdot \frac{n}{9} = \Omega(\frac{n^3}{(\ln n)^2})$ as $d_{\mathbf{x}} \in \left[\frac{2n}{13}, \frac{n}{3}\right]$ and we have nothing to prove in this case. If Case 2 is true then we can find a 'few' affine polynomials such that modulo these the circuit is free of 'heavy' product terms. This is stated formally in the lemma below and the corollary thereafter, and is directly inspired by a similar argument in [SW99, SW01]. However, the threshold chosen to define 'heavy' product gates in [SW99, SW01] is linear in n, whereas the one here has an extra $\frac{d_{\mathbf{x}}}{(\ln n)^2}$ factor that finally accounts for the improvement in the lower bound. As mentioned in Section 1, this is the leeway to the formal degree of the circuit provided by the analysis with shifted partials.

Lemma 7. *Suppose the number of product terms of $\mathtt{C}$, with $\mathbf{x}$-degree greater or equal to $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$, is bounded by $\frac{n}{9}$. Then, there is a set $S \subseteq \left[\frac{10n}{9}\right]$ of size $\frac{n}{9}$ and a set of affine polynomials,*

$$\mathcal{L}_S = \{x_i - h_i\}_{i \in S}, \text{ where } h_i \text{ is an affine polynomial in } \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}] \text{ for every } i \in S,$$

such that $f_{\langle \mathcal{L}_S \rangle} \in \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}]$ is computed by a depth three circuit, say $\mathtt{C}_{\langle \mathcal{L}_S \rangle}$, satisfying the following:

1. *top fan-in of $\mathtt{C}_{\langle \mathcal{L}_S \rangle}$ is upper bounded by the top fan-in of $\mathtt{C}$,*
2. *every product term of $\mathtt{C}_{\langle \mathcal{L}_S \rangle}$ has $\mathbf{x}$-degree upper bounded by $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$.*

The proof of the lemma is relatively straightforward and we defer the proof to Section 4.2.

Corollary 8. *Polynomial $f_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0} \in \mathbb{F}[\mathbf{y}, \mathbf{x}_{\bar{S}}]$ is computed by a depth three circuit, say $\mathtt{C}_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0}$, with top fan-in bounded by the top fan-in of $\mathtt{C}$ and every product term of $\mathtt{C}_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0}$ has $\mathbf{x}$-degree bounded by $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$.*

Let us denote the circuit $\mathtt{C}_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0}$ by $\mathtt{D}$. Let $\mathtt{D} = \sum_{i=1}^{s} P_i$, where a term P_i is a product of affine polynomials in $\mathbb{F}[\mathbf{y}, \mathbf{x}_{\bar{S}}]$. Note that the pruned circuit $\mathtt{D}$ has only $\mathbf{y}$ and $\mathbf{x}_{\bar{S}}$ variables.

4.1 Upper bounding the measure for the pruned circuit D

Lemma 9. *Let $k, \ell \in \mathbb{N}$ be as in Theorem 6, and S be a set as in Lemma 7. Then*

$$\mathrm{SP}_{k,\ell,\bar{S}}(\mathtt{D}) \le s \cdot \binom{\lceil 32 c_0 d_{\mathbf{x}} \rceil}{k} \cdot \binom{n + \ell + kt}{n}, \text{ where } t = \left\lceil \frac{n}{32 \cdot (\ln n)^2} \right\rceil.$$

Proof. By the sub-additive property of the measure (from Proposition 5), it is sufficient to show that

$$\mathrm{SP}_{k,\ell,\bar{S}}(P) \le \binom{\lceil 32 c_0 d_{\mathbf{x}} \rceil}{k} \cdot \binom{n + \ell + kt}{n}, \tag{2}$$

for any product term P of circuit $\mathtt{D}$. Let t be as in the lemma statement. By Corollary 8, $\mathbf{x}$-degree of every product term P is bounded by $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$. Let $P = l_1 \cdots l_w \cdot R(\mathbf{y})$ where every affine polynomial l_j has some $\mathbf{x}_{\bar{S}}$-variable present in it and $R(\mathbf{y}) \in \mathbb{F}[\mathbf{y}]$ is $\mathbf{x}$-free; naturally, $w \le \left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$. Group the affine polynomials $l_1, \ldots, l_w$ (arbitrarily) into blocks of size t, and multiply the affine polynomials within each block. Only one block might have size less than t. After this "grouping", we have

$$P = Q_1 \cdots Q_{\lceil \frac{w}{t} \rceil} \cdot R(\mathbf{y}),$$

where every $Q_j \in \mathbb{F}[\mathbf{y}, \mathbf{x}_{\bar{S}}]$ has $\mathbf{x}$-degree bounded by t. Observe the following.

Observation 10. **i.** *Every element of $\sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k} P)$ is in the $\mathbb{F}$-span of the set,*

$$\left\{ \sigma_{\mathbf{y}} \left(\prod_{j \in W} Q_j \right) \cdot \eta \quad : \quad W \subseteq \left[\left\lceil \frac{w}{t} \right\rceil \right], |W| = \left\lceil \frac{w}{t} \right\rceil - k, \right.$$
$$\left. \text{and } \eta \text{ is a monomial in } \mathbf{x}_{\bar{S}}\text{-variables of degree} \leq kt \right\}.$$

ii. *Hence, every element of $\mathbf{x}_{\bar{S}}^{\leq \ell} \cdot \sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k} P)$ is in the $\mathbb{F}$-span of the set,*

$$\left\{ \sigma_{\mathbf{y}} \left(\prod_{j \in W} Q_j \right) \cdot \eta \quad : \quad W \subseteq \left[\left\lceil \frac{w}{t} \right\rceil \right], |W| = \left\lceil \frac{w}{t} \right\rceil - k, \right.$$
$$\left. \text{and } \eta \text{ is a monomial in } \mathbf{x}_{\bar{S}}\text{-variables of degree} \leq \ell + kt \right\}.$$

Therefore,

$$\mathrm{SP}_{k,\ell,\bar{S}}(P) \leq \binom{\lceil \frac{w}{t} \rceil}{k} \cdot \binom{n + \ell + kt}{n}, \text{ as } |\bar{S}| = n.$$

Now observe that

$$\left\lceil \frac{w}{t} \right\rceil \leq \left\lceil \frac{\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor}{\left\lceil \frac{n}{32 \cdot (\ln n)^2} \right\rceil} \right\rceil \leq \left\lceil \frac{\frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2}}{\frac{n}{32 \cdot (\ln n)^2}} \right\rceil = \lceil 32 c_0 d_{\mathbf{x}} \rceil.$$

This proves the lemma. □

4.2 Proof of Lemma 7: Pruning heavy product gates

For any affine polynomial $l \in \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]$, let $l_{\mathbf{x}=0}$ be the affine polynomial in $\mathbb{F}[\mathbf{u}, \mathbf{y}]$ obtained by setting all the $\mathbf{x}$-variables to zero in l. Let $l(\mathbf{x}) \stackrel{\text{def}}{=} l - l_{\mathbf{x}=0}$, which is a homogeneous linear polynomial (or a linear form) in $\mathbb{F}[\mathbf{x}]$. Focus on the product terms in $\mathtt{C} = \sum_{i=1}^{s} T_i$ that have $\mathbf{x}$-degree greater than or equal to $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$. Let these product terms be $T_1, \ldots, T_m$, where $m \leq \frac{n}{9}$ (as is the premise of the lemma statement).

Let $\mathcal{L} = \{l_1, \ldots, l_{m'}\}$ be a set of affine polynomials in $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]$ such that

(a) for $i \neq j$, l_i and l_j are factors of two distinct product terms T_a and T_b where $a, b \in [m]$,

(b) the linear forms $l_1(\mathbf{x}), \ldots, l_{m'}(\mathbf{x})$ are $\mathbb{F}$-linearly independent, and

(c) $\mathcal{L}$ is *maximal* in the sense that there is no other $\mathcal{L}' \supset \mathcal{L}$ satisfying (a) and (b).

Condition (a) implies that $m' \leq m$. Such a set $\mathcal{L}$ exists and can be constructed greedily by picking at most one affine polynomial from each product term T_i, $i \leq m$, until we can no longer add affine polynomials such that (a) and (b) are simultaneously satisfied. The following observation is easy to verify owing to condition (b).

Observation 11. *We can find a set $S \subseteq \left[\frac{10n}{9}\right]$ of size m' such that there is a basis*

$$\mathcal{L}_S = \{x_i - h_i\}_{i \in S}, \text{ where } h_i \text{ is an affine polynomial in } \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}] \text{ for every } i \in S,$$

of $\text{span}_{\mathbb{F}}\ \mathcal{L}$*. Hence $\langle \mathcal{L} \rangle = \langle \mathcal{L}_S \rangle$ and $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]/\langle \mathcal{L} \rangle = \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]/\langle \mathcal{L}_S \rangle \cong \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}]$.*

Proof. Follows from Gaussian elimination on the coefficient vectors of the linear forms $l_1(\mathbf{x}), \ldots, l_{m'}(\mathbf{x})$. □

The next observation helps complete the proof of the lemma.

Observation 12. *Let T be any product term out of $T_1, \ldots, T_m$. If the set $\mathcal{L}$ contains any affine factor of T then $T_{\langle \mathcal{L} \rangle} = 0$. Otherwise, for every affine polynomial l dividing T, $l_{\langle \mathcal{L} \rangle} \in \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]/\langle \mathcal{L}_S \rangle$ is an affine polynomial in $\mathbb{F}[\mathbf{u}, \mathbf{y}]$ i.e. $l_{\langle \mathcal{L} \rangle}$ is $\mathbf{x}$-free and hence $T_{\langle \mathcal{L} \rangle} \in \mathbb{F}[\mathbf{u}, \mathbf{y}]$ has $\mathbf{x}$-degree zero. Also, for every $T \in \{T_1, \ldots, T_s\}$, $\mathbf{x}$-degree of $T_{\langle \mathcal{L} \rangle}$ is less or equal to $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$.*

Proof. The 'if' part is trivial. To see the 'otherwise' part, observe that $l(\mathbf{x})$ must be linearly dependent on $l_1(\mathbf{x}), \ldots, l_{m'}(\mathbf{x})$ as $\mathcal{L}$ is maximal by condition (c). The 'also' part is also easy to see as $T_{\langle \mathcal{L} \rangle} = T_{\langle \mathcal{L}_S \rangle}$ is obtained from T by replacing x_i by h_i for every $i \in S$. □

From the above observations it follows that $f_{\langle \mathcal{L} \rangle} = f_{\langle \mathcal{L}_S \rangle} \in \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}]$ is computed by a depth three circuit, say $\mathtt{C}_{\langle \mathcal{L}_S \rangle}$, with top fan-in upper bounded by the top fan-in of $\mathtt{C}$ and every product term of $\mathtt{C}_{\langle \mathcal{L}_S \rangle}$ has $\mathbf{x}$-degree less than or equal to $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$.

Finally, the proof of the lemma is complete by observing that if $m' < \frac{n}{9}$, we can pick some more $\mathbf{x}$-variables arbitrarily from $\mathbf{x}_{\bar{S}}$ and include them in $\mathcal{L}_S$ so that $|S|$ becomes exactly $\frac{n}{9}$.

5 Putting together: Proof of Theorem 1

Let $\mathtt{C}$ be a depth three circuit computing f_n. Then, as explained in Section 4, we have two cases to handle. In Case 1, the multiplicative complexity of $\mathtt{C}$ is already $\Omega(\frac{n^3}{(\ln n)^2})$ and we have nothing to prove. Whereas, in Case 2, the circuit can be pruned of heavy product gates so that the polynomial $f_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0} \in \mathbb{F}[\mathbf{y}, \mathbf{x}_{\bar{S}}]$ is computed by a depth three circuit, say $\mathtt{D}$, whose top fan-in is upper bounded by the top fan-in of $\mathtt{C}$ (by Corollary 8). Moreover, every product term of $\mathtt{D}$ has $\mathbf{x}$-degree bounded by $\left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$ so that Lemma 9 is applicable now.

Lemma 13. *In Case 2, the top fan-in of circuit $\mathtt{D}$ (hence also the top fan-in of circuit $\mathtt{C}$) is $\omega(n^3)$.*

Proof. By Theorem 6 and Lemma 9, the top fan-in s of $\mathtt{D}$ can be lower bounded as follows:

$$s \geq \frac{1}{2} \cdot \frac{q^k \cdot \binom{n+\ell}{n}}{\binom{\lceil 32 c_0 d_{\mathbf{x}} \rceil}{k} \cdot \binom{n+\ell+kt}{n}},$$

where $k=\lfloor \ln n\rfloor$, q is the smallest prime greater than or equal to $\lceil \frac{n}{1000\ln n}\rceil$, $\ell=\lfloor \frac{n^2}{32\cdot k\cdot \ln q}\rfloor$, $d_{\mathbf{x}}$ is an integer in $\left[\frac{2n}{13},\frac{n}{3}\right]$, and $t=\left\lceil \frac{n}{32\cdot(\ln n)^2}\right\rceil$. The ratio

$$\begin{aligned}
\frac{\binom{n+\ell}{n}}{\binom{n+\ell+kt}{n}} &= \frac{(n+\ell)!}{(n+\ell+kt)!}\cdot\frac{(\ell+kt)!}{\ell!}\\
&= \frac{(\ell+1)\cdots(\ell+kt)}{(n+\ell+1)\cdots(n+\ell+kt)}\\
&= \frac{1}{(1+\frac{n}{\ell+1})\cdots(1+\frac{n}{\ell+kt})}\\
&\geq \frac{1}{(1+\frac{n}{\ell+1})^{kt}}\\
&\geq e^{-\frac{n}{\ell+1}\cdot kt}\\
&= e^{-\frac{n}{\lfloor\frac{n^2}{32\cdot k\cdot\ln q}\rfloor+1}\cdot\lfloor\ln n\rfloor\cdot\left\lceil\frac{n}{32\cdot(\ln n)^2}\right\rceil}
\end{aligned}$$

Let us analyse the quantity $\frac{n}{\lfloor\frac{n^2}{32\cdot k\cdot\ln q}\rfloor+1}\cdot\lfloor\ln n\rfloor\cdot\left\lceil\frac{n}{32\cdot(\ln n)^2}\right\rceil$.

$$\begin{aligned}
\frac{n}{\lfloor\frac{n^2}{32\cdot k\cdot\ln q}\rfloor+1}\cdot\lfloor\ln n\rfloor\cdot\left\lceil\frac{n}{32\cdot(\ln n)^2}\right\rceil &\leq \frac{n\cdot\ln n\cdot(\frac{n}{32\cdot(\ln n)^2}+1)}{\frac{n^2}{32\cdot k\cdot\ln q}}\\
&= \frac{n\cdot\ln n\cdot\frac{n}{32\cdot(\ln n)^2}\cdot(1+\frac{32\cdot(\ln n)^2}{n})}{\frac{n^2}{32\cdot k\cdot\ln q}}\\
&= \frac{k\cdot\ln q}{\ln n}\cdot(1+\frac{32\cdot(\ln n)^2}{n})\\
&\leq 1.001\cdot\frac{k\cdot\ln q}{\ln n}, \qquad \text{for sufficiently large } n.
\end{aligned}$$

Hence $\binom{n+\ell}{n}/\binom{n+\ell+kt}{n}\geq q^{-\frac{1.001\cdot k}{\ln n}}\geq n^{-1.001}$, as $k\leq\ln n$ and $q\leq n$. Therefore,

$$\begin{aligned}
s &\geq \frac{1}{2}\cdot n^{-1.001}\cdot\frac{q^k}{\binom{\lceil 32c_0d_x\rceil}{k}}\\
&\geq \frac{1}{2}\cdot n^{-1.001}\cdot\left(\frac{qk}{e\cdot\lceil 32c_0d_{\mathbf{x}}\rceil}\right)^k\\
&\geq \frac{1}{2}\cdot n^{-1.001}\cdot\left(\frac{q\cdot(\ln n-1)}{e\cdot(32c_0d_{\mathbf{x}}+1)}\right)^k, \quad \text{as } k=\lfloor\ln n\rfloor\\
&\geq \frac{1}{2}\cdot n^{-1.001}\cdot\left(\frac{\frac{n}{1000\cdot\ln n}\cdot\ln n\cdot(1-\frac{1}{\ln n})}{e\cdot 32c_0\cdot\frac{n}{3}\cdot(1+\frac{1}{32c_0d_{\mathbf{x}}})}\right)^k, \qquad \text{as } q\geq\left\lceil\frac{n}{1000\ln n}\right\rceil \text{ and } d_{\mathbf{x}}\leq\frac{n}{3}\\
&= \frac{1}{2}\cdot n^{-1.001}\cdot\left(\frac{3}{32000\cdot e\cdot c_0}\cdot\frac{1-\frac{1}{\ln n}}{1+\frac{1}{32c_0d_{\mathbf{x}}}}\right)^k
\end{aligned}$$

$$\begin{aligned}
\Rightarrow s &\geq \frac{1}{2} \cdot n^{-1.001} \cdot \left(\frac{3 \cdot 0.99}{32000 \cdot e \cdot c_0}\right)^k, \quad \text{for large enough } n \\
&= \frac{1}{2} \cdot n^{-1.001} \cdot e^{5k}, \quad \text{if we choose } c_0 = \frac{3 \cdot 0.99}{32000 \cdot e^6} \\
&\geq \frac{1}{2e^5} \cdot n^{-1.001} \cdot e^{5 \cdot \ln n} = \frac{1}{2e^5} \cdot n^{3.999} = \omega(n^3).
\end{aligned}$$

□

Thus, in Case 2, the top fan-in of D (and hence C) must be $\omega(n^3)$ and therefore putting Case 1 and 2 together, the multiplicative complexity of C is $\min\{\Omega(\frac{n^3}{(\ln n)^2}), \omega(n^3)\} = \Omega(\frac{n^3}{(\ln n)^2})$ for sufficiently large n.

5.1 Proof of Theorem 2

The proof follows from Lemma 13. Suppose $f = f_n$ is computed by a symmetric circuit where $l_1, \ldots, l_m$ are the bottom level affine polynomials. Naturally, $f = \text{ESYM}_m^d(l_1, \ldots, l_m)$ for some d, and hence (by Ben-Or's interpolation trick over any field of size more than m) f is also computed by a depth three circuit C with top fan-in $m+1$ and degree of every product term bounded by m. If $m \geq \left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$ then we have nothing to prove. Suppose $m < \left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$. Then the condition of Case 2 (in Section 4) is satisfied as *every* product term of C has $\mathbf{x}$-degree (in fact, total degree) bounded by $m < \left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor$. But then, Lemma 13 tells us that C has top fan-in $\omega(n^3)$ which contradicts with the fact that the top fan-in is $m + 1 = O(n^2)$. So, it must be that $m \geq \left\lfloor \frac{c_0 n d_{\mathbf{x}}}{(\ln n)^2} \right\rfloor = \Omega\left(\frac{n^2}{(\ln n)^2}\right)$.

6 The polynomial family and proof of Theorem 6

6.1 Construction of the Nisan-Wigderson polynomial family

Let $\mathbf{z} = \{z_1, \ldots, z_n\}$ be a set of n formal variables. For any two multilinear monomials m_1 and m_2 in the $\mathbf{z}$-variables of degree $d_{\mathbf{z}}$ each, let $|m_1 \cap m_2|$ be the number of variables common between m_1 and m_2. Define *distance* between the monomials m_1, m_2 as,

$$\Delta(m_1, m_2) \stackrel{\text{def}}{=} d_{\mathbf{z}} - |m_1 \cap m_2|.$$

As in the statement of Theorem 6, let q be the smallest prime greater or equal to $\lceil \frac{n}{1000 \ln n} \rceil$ and $k = \lfloor \ln n \rfloor$. The following lemma plays a central role in the construction of the polynomial family $\{f_n\}_{n \geq 1}$. We will prove it in Section 6.2.

Lemma 14. *There is a family of polynomials $\{g_n(\mathbf{z})\}_{n \geq 1}$ in* VNP *such that $g_n(\mathbf{z})$ is a homogeneous multilinear polynomial of degree $d_{\mathbf{z}} \in \left[\frac{2n}{13}, \frac{n}{3}\right]$ in n $\mathbf{z}$-variables, and $\Delta(m_1, m_2) \geq \frac{n}{16}$ for any pair of distinct monomials m_1 and m_2 of g_n. Further, g_n is a sum of q^k distinct monomials.*

The family $\{f_n\}_{n \geq 1}$. Let $(m_1, \ldots, m_{q^k})$ be an ordered sequence of monomials of the polynomial $g_n(\mathbf{z})$ from the above lemma under lexicographic monomial ordering $z_1 \succ \ldots \succ z_n$. Let $\mathbf{w} = \{w_1, \ldots, w_n\}$ be n formal variables different from $\mathbf{z}$. The number of multilinear monomials in $\mathbf{w}$-variables of degree k is $\binom{n}{k} \geq (\frac{n}{k})^k = \left(\frac{n}{\lfloor \ln n \rfloor}\right)^k \geq q^k$. Under lexicographic monomial ordering

$w_1 \succ \ldots \succ w_n$, let $(\beta_1, \ldots, \beta_{q^k})$ be the ordered sequence of the first q^k monomials among all multilinear monomials in the $\mathbf{w}$-variables of degree k. Define the polynomial $F_n(\mathbf{w}, \mathbf{z})$ as,

$$F_n(\mathbf{w}, \mathbf{z}) \stackrel{\text{def}}{=} \sum_{j=1}^{q^k} \beta_j m_j. \tag{3}$$

Now let $\mathbf{u} = \{u_1, \ldots, u_{\frac{10n}{9}}\}$, $\mathbf{y} = \{y_1, \ldots, y_{\frac{10n}{9}}\}$ and $\mathbf{x} = \{x_1, \ldots, x_{\frac{10n}{9}}\}$ be the sets of variables on which $f_n(\mathbf{u}, \mathbf{y}, \mathbf{x})$ is defined as follows.

$$f_n(\mathbf{u}, \mathbf{y}, \mathbf{x}) \stackrel{\text{def}}{=} \sum_{\substack{A \subseteq \left[\frac{10n}{9}\right] \\ |A| = n}} \prod_{i \in A} u_i \cdot F_n(\mathbf{y}_A, \mathbf{x}_A). \tag{4}$$

We assume the lexicographic order $x_1 \succ \ldots \succ x_{\frac{10n}{9}}$ and $y_1 \succ \ldots \succ y_{\frac{10n}{9}}$. The polynomial $F_n(\mathbf{y}_A, \mathbf{x}_A)$ is obtained by substituting the $\mathbf{y}_A$-variables $\{y_i : i \in A\}$ in place of the $\mathbf{w}$-variables and $\mathbf{x}_A$-variables $\{x_i : i \in A\}$ in place of the $\mathbf{z}$-variables such that the underlying lexicographic orders, $z_1 \succ \ldots \succ z_n$ and $w_1 \succ \ldots \succ w_n$, are obeyed. Note that $d_{\mathbf{y}} = \deg_{\mathbf{y}} f_n = k$, $d_{\mathbf{u}} = \deg_{\mathbf{u}} f_n = n$ and $d_{\mathbf{x}} = \deg_{\mathbf{x}} f_n = \deg_{\mathbf{z}} g_n = d_{\mathbf{z}} \in \left[\frac{2n}{13}, \frac{n}{3}\right]$. Further, the polynomial family $\{f_n\}_{n \geq 1}$ is in VNP: It would be clear from the proof of Lemma 14 that the computational problem of finding the 'index' of a given monomial in g_n can be solved in $\mathsf{poly}(n)$ time, which in turn implies the coefficient of a given monomial in f_n can be found in $\mathsf{poly}(n)$ time. The *index* of a monomial m in g_n is the position of m in the lexicographically ordered list of q^k monomials of g_n.

6.2 Proof of Lemma 14: Composing two Nisan-Wigderson families

In Lemma 14, we need a family whose monomials are pairwise distant, such that the degree is linear in the number of variables and the family is in VNP. The Nisan-Wigderson polynomial family in [KSS14] is in VNP but its degree is not linear. On the other hand, one can greedily get a family such that the degree is linear but which is not known to be in VNP. We show that by 'composing' these two families one can get both the desired properties.

A "greedy" Nisan-Wigderson family. Let $n_0 = 480 \cdot \lfloor \ln n \rfloor$. Since, $2 \cdot \left\lceil \frac{n}{1000 \ln n} \right\rceil < \left\lfloor \frac{n}{n_0} \right\rfloor$ for sufficiently large n, we can find a prime $q \in \left[\left\lceil \frac{n}{1000 \ln n} \right\rceil, 2 \cdot \left\lceil \frac{n}{1000 \ln n} \right\rceil\right]$ and a collection of disjoint subsets of $\mathbf{z}$-variables, $Z_1, \ldots, Z_q$, such that $|Z_i| = n_0$ for every $i \in [q]$.

Proposition 15. *For every set Z_i, $i \in [q]$, there is a set M_{Z_i} of q multilinear monomials of degree $\frac{n_0}{3}$ each in the Z_i-variables such that for every two distinct monomials γ_1 and γ_2 in M_{Z_i}, $\Delta(\gamma_1, \gamma_2) \geq \frac{2n_0}{15}$. Further, the set M_{Z_i} can be constructed deterministically from Z_i in $\mathsf{poly}(n)$ time.*

Proof. We show that the following greedy procedure (similar to the well-known greedy construction of Nisan-Wigderson combinatorial set-system) works in forming the set M_{Z_i}.

Greedy construction of M_{Z_i}

1. Initialize $M_{Z_i} = \emptyset$.

2. Do until $|M_{Z_i}| = q$

3. Pick the lexicographically smallest multilinear monomial γ of degree $\frac{n_0}{3}$ such that $\gamma \notin M_{Z_i}$ and $\Delta(\gamma, \eta) \geq \frac{2n_0}{15}$ for every monomial $\eta \in M_{Z_i}$. Put γ in M_{Z_i}.

The following claim shows that if $|M_{Z_i}| < q$ then step 3 always succeeds in adding a new monomial to M_{Z_i} in time $\binom{|Z_i|}{n_0/3} = \binom{n_0}{n_0/3} = \mathsf{poly}(n)$ (by exhaustive search), so that the total running time of the above greedy algorithm is also $\mathsf{poly}(n)$.

Claim 16. *Let M_{Z_i} be a set of multilinear monomials of degree $\frac{n_0}{3}$ in the Z_i-variables such that $|M_{Z_i}| < q$. Then there exists a multilinear monomial $\gamma \notin M_{Z_i}$ of degree $\frac{n_0}{3}$ such that $\Delta(\gamma, \eta) \geq \frac{2n_0}{15}$ for every $\eta \in M_{Z_i}$.*

Proof. The proof is a standard application of probabilistic argument. Pick every variable independently from Z_i with probability $\frac{1}{2}$ and multiply them to form a monomial γ. Then $\mathcal{E}[\deg(\gamma)] = \frac{n_0}{2}$. Applying Chernoff bound,

$$\Pr\left[\deg(\gamma) < \frac{n_0}{3}\right] < \frac{1}{n^{13}}.$$

Hence, with probability greater than $1 - \frac{1}{n^{13}}$, $\deg(\gamma) \geq \frac{n_0}{3}$. Let η be any particular existing monomial of degree $\frac{n_0}{3}$ in M_{Z_i}. Recall, $|\gamma \cap \eta|$ denotes the number of common variables between γ and η. Then $\mathcal{E}[|\gamma \cap \eta|] = \frac{n_0}{6}$. By Chernoff bound,

$$\Pr\left[|\gamma \cap \eta| \geq \frac{n_0}{5}\right] < \frac{1}{n^{1.06}}.$$

Applying union bound,

$$\Pr\left[|\gamma \cap \eta| \geq \frac{n_0}{5} \text{ for any } \eta \in M_{Z_i}\right] < \frac{|M_{Z_i}|}{n^{1.06}} < \frac{q}{n^{1.06}}.$$

Thus, with probability greater than $1 - \frac{1}{n^{13}} - \frac{q}{n^{1.06}}$, $\deg(\gamma) \geq \frac{n_0}{3}$ and $|\gamma \cap \eta| < \frac{n_0}{5}$ for every $\eta \in M_{Z_i}$. We can drop some extra variables from γ to make sure that $\deg(\gamma) = \frac{n_0}{3}$. This dropping process does not increase the number of common variables between γ and η. Hence, with probability at least $1 - \frac{1}{n^{13}} - \frac{q}{n^{1.06}}$, $\deg(\gamma) = \frac{n_0}{3}$ and $\Delta(\gamma, \eta) \geq \frac{n_0}{3} - \frac{n_0}{5} = \frac{2n_0}{15}$ for every $\eta \in M_{Z_i}$. Since $q < n$, there exists a monomial γ with the desired properties. □

This proves the proposition. □

By Proposition 15, we have q sets of monomials $M_{Z_1}, \ldots, M_{Z_q}$ on disjoint sets of variables such that each set contains q monomials of degree $\frac{n_0}{3}$ with large pairwise distance. Order the monomials in M_{Z_i} in lexicographic order (following $z_1 \succ \ldots \succ z_n$) and denote the j-th monomial in M_{Z_i} by $\gamma_{ij}(\mathbf{z})$. Observe that $\gamma_{i_1 j_1}(\mathbf{z})$ and $\gamma_{i_2 j_2}(\mathbf{z})$ are variable disjoint for $i_1 \neq i_2$, and $\Delta(\gamma_{ij_1}, \gamma_{ij_2}) \geq \frac{2n_0}{15}$ for $j_1 \neq j_2$.

$$\begin{aligned} M_{Z_1} &= \{\gamma_{11}, \ldots, \gamma_{1q}\} \\ &\vdots \\ M_{Z_q} &= \{\gamma_{q1}, \ldots, \gamma_{qq}\} \end{aligned} \tag{5}$$

Summing the monomials of M_{Z_i}, for any i, gives a polynomial in n_0 variables and of degree $\frac{n_0}{3}$. This naturally gives rise to a family of Nisan-Wigderson polynomials where degree is linearly related to the number of variables.

An explicit Nisan-Wigderson family. Consider the following instance of Nisan-Wigderson polynomials (as defined in [KSS14]). Let $\mathbf{v} = \{v_{ij} : i, j \in [q]\}$ be a set of q^2 formal variables. Identify the elements of the prime field $\mathbb{F}_q$ naturally with $[q]$.

$$\mathsf{NW}_n(\mathbf{v}) \stackrel{\text{def}}{=} \sum_{\substack{h(r)\in\mathbb{F}_q[r] \\ \deg_r(h)<k}} \prod_{i\in[q]} v_{ih(i)}.$$

$\mathsf{NW}_n(\mathbf{v})$ is a polynomial in q^2 variables of degree q.

Composing the two families. Replace v_{ij} by $\gamma_{ij}(\mathbf{z})$ from Equation 5 in $\mathsf{NW}_n(\mathbf{v})$ to get the polynomial $g_n(\mathbf{z})$ mentioned in the statement of Lemma 14.

$$g_n(\mathbf{z}) \stackrel{\text{def}}{=} \sum_{\substack{h(r)\in\mathbb{F}_q[r] \\ \deg_r(h)<k}} \prod_{i\in[q]} \gamma_{ih(i)}(\mathbf{z}).$$

Note that $g_n(\mathbf{z})$ has q^k monomials as NW_n has q^k monomials. Moreover, g_n is multilinear and homogeneous of degree $d_{\mathbf{z}} = q \cdot \frac{n_0}{3}$. It is easy to check that $d_{\mathbf{z}} \in [\frac{2n}{13}, \frac{n}{3}]$ as $q \in \left[\left\lceil \frac{n}{1000 \ln n} \right\rceil, 2 \cdot \left\lceil \frac{n}{1000 \ln n} \right\rceil\right]$.

Claim 17. *For any two distinct monomials m_1, m_2 of $g_n(\mathbf{z})$, $\Delta(m_1, m_2) \geq \frac{n}{16}$.*

Proof. Consider any two distinct monomials $m'_1 = \prod_{i\in[q]} v_{ih_1(i)}$ and $m'_2 = \prod_{i\in[q]} v_{ih_2(i)}$ of $\mathsf{NW}_n(\mathbf{v})$, where $h_1(r), h_2(r) \in \mathbb{F}_q[r]$ are distinct univariate polynomials of degree less than k. There are at least $q - k$ indices, say $\{i_1, \ldots, i_{q-k}\} \in [q]$ such that $v_{i_p h_1(i_p)} \neq v_{i_p h_2(i_p)}$, i.e. $h_1(i_p) \neq h_2(i_p)$ for every $p \in [q-k]$. Thus, by Proposition 15, $\Delta(\gamma_{i_p h_1(i_p)}, \gamma_{i_p h_2(i_p)}) \geq \frac{2n_0}{15}$ for every $p \in [q-k]$. Let $m_1 = \prod_{i\in[q]} \gamma_{ih_1(i)}$ and $m_2 = \prod_{i\in[q]} \gamma_{ih_2(i)}$. Therefore,

$$\begin{aligned}
\Delta(m_1, m_2) &\geq (q-k) \cdot \frac{2n_0}{15} \\
&\geq \left(\left\lceil \frac{n}{1000 \ln n} \right\rceil - \lfloor \ln n \rfloor\right) \cdot \frac{2}{15} \cdot 480 \cdot \lfloor \ln n \rfloor \\
&\geq \left(\frac{n}{1000 \ln n} - \ln n\right) \cdot 64 \cdot (\ln n - 1) \\
&= \left(\frac{8n}{125} - 64 \cdot (\ln n)^2\right) \cdot \left(1 - \frac{1}{\ln n}\right) \\
&\geq \frac{n}{16}, \quad \text{for sufficiently large } n.
\end{aligned}$$

□

Finally, it is not hard to show from the explicit definition of $\mathsf{NW}_n(\mathbf{v})$ that the problem of finding the index of a given monomial in $g_n(\mathbf{z})$ can be solved in $\mathsf{poly}(n)$ time, as the monomial sets in Equation 5 can be constructed a priori in $\mathsf{poly}(n)$ time. This also shows that the family $\{g_n(\mathbf{z})\}_{n\geq 1}$ is in VNP.

6.3 Proof of Theorem 6: The measure on the polynomial family

In this section, we show that the relevant measure is high for the family of polynomials (defined in Equation 4) even when restricted to an affine subspace. As in Section 3 (Equation 1), let $S \subseteq \left[\frac{10n}{9}\right]$ be a set of size $\frac{n}{9}$. Let

$$\mathcal{L}_S = \{x_i - h_i\}_{i \in S}$$

be any set of $\frac{n}{9}$ affine polynomials in $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]$ such that $h_i \in \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}_{\bar{S}}]$ for every $i \in S$, where $\bar{S} = \left[\frac{10n}{9}\right] \setminus S$. Let $f = f_n(\mathbf{u}, \mathbf{y}, \mathbf{x})$ (as defined in Equation 4).

Observation 18. $f_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0} = F_n(\mathbf{y}_{\bar{S}}, \mathbf{x}_{\bar{S}}) \in \mathbb{F}[\mathbf{y}, \mathbf{x}_{\bar{S}}]$.

Proof. The polynomial $f_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0}$ is obtained from f by substituting every x_i by h_i for every $i \in S$, and then setting $u_j = 0$ for every $j \in S$ and $u_j = 1$ otherwise. Since the only $\mathbf{x}$-variables occurring in $F_n(\mathbf{y}_{\bar{S}}, \mathbf{x}_{\bar{S}})$ are from $\mathbf{x}_{\bar{S}}$, it remains untouched by the above substitutions. Finally, the setting of the $\mathbf{u}$-variables retains only $F_n(\mathbf{y}_{\bar{S}}, \mathbf{x}_{\bar{S}})$ from the sum in Equation 4. □

So, we need to show that

$$\mathrm{SP}_{k,\ell,\bar{S}}(F_n(\mathbf{y}_{\bar{S}}, \mathbf{x}_{\bar{S}})) \geq \frac{1}{2} \cdot q^k \cdot \binom{n+\ell}{n}.$$

This part of the argument bears close resemblance to and is inspired by similar arguments in [FLMS14, CM14]. We begin with the following observation.

Observation 19. *The set* $\boldsymbol{\partial}_{\mathbf{y}}^{=k} F_n(\mathbf{y}_{\bar{S}}, \mathbf{x}_{\bar{S}})$ *consists of exactly the monomials of* $g_n(\mathbf{x}_{\bar{S}})$*. Hence,* $\sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k} F_n(\mathbf{y}_{\bar{S}}, \mathbf{x}_{\bar{S}}))$ *also consists of exactly the monomials of* $g_n(\mathbf{x}_{\bar{S}})$*.*

Proof. Follows easily from the definition of the polynomial F_n in Equation 3. □

Reusing notation, let the monomials of $g_n(\mathbf{x}_{\bar{S}})$ be $\{m_1, \ldots, m_{q^k}\}$ – these are monomials in $\mathbf{x}_{\bar{S}}$-variables. By Lemma 14, $\Delta(m_i, m_j) \geq \frac{n}{16}$ for every $i \neq j$ and $\frac{2n}{13} \leq \deg(m_i) \leq \frac{n}{3}$ for every $i \in [q^k]$. Let

$$B_i \stackrel{\text{def}}{=} \mathbf{x}_{\bar{S}}^{\leq \ell} \cdot m_i, \qquad \text{for } i \in [q^k].$$

Then,

$$\begin{aligned} \dim(\mathbf{x}_{\bar{S}}^{\leq \ell} \cdot \sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k} F_n(\mathbf{y}_{\bar{S}}, \mathbf{x}_{\bar{S}}))) &= |B_1 \cup \ldots \cup B_{q^k}| \\ \Rightarrow \mathrm{SP}_{k,\ell,\bar{S}}(F_n(\mathbf{y}_{\bar{S}}, \mathbf{x}_{\bar{S}})) &\geq \sum_{i=1}^{q^k} |B_i| - \frac{1}{2} \cdot \sum_{\substack{i,j \\ i \neq j}} |B_i \cap B_j| \\ &= q^k \cdot \binom{n+\ell}{n} - \frac{1}{2} \cdot \sum_{\substack{i,j \\ i \neq j}} |B_i \cap B_j|, \end{aligned} \tag{6}$$

as $|\bar{S}| = n$ and $|B_i| = \binom{n+\ell}{n}$.

Proposition 20. *For every* $i, j \in [q^k]$ *and* $i \neq j$*,* $|B_i \cap B_j| \leq \binom{n+\ell-n/16}{n}$*.*

Proof. If a monomial m belongs to both B_i and B_j then $m = s_1 \cdot m_i = s_2 \cdot m_j$ where $\deg(s_1), \deg(s_2) \leq \ell$. Since $\Delta(m_i, m_j) \geq n/16$,

$$m = s' \cdot \frac{m_j}{\gcd(m_i, m_j)} \cdot m_i, \qquad \text{where } \deg(s') \leq \ell - \frac{n}{16}.$$

Hence, the number of such monomials m is bounded by $\binom{n+\ell-n/16}{n}$. □

Therefore, by Equation 6,

$$\begin{aligned} \mathrm{SP}_{k,\ell,\bar{S}}(F_n(\mathbf{y}_{\bar{S}}, \mathbf{x}_{\bar{S}})) &\geq q^k \cdot \binom{n+\ell}{n} - \frac{q^{2k}}{2} \cdot \binom{n+\ell-n/16}{n} \\ &\geq \frac{1}{2} \cdot q^k \cdot \binom{n+\ell}{n}, \quad \text{(by the following Claim 21)} \end{aligned}$$

Claim 21. $\binom{n+\ell}{n} / \binom{n+\ell-n/16}{n} \geq q^k$.

Proof.

$$\begin{aligned} \frac{\binom{n+\ell}{n}}{\binom{n+\ell-n/16}{n}} &= \frac{(n+\ell)! \cdot (\ell - \frac{n}{16})!}{(n+\ell-\frac{n}{16})! \cdot \ell!} \\ &= \frac{(n+\ell-\frac{n}{16}+1) \cdot (n+\ell-\frac{n}{16}+2) \cdots (n+\ell-\frac{n}{16}+\frac{n}{16})}{(\ell-\frac{n}{16}+1) \cdot (\ell-\frac{n}{16}+2) \cdots (\ell-\frac{n}{16}+\frac{n}{16})} \\ &= \left(\frac{n}{\ell-\frac{n}{16}+1}+1\right) \cdot \left(\frac{n}{\ell-\frac{n}{16}+2}+1\right) \cdots \left(\frac{n}{\ell-\frac{n}{16}+\frac{n}{16}}+1\right) \end{aligned}$$

$$\begin{aligned} \Rightarrow \frac{\binom{n+\ell}{n}}{\binom{n+\ell-n/16}{n}} &\geq \left(\frac{n}{\ell}+1\right)^{\frac{n}{16}} \\ &\geq e^{\frac{n}{2\ell} \cdot \frac{n}{16}} \qquad (\text{as } \ell = \lfloor n^2/(32 \cdot k \cdot \ln q) \rfloor > n) \\ &= e^{\frac{n^2}{32 \cdot \lfloor \frac{n^2}{32k \ln q} \rfloor}} \\ &\geq e^{\frac{n^2}{32 \cdot \frac{n^2}{32k \ln q}}} = q^k \end{aligned}$$

□

This completes the proof of Theorem 6.

7 Lower bound on the number of gates

In this section, we will prove Theorem 3. Let us recall its statement:

Theorem 3 (restated). (Depth three circuit lower bound on the number of gates) *There is a family of homogeneous multilinear polynomials $\{f_n\}_{n \geq 1}$ in* VNP*, where f_n is a $\Theta(n)$-variate polynomial of degree $\Theta(n)$ such that any depth three circuit computing f_n has $\Omega\left(\frac{n^3}{(\ln n)^3}\right)$ gates.*

To prove this theorem, we will show a lower bound for $\Sigma\Pi^{\{w\}}\Sigma$-circuits defined as follows.

Definition 1. *Let* $\mathtt{D} = \sum_{i=1}^{s} P_i$ *be a depth three circuit computing a polynomial in* $\mathbb{F}[\mathbf{y}, \mathbf{x}]$*. If each* $P_i = l_{i1}^{e_{i1}} l_{i2}^{e_{i2}} \cdots l_{ir_i}^{e_{ir_i}} \cdot R_i(\mathbf{y})$*, where* $r_i \leq w$ *and* l_{ij}*'s are affine forms in* $\mathbb{F}[\mathbf{y}, \mathbf{x}]$ *and* $R_i(\mathbf{y})$ *is a product of affine forms in only* $\mathbf{y}$*-variables, then* $\mathtt{D}$ *is a* $\Sigma\Pi^{\{w\}}\Sigma$*-circuit.*

7.1 Transforming a small $\Sigma\Pi\Sigma$-circuit into a $\Sigma\Pi^{\{w\}}\Sigma$-circuit

Let $\mathtt{C} = \sum_{i=1}^{s} T_i$ be a depth three circuit computing the polynomial $f = f_n$ (defined in Section 6), where T_i is a product term of $\mathtt{C}$. In particular, T_i is of the form $T_i(\mathbf{u}, \mathbf{y}, \mathbf{x}) = l_{i1}^{e_{i1}} l_{i2}^{e_{i2}} \cdots l_{ir_i}^{e_{ir_i}} \cdot R_i$ where R_i is a product of affine forms in $\mathbb{F}[\mathbf{u}, \mathbf{y}]$ and the l_{ij}'s are distinct affine forms in $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]$. We show how to transform, via restrictions, such a circuit into a circuit such that for any product term T_i, the number r_i of distinct affine forms l_{ij} that divide T_i is small. Let us define for all $i \in [s]$

$$\mathcal{L}_i \stackrel{\text{def}}{=} \{l \mid l \text{ is an affine form that depends on at least one } \mathbf{x}\text{-variable, and } l \text{ divides } T_i\}.$$

A product term T_i will be called *heavy* if $|\mathcal{L}_i| \geq \left\lfloor \frac{c_0 n^2}{(\ln n)^2} \right\rfloor = w$ (say), where c_0 is a constant which will be fixed later. Notice that now a heavy product gate does not have the same meaning as in Section 4. In Section 4, a gate was heavy if its $\mathbf{x}$-degree is large; here it is heavy if the number of distinct $\mathbf{x}$-dependent affine forms which divide it is large. Let us also define $H = \{i \mid |\mathcal{L}_i| \geq w\}$ be the set of the indexes of the heavy product terms and $\mathcal{L} = \bigcup_{i \in H} \mathcal{L}_i$ be the set of affine forms appearing inside at least one heavy product term. Then at least one of the following three cases is obviously true.

- Case 1: The number of heavy product terms of $\mathtt{C}$ is large: $|H| > \left\lfloor \frac{c_0 n^3}{30 \cdot (\ln n)^3} \right\rfloor$.
- Case 2: The number of affine forms dividing at least one of the heavy product terms of $\mathtt{C}$ is large: $|\mathcal{L}| > \left\lfloor \frac{c_0 n^3}{30 \cdot (\ln n)^3} \right\rfloor$.
- Case 3: The sizes of both H and $\mathcal{L}$ are small enough: $|H|, |\mathcal{L}| \leq \left\lfloor \frac{c_0 n^3}{30 \cdot (\ln n)^3} \right\rfloor$.

If Case 1 is true then the number of product terms (which is the number of product gates at the second layer of $\mathtt{C}$) is also lower bounded by $\left\lfloor \frac{c_0 n^3}{30 \cdot (\ln n)^3} \right\rfloor$. If Case 2 is true then the number of affine forms computed at the first layer of $\mathtt{C}$ is at least $\left\lfloor \frac{c_0 n^3}{30 \cdot (\ln n)^3} \right\rfloor$. In these two cases, it directly implies that the number of gates of $\mathtt{C}$ is $\Omega\left(\frac{n^3}{(\ln n)^3}\right)$ and we have nothing to prove.

If Case 3 is true then we can find a 'few' affine forms such that modulo these, the circuit is free of 'heavy' product terms. We give an algorithm which finds these affine forms.

Pruning procedure: Given a depth three circuit $\mathtt{C}$ which computes a function $f \in \mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x}]$, we first compute the corresponding sets $\mathcal{L}_i$, H and $\mathcal{L}$. After renumbering the elements of $\mathcal{L} = \{l_1, \ldots, l_p\}$, we define

$$e(j) \stackrel{\text{def}}{=} \# \text{ of indices } i \text{ in } H \text{ such that } T_i \text{ is divisible by } l_j.$$

Then we choose $l = l_j$, an affine form of $\mathcal{L}$ which maximizes $e(j)$. By definition of the sets $\mathcal{L}_i$, l is of the form $l = c \cdot x - h$ where c is a non-zero constant in $\mathbb{F}$, x is a variable from $\mathbf{x}$ and h is an affine form in $\mathbb{F}[\mathbf{u}, \mathbf{y}, \mathbf{x} \setminus \{x\}]$. Let $\mathtt{C}_{\langle\{l\}\rangle}$ be the circuit we get by replacing in $\mathtt{C}$ the variable x by the affine form h/c. So $\mathtt{C}_{\langle\{l\}\rangle}$ computes the function $f_{|x \leftarrow h/c}$ which equals $f_{\langle\{l\}\rangle}$ (as in Observation 11). The procedure outputs $\mathtt{C}_{\langle\{l\}\rangle}$.

Let us remark that during this algorithm, we just modify the gates at the bottom level: each affine form l' is transformed into $l'_{\langle\{l\}\rangle}$. In particular, the cardinals of the sets $\mathcal{L}_i$, H and $\mathcal{L}$ can only decrease at each iteration of this procedure.

Claim 22. *Unless $|\mathcal{L}|$ or $|H| > \left\lfloor \frac{c_0 n^3}{30 \cdot (\ln n)^3} \right\rfloor$, the "Pruning procedure" will kill of all heavy product gates within $n/9$ iterations.*

Proof. Let $\mathcal{L}_i^t$, H^t and $\mathcal{L}^t$ be respectively the sets $\mathcal{L}_i$, H and $\mathcal{L}$ after t iterations of the "Pruning procedure". In particular $H^0 = H$ and $\mathcal{L}^0 = \mathcal{L}$. We reuse symbol and denote the i-th product term of the resulting circuit after every iteration also by T_i. At each iteration t, we define as before

$$e_t(j) \stackrel{\text{def}}{=} \#\text{ of indices } i \text{ in } H^{t-1} \text{ such that } T_i \text{ is divisible by } l_j.$$

Counting the number of affine forms with multiplicity which appear in the heavy product terms,

$$\sum_{l_j \in \mathcal{L}^{t-1}} e_t(j) \geq \left\lfloor \frac{c_0 n^2}{(\ln n)^2} \right\rfloor \cdot \left|H^{t-1}\right|.$$

Thus, by an averaging argument, if l_j is an affine form in $\mathcal{L}^{t-1}$ which maximizes $e_t(j)$, then for sufficiently large n

$$e_t(j) \geq \left\lfloor \frac{c_0 n^2}{(\ln n)^2} \right\rfloor \cdot \frac{\left|H^{t-1}\right|}{\left|\mathcal{L}^{t-1}\right|} \geq \frac{29 \ln n}{n} \cdot \left|H^{t-1}\right| \qquad (\text{since } \left|\mathcal{L}^{t-1}\right| \leq \left|\mathcal{L}^0\right| \leq \left\lfloor \frac{c_0 n^3}{30 \cdot (\ln n)^3} \right\rfloor).$$

Moreover, each one of the $e_t(j)$ corresponding heavy product terms evaluates to zero modulo l_j. Consequently, the number of remaining heavy gate after the t^{th} iteration is bounded by:

$$\left|H^t\right| \leq \left|H^{t-1}\right| - \frac{29 \ln n}{n} \left|H^{t-1}\right| \leq \left|H^{t-1}\right| \left(1 - \frac{29 \ln n}{n}\right).$$

So, after t iterations we obtain

$$\begin{aligned} \left|H^t\right| &\leq \left|H^0\right| \left(1 - \frac{29 \ln n}{n}\right)^t \\ &\leq \left\lfloor \frac{c_0 n^3}{30 \cdot (\ln n)^3} \right\rfloor \cdot e^{-29t \ln n/n} \\ &\leq \left\lfloor \frac{c_0 n^3}{30 \cdot (\ln n)^3} \right\rfloor \cdot \frac{1}{n^{29t/n}}. \end{aligned}$$

Consequently, after $n/9$ iterations $H^{n/9}$ is empty for sufficiently large n. □

As in Lemma 7, at the end of the "Pruning procedure" there will be a set $S \subseteq \left[\frac{10n}{9}\right]$ of size $\frac{n}{9}$ and a set of affine forms $\mathcal{L}_S$ such that the following holds.

Corollary 23. *The polynomial* $f_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0} \in \mathbb{F}[\mathbf{y}, \mathbf{x}_{\bar{S}}]$ *is computed by a depth three circuit, say* $\mathtt{C}_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0}$*, with top fan-in bounded by the top fan-in of* $\mathtt{C}$ *and every product term of* $\mathtt{C}_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0}$ *has at most* w *children (addition gates of the first layer) which depend on* $\mathbf{x}$*-variables.*

Let us denote the circuit $\mathtt{C}_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0}$ by $\mathtt{D}$. Note that this pruned circuit $\mathtt{D}$ has only $\mathbf{y}$ and $\mathbf{x}_{\bar{S}}$ variables, and it is indeed a $\Sigma\Pi^{\{w\}}\Sigma$-circuit, where $w = \left\lfloor \frac{c_0 n^2}{(\ln n)^2} \right\rfloor$.

7.2 Upper bounding the shifted partials measure for a $\Sigma\Pi^{\{w\}}\Sigma$-circuit

Let $\mathtt{D}$ be any $\Sigma\Pi^{\{w\}}\Sigma$-circuit computing a polynomial in $\mathbb{F}[\mathbf{y}, \mathbf{x}_{\bar{S}}]$. More precisely, $\mathtt{D} = \sum_{i=1}^{s} P_i$, where $P_i = l_{i1}^{e_{i1}} \cdot l_{i2}^{e_{i2}} \cdot \ldots \cdot l_{ir_i}^{e_{ir_i}} \cdot R_i(\mathbf{y})$, every l_{ij} is an affine form in $\mathbb{F}[\mathbf{y}, \mathbf{x}_{\bar{S}}]$ and $R_i(\mathbf{y})$ is a product of affine forms in $\mathbb{F}[\mathbf{y}]$. By "padding" every product term P_i with some extra affine forms that equal 1, we can assume without loss of generality that $r_i = w$ for all $i \in [s]$. Let us consider any one product term, say $P = l_1^{e_1} \cdot l_2^{e_2} \cdot \ldots \cdot l_w^{e_w} \cdot R(\mathbf{y})$, and partition the set of affine forms $l_1, \ldots, l_w$ into groups of size at most t:

$$[w] = S_1 \uplus S_2 \uplus \cdots \uplus S_{\lceil \frac{w}{t} \rceil}, \text{ where } \forall i, \ |S_i| \leq t.$$

We obtain by multiplying the powers of affine forms of a part together,

$$P = Q_1 \cdot \cdots \cdot Q_{\lceil \frac{w}{t} \rceil} \cdot R(\mathbf{y}), \text{ where } Q_j = \prod_{p \in S_j} l_p^{e_p}.$$

We will prove the following upper bound on the measure applied on a $\Sigma\Pi^{\{w\}}\Sigma$-circuit $\mathtt{D}$.

Lemma 24. *For any* $k, \ell \in \mathbb{N}$*,*

$$\mathrm{SP}_{k,\ell,\bar{S}}(\mathtt{D}) \leq s \cdot \binom{\lceil 32 c_0 n \rceil + k}{k} \cdot \binom{n + \ell + k \cdot (t-1)}{n}, \text{ where } t = \left\lceil \frac{n}{32 \cdot (\ln n)^2} \right\rceil.$$

To prove the lemma, let us start by studying the derivatives of the polynomials Q_i.

Definition 2. *Let* $Q = l_1^{e_1} l_2^{e_2} \cdots l_t^{e_t}$ *be one of the polynomials* Q_i*. For any integer* $k \geq 0$*, define*

$$Q^{(k)} \stackrel{\text{def}}{=} l_1^{\max(e_1 - k, 0)} \cdot l_2^{\max(e_2 - k, 0)} \cdots l_t^{\max(e_t - k, 0)}.$$

Claim 25. *Let* $\mathbf{z} = \mathbf{y} \cup \mathbf{x}_{\bar{S}}$ *be the variables of* Q*. For any non-negative integer* k*,*

$$\boldsymbol{\partial}_{\mathbf{y}}^{=k} Q \subseteq \mathbb{F}\text{-span}\left\{ \mathbf{z}^{\leq k(t-1)} \cdot Q^{(k)} \right\}.$$

Proof. We will prove the claim by induction on k.

- If $k = 0$ then $\boldsymbol{\partial}_{\mathbf{y}}^{=0} Q = \{Q\} \subseteq \mathbb{F}\text{-span}\left\{\mathbf{z}^{\leq 0} \cdot Q^{(0)}\right\}$.
- Let us assume the claim is true for a given k, we will show that for any monomial Υ in the $\mathbf{y}$-variables of degree $k+1$, we have

$$\frac{\boldsymbol{\partial} Q}{\boldsymbol{\partial} \Upsilon} \in \mathbb{F}\text{-span}\left\{ \mathbf{z}^{\leq (k+1)(t-1)} \cdot Q^{(k+1)} \right\}.$$

Let y be a variable dividing Υ. So $\Upsilon = y \cdot \Upsilon'$ where Υ' is a degree-k monomial on $\mathbf{y}$-variables. By induction hypothesis

$$\frac{\boldsymbol{\partial} Q}{\boldsymbol{\partial} \Upsilon'} = Q^{(k)} \cdot g$$

where g is a polynomial in the $\mathbf{z}$-variables of total degree at most $k(t-1)$. Let us consider

$$J = \{j \in [t] \mid e_j > k\}.$$

As, $Q^{(k)} = \prod_{j \in J} l_j^{e_j - k}$, we get

$$\begin{aligned}
\frac{\boldsymbol{\partial} Q}{\boldsymbol{\partial} \Upsilon} &= \frac{\boldsymbol{\partial}}{\boldsymbol{\partial} y} \left(Q^{(k)} \cdot g \right) \\
&= \frac{\boldsymbol{\partial} g}{\boldsymbol{\partial} y} \cdot \prod_{j \in J} l_j^{e_j - k} + g \cdot \sum_{j \in J} (e_j - k) \frac{\boldsymbol{\partial} l_j}{\boldsymbol{\partial} y} \cdot l_j^{e_j - k - 1} \cdot \prod_{p \in J, p \neq j} l_p^{e_p - k} \\
&= \left(\frac{\boldsymbol{\partial} g}{\boldsymbol{\partial} y} \cdot \prod_{j \in J} l_j + g \cdot \sum_{j \in J} (e_j - k) \frac{\boldsymbol{\partial} l_j}{\boldsymbol{\partial} y} \prod_{p \in J, p \neq j} l_p \right) \cdot \prod_{j \in J} l_j^{e_j - k - 1}.
\end{aligned}$$

Since $|J| \leq t$, the degree of the parenthesis is bounded by $(k+1)(t-1)$. Moreover, $Q^{(k+1)}$ equals $\prod_{j \in J} l_j^{e_j - k - 1}$. That proves the induction.

□

Proof of Lemma 24. By the sub-additivity of the measure (Proposition 5), it is sufficient to prove that for any product term P of circuit $\mathtt{D}$ we have

$$\mathrm{SP}_{k,\ell}(P) \leq \binom{\lceil \frac{w}{t} \rceil + k}{k} \cdot \binom{n + \ell + k \cdot (t-1)}{n}. \tag{7}$$

Let t be as in the lemma statement. Let us recall that we have

$$P = Q_1 \cdots Q_{\lceil \frac{w}{t} \rceil} \cdot R(\mathbf{y}).$$

By letting $m = \lceil \frac{w}{t} \rceil$,

$$\begin{aligned}
\sigma_{\mathbf{y}} \left(\boldsymbol{\partial}_{\mathbf{y}}^{=k} P \right) &\subseteq \mathbb{F}\text{-span} \left\{ \left(\sigma_{\mathbf{y}} (\boldsymbol{\partial}_{\mathbf{y}}^{=k_1} Q_1) \right) \cdot \ldots \cdot \left(\sigma_{\mathbf{y}} (\boldsymbol{\partial}_{\mathbf{y}}^{=k_m} Q_m) \right) \mid k_1 + \ldots + k_m \leq k \right\} \\
&\subseteq \mathbb{F}\text{-span} \left\{ \left(\mathbf{x}_{\bar{S}}^{\leq k_1 (t-1)} \cdot Q_1^{(k_1)} \right) \cdot \ldots \cdot \left(\mathbf{x}_{\bar{S}}^{\leq k_m (t-1)} \cdot Q_m^{(k_m)} \right) \mid k_1 + \ldots + k_m \leq k \right\} \\
&\subseteq \mathbb{F}\text{-span} \left\{ \mathbf{x}_{\bar{S}}^{\leq k(t-1)} \cdot Q_1^{(k_1)} \cdot \ldots \cdot Q_m^{(k_m)} \mid k_1 + \ldots + k_m \leq k \right\}.
\end{aligned}$$

Furthermore, the number of monomials in the $\mathbf{x}_{\bar{S}}$-variables of degree at most $\ell + k(t-1)$ is bounded by $\binom{n+\ell+k(t-1)}{n}$, as $|\bar{S}| = n$, and the number of m-uplet $(k_1, \ldots, k_m) \in \mathbb{N}^m$ such that $k_1 + \ldots + k_m \leq k$ is bounded by $\binom{m+k}{k}$. Finally,

$$\frac{w}{t} \leq \left\lfloor \frac{c_0 n^2}{(\ln n)^2} \right\rfloor \cdot \left(\left\lceil \frac{n}{32 \cdot (\ln n)^2} \right\rceil \right)^{-1} \leq 32 c_0 n.$$

This proves the lemma. □

7.3 Putting together: Proof of Theorem 3

Let $\mathtt{C}$ be a depth three circuit computing f_n. Then, as explained in Section 7.1, we have three cases to handle. In Cases 1 and 2, the number of gates of $\mathtt{C}$ is already $\Omega(\frac{n^3}{(\ln n)^3})$ and we have nothing to prove. Whereas, in Case 3, the circuit can be pruned of heavy product gates so that the polynomial $f_{\langle \mathcal{L}_S \rangle, \mathbf{u}_S = 0} \in \mathbb{F}[\mathbf{y}, \mathbf{x}_{\bar{S}}]$ is computed by a depth three circuit, say $\mathtt{D}$, whose top fan-in is upper bounded by the top fan-in of $\mathtt{C}$ (by Corollary 23). Moreover, every product term of $\mathtt{D}$ has at most $w = \left\lfloor \frac{c_0 n^2}{(\ln n)^2} \right\rfloor$ children which depend on the $\mathbf{x}$-variables. So Lemma 24 is applicable now.

Lemma 26. *In Case 3, the top fan-in of circuit $\mathtt{D}$ (hence also the top fan-in of circuit $\mathtt{C}$) is $\Omega(n^3)$.*

Proof. By Theorem 6 and Lemma 24, the top fan-in s of $\mathtt{D}$ can be lower bounded as follows:

$$s \geq \frac{1}{2} \cdot \frac{q^k \cdot \binom{n+\ell}{n}}{\binom{\lceil 32 c_0 n \rceil + k}{k} \cdot \binom{n+\ell+k(t-1)}{n}},$$

where $k = \lfloor \ln n \rfloor$, q is the smallest prime greater than or equal to $\lceil \frac{n}{1000 \ln n} \rceil$, $\ell = \lfloor \frac{n^2}{32 \cdot k \cdot \ln q} \rfloor$, and $t = \left\lceil \frac{n}{32 \cdot (\ln n)^2} \right\rceil$. Analyzing the above ratio and setting c_0 appropriately as in Lemma 13, it follows that $s = \Omega(n^3)$. □

Thus, in Case 3, the top fan-in of $\mathtt{D}$ (and hence the number of multiplication gates in $\mathtt{C}$) must be $\Omega(n^3)$ and therefore putting the three cases together, the number of gates of $\mathtt{C}$ is $\min\{\Omega(\frac{n^3}{(\ln n)^3}), \Omega(n^3)\} = \Omega(\frac{n^3}{(\ln n)^3})$ for sufficiently large n.

8 Homogeneous depth three circuits with large degree

We prove Theorem 4 in this section. The measure remains the same as before, but the notation is simplified a little bit (as we do not need to include a subset of variables in the definition of the measure). For any polynomial $g \in \mathbb{F}[\mathbf{y}, \mathbf{x}]$, define the measure $\mathrm{SP}_{k,\ell} : \mathbb{F}[\mathbf{y}, \mathbf{x}] \to \mathbb{N}$ as

$$\mathrm{SP}_{k,\ell}(g) \stackrel{\text{def}}{=} \dim(\mathbf{x}^{\leq \ell} \cdot \sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k} g)).$$

Like before, the measure is sub-additive, i.e. for $g_1, g_2 \in \mathbb{F}[\mathbf{y}, \mathbf{x}]$ and $k, \ell \in \mathbb{N}$,

$$\mathrm{SP}_{k,\ell}(g_1 + g_2) \leq \mathrm{SP}_{k,\ell}(g_1) + \mathrm{SP}_{k,\ell}(g_2).$$

Moreover, the measure is invariant under multiplication by any fixed polynomial from $\mathbb{F}[\mathbf{x}]$ (the proof of the following lemma is very simple and is given in Appendix A):

Lemma 27. *For any $g \in \mathbb{F}[\mathbf{y}, \mathbf{x}]$, $h \in \mathbb{F}[\mathbf{x}]$ and $k, \ell \in \mathbb{N}$, $\mathrm{SP}_{k,\ell}(h \cdot g) = \mathrm{SP}_{k,\ell}(g)$.*

The outline of the proof of Theorem 4 also remains the same: we show a suitable upper bound on the measure for the circuit, and a lower bound for the target family of polynomials. The target family of polynomials is basically a *multi-r-ic* variant of the iterated matrix multiplication polynomial defined and analysed in [KST15] – we will recall some parts of the analysis from there to lower bound the measure for the family of polynomials. Furthermore, this polynomial can be computed by an algebraic branching program of size polynomial in the number of variables and degree of the polynomial.

Definition 3 (Algebraic Branching Program)**.** *An Algebraic Branching Program(ABP) in the variables $X = \{x_1, x_2, ..., x_n\}$ is a directed acyclic graph with a source vertex s and a sink vertex t. It has $(d+1)$ sets or layers of vertices $V_1, V_2, ..., V_{d+1}$, where V_1 and V_{d+1} contain only s and t respectively. The width of an ABP is the maximum number of vertices in any of the $(d+1)$ layers. All the edges in an ABP are such that an edge starts from a vertex in V_i and is directed to a vertex in V_{i+1}, where V_i belongs to the set $\{V_1, V_2, ..., V_d\}$. The edges in an ABP are labeled by affine polynomials over a base field $\mathbb{F}$. The weight of the path between any two vertices u and v in an ABP is computed by taking the product of the edge labels on the path from u to v. An ABP computes the sum of the weights of all the paths from s to t.*

8.1 Upper bound for the circuit

Let $\mathtt{C}$ be any homogeneous depth three circuit computing a polynomial in n variables $\mathbf{y} \uplus \mathbf{x}$ and of degree d. More precisely, by identifying the circuit with the polynomial it computes,

$$\mathtt{C} = T_1 + T_2 + \ldots + T_s,$$

where the T_i's are products of d homogeneous linear polynomials i.e. $T_i = l_{i1} \cdot l_{i2} \cdot \ldots \cdot l_{id}$, where every l_{ij} is a linear form. Let us consider any one product term, say T. By grouping t linear forms together and multiplying the linear forms within each group, we obtain

$$T = Q_1 \cdot \cdots \cdot Q_{\lceil \frac{d}{t} \rceil},$$

where $\deg(Q_j) \leq t$ for every $j \in \left[\lceil \frac{d}{t} \rceil\right]$. By sub-additivity of the measure and following a similar argument as in Section 4.1, we get the following lemma.

Lemma 28. *For any $k, \ell \in \mathbb{N}$ and $t \leq d$,*

$$\mathrm{SP}_{k,\ell}(\mathtt{C}) \leq s \cdot \binom{\lceil \frac{d}{t} \rceil}{k} \cdot \binom{|\mathbf{x}| + \ell + kt}{|\mathbf{x}|}. \tag{8}$$

8.2 Lower bound for the polynomial family

The polynomial family. We define a polynomial on n variables $\mathbf{y} \uplus \mathbf{x}$ and of degree d, where d is any integer greater or equal to n.

For $w, k, r, \alpha \in \mathbb{N}$, consider the following polynomial.

$$F_{w,k,r,\alpha}(\mathbf{y}, \mathbf{x}) \stackrel{\text{def}}{=} g_1(\mathbf{y}_1, \mathbf{x}_1) \cdot g_2(\mathbf{y}_2, \mathbf{x}_2) \cdot \ldots \cdot g_k(\mathbf{y}_k, \mathbf{x}_k),$$

where the g_i's are polynomials over the indicated (disjoint) subsets of variables $\mathbf{y} = \mathbf{y}_1 \uplus \ldots \uplus \mathbf{y}_k$ and $\mathbf{x} = \mathbf{x}_1 \uplus \ldots \uplus \mathbf{x}_k$, and defined as,

$$g_i(\mathbf{y}_i, \mathbf{x}_i) \stackrel{\text{def}}{=} \sum_{a,b \in [w]} y_{i,a,b} \cdot \prod_{c \in [\alpha]} x_{i,c,a}^r \cdot x_{i,c+\alpha,b}^r.$$

The number of $\mathbf{y}$-variables is $|\mathbf{y}| = kw^2$ and the number of $\mathbf{x}$-variables is $|\mathbf{x}| = 2k\alpha w$. The total number of variables in $F_{w,k,r,\alpha}$ is $(w^2 + 2\alpha w) \cdot k$, and it has degree $\tilde{d} = (2\alpha r + 1) \cdot k$. Our target polynomial is almost $F_{w,k,r,\alpha}$, except that we multiply it with a suitable power of a variable just to match its degree with the given degree parameter d which is any number more than the number of variables.

Let $n = (w^2 + 2\alpha w) \cdot k$ and $d \geq n$ be a given degree parameter. In the analysis, we eventually fix α and w to integer constants (in Equation 12) so that $n = \Theta(k)$. Set $r = \lceil \frac{d}{3\alpha k} \rceil$ and x be any arbitrarily fixed variable in $\mathbf{x}$. Our polynomial family $\{f_{n,d}\}$ is defined by

$$f_{n,d} \stackrel{\text{def}}{=} x^{d-\tilde{d}} \cdot F_{w,k,r,\alpha}. \tag{9}$$

This polynomial is well defined, i.e. $d \geq \tilde{d}$, as soon as $w \geq 3$. Observe that $f_{n,d}$ has the same set of n variables as $F_{w,k,r,\alpha}$ and has degree d. Let us record the values for k and r for the analysis later.

$$k = \frac{n}{w^2 + 2\alpha w} \quad \text{and} \quad r = \left\lceil \frac{d}{3\alpha k} \right\rceil. \tag{10}$$

Also, note that $f_{n,d}$ can be computed by a $\mathsf{poly}(n, d)$ size ABP.

The measure on the polynomial family. The following lemma was essentially proved in [KST15] (see Section 7.5 in there) with slightly different notations. For completeness, we include a proof in Appendix B.

Lemma 29. *Let $0 < \delta \leq 1/5$ be a constant and $w \geq 3$.*
1) Then

$$\mathrm{SP}_{k,\ell}(F_{w,k,r,\alpha}) \geq M \cdot \binom{|\mathbf{x}| + \ell}{|\mathbf{x}|} - \frac{M^2}{2} \cdot \binom{|\mathbf{x}| + \ell - \lceil \delta k \rceil \cdot \alpha r}{|\mathbf{x}|}, \tag{11}$$

where $M = \left(\left\lfloor \frac{w^{2-\delta}}{2} \right\rfloor \right)^k$.

2) Moreover, if $\ell \geq |\mathbf{x}|$ and $2 \cdot |\mathbf{x}| \cdot \alpha r \geq \ell\beta \cdot \ln w$ where $\beta \geq 4(2-\delta)/\delta$ is a constant then we can also conclude that (11) is lower bounded by $M \cdot \binom{|\mathbf{x}|+\ell}{|\mathbf{x}|}/2$.

For the choice of parameters in Equation (12) below, $\frac{w^{2-\delta}}{2}$ is an integer. Hence, $M = \left(\frac{w^{2-\delta}}{2}\right)^k$.

Corollary 30. *If the conditions of Lemma 29 are satisfied then it follows from Lemma 27 that*

$$\mathrm{SP}_{k,\ell}(f_{n,d}) \geq \frac{M}{2} \cdot \binom{|\mathbf{x}| + \ell}{|\mathbf{x}|}.$$

8.3 Putting together: Proof of Theorem 4

Let us choose

$$t = \lfloor 2\varepsilon\alpha r \rfloor, \quad \text{and} \quad \ell = \left\lfloor \frac{|\mathbf{x}| \cdot t}{\varepsilon\beta \cdot \ln w} \right\rfloor$$

with the following parameters

$$\alpha = 18, \quad \delta = \frac{1}{5}, \quad \beta = 36, \quad \varepsilon = \frac{1}{200}, \quad \text{and} \quad w = 2^{10}. \tag{12}$$

We can notice that $t > 0$ and $\lceil d/t \rceil \leq (2k)/\varepsilon$. Furthermore, the conditions $\ell \geq |\mathbf{x}|$, $2 \cdot |\mathbf{x}| \cdot \alpha r \geq \ell\beta \cdot \ln w$, and $\beta \geq 4(2-\delta)/\delta$ are satisfied. Hence, if C is a homogeneous depth three circuit computing $f_{n,d}$, then by Lemma 28 and Corollary 30,

$$s \cdot \binom{\lceil \frac{d}{t} \rceil}{k} \cdot \binom{|\mathbf{x}| + \ell + kt}{|\mathbf{x}|} \geq \mathrm{SP}_{k,\ell}(f_{n,d}) \geq \frac{M}{2} \cdot \binom{|\mathbf{x}| + \ell}{|\mathbf{x}|}.$$

Consequently,

$$\begin{aligned}
s &\geq \frac{M \cdot \binom{|\mathbf{x}|+\ell}{|\mathbf{x}|}}{2 \cdot \binom{\lceil d/t \rceil}{k} \cdot \binom{|\mathbf{x}|+\ell+kt}{|\mathbf{x}|}} \\
&\geq \frac{M}{2 \cdot \binom{2k/\varepsilon}{k}} \cdot \frac{(\ell+1)\cdots(\ell+tk)}{(|\mathbf{x}|+\ell+1)\cdots(|\mathbf{x}|+\ell+tk)} \\
&= \frac{M}{2 \cdot \binom{400k}{k}} \cdot \frac{1}{(1+\frac{|\mathbf{x}|}{\ell+1})\cdots(1+\frac{|\mathbf{x}|}{\ell+tk})} \\
&\geq \frac{(w^{2-\delta})^k}{2^{k+1} \cdot \binom{400k}{k}} \cdot \frac{1}{(1+\frac{|\mathbf{x}|}{\ell+1})^{tk}} \\
&\geq \frac{(w^{2-\delta})^k}{2^{k+1} \cdot (400e)^k} \cdot e^{-\frac{|\mathbf{x}|}{\ell+1} \cdot tk} \\
&\geq \frac{1}{2} \cdot \left(\frac{w^{2-\delta} \cdot e^{-\frac{|\mathbf{x}|}{\ell+1} \cdot t}}{2200} \right)^k \\
&\geq \frac{1}{2} \cdot \left(\frac{w^{2-\delta-\varepsilon\beta}}{2200} \right)^k \\
&= 2^{\Omega(k)} = 2^{\Omega(n)}.
\end{aligned}$$

This completes the proof of Theorem 4.

References

[ASSS12] Manindra Agrawal, Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. Jacobian hits circuits: hitting-sets, lower bounds for depth-d occur-k formulas & depth-3 transcendence degree-k circuits. In *STOC*, pages 599–614, 2012.

[AV08] Manindra Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 67–75, 2008.

[B0̈0] Peter Bürgisser. *Completeness and Reduction in Algebraic Complexity Theory.* Springer, 2000.

[BS83] W. Baur and V. Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22(3):317–330, 1983.

[CM14] Suryajith Chillara and Partha Mukhopadhyay. Depth-4 lower bounds, determinantal complexity: A unified approach. In *31st International Symposium on Theoretical Aspects of Computer Science (STACS 2014), STACS 2014, March 5-8, 2014, Lyon, France*, pages 239–250, 2014.

[FLMS14] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *STOC*, pages 128–135, 2014.

[FS13] Michael A. Forbes and Amir Shpilka. Quasipolynomial-Time Identity Testing of Non-commutative and Read-Once Oblivious Algebraic Branching Programs. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 243–252, 2013.

[GK98] Dima Grigoriev and Marek Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 577–582, 1998.

[GKKS13a] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth three. In *Foundations of Computer Science (FOCS)*, pages 578–587, 2013.

[GKKS13b] Ankit Gupta, Neeraj Kayal, Pritish Kamath, and Ramprasad Saptharishi. Approaching the chasm at depth four. In *Conference on Computational Complexity (CCC)*, pages 65–73, 2013.

[GR98] Dima Grigoriev and Alexander A. Razborov. Exponential complexity lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. In *39th Annual Symposium on Foundations of Computer Science, FOCS '98, November 8-11, 1998, Palo Alto, California, USA*, pages 269–278, 1998.

[JR07] Maurice J. Jansen and Kenneth W. Regan. "Resistant" Polynomials and Stronger Lower Bounds for Depth-Three Arithmetical Formulas. In *Computing and Combinatorics, 13th Annual International Conference, COCOON 2007, Banff, Canada, July 16-19, 2007, Proceedings*, pages 470–481, 2007.

[JS82] Mark Jerrum and Marc Snir. Some exact complexity results for straight-line computations over semirings. *J. ACM*, 29(3):874–897, 1982.

[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.

[KLSS14] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An Exponential Lower Bound for Homogeneous Depth Four Arithmetic Formulas. In *Foundations of Computer Science (FOCS)*, pages 61–70, 2014.

[Koi12] Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theor. Comput. Sci.*, 448:56–65, 2012.

[KS14] Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 364–373, 2014.

[KS15a] Neeraj Kayal and Chandan Saha. Lower Bounds for Depth Three Arithmetic Circuits with small bottom fanin. In *Conference on Computational Complexity*, pages 158–208, 2015.

[KS15b] Neeraj Kayal and Chandan Saha. Multi-k-ic depth three circuit lower bound. In *32nd International Symposium on Theoretical Aspects of Computer Science (STACS 2015)*, volume 30, pages 527–539, 2015.

[KS15c] Mrinal Kumar and Ramprasad Saptharishi. An exponential lower bound for homogeneous depth-5 circuits over finite fields. *CoRR*, abs/1507.00177, 2015.

[KS15d] Mrinal Kumar and Shubhangi Saraf. Arithmetic circuits with locally low algebraic rank. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:194, 2015.

[KS15e] Mrinal Kumar and Shubhangi Saraf. Sums of products of polynomials in few variables : lower bounds and polynomial identity testing. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:71, 2015.

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

[KST15] Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. On the size of homogeneous and of depth four formulas with low individual degree. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:181, 2015.

[LMS15] Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for non-commutative skew circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:22, 2015.

[Mah13] Meena Mahajan. Algebraic complexity classes. *CoRR*, abs/1307.3863, 2013.

[MP08] Guillaume Malod and Natacha Portier. Characterizing Valiant's algebraic complexity classes. *J. Complex.*, 24(1):16–38, 2008.

[Nis91] Noam Nisan. Lower bounds for non-commutative computation (extended abstract). In *STOC*, pages 410–418, 1991.

[NW96] Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1996.

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

[Raz09] Ran Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size. *J. ACM*, 56(2), 2009.

[Raz10a] Ran Raz. Elusive functions and lower bounds for arithmetic circuits. *Theory of Computing*, 6(1):135–177, 2010.

[Raz10b] Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 659–666, 2010.

[RY08] Ran Raz and Amir Yehudayoff. Balancing syntactically multilinear arithmetic circuits. *Computational Complexity*, 17(4):515–535, 2008.

[RY09] Ran Raz and Amir Yehudayoff. Lower Bounds and Separations for Constant Depth Multilinear Circuits. *Computational Complexity*, 18(2):171–207, 2009.

[Shp01] Amir Shpilka. Affine projections of symmetric polynomials. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity, Chicago, Illinois, USA, June 18-21, 2001*, pages 160–171, 2001.

[SS97] Victor Shoup and Roman Smolensky. Lower Bounds for Polynomial Evaluation and Interpolation Problems. *Computational Complexity*, 6(4):301–311, 1997.

[Str73] V. Strassen. Die Berechnungskomplexität von elementarsymmetrischen Funktionen und von Interpolationskoeffizienten. *Numerische Mathematik*, 20:238251, 1973.

[SW99] Amir Shpilka and Avi Wigderson. Depth-3 arithmetic formulae over fields of characteristic zero. In *IEEE Conference on Computational Complexity*, pages 87–, 1999. Available at http://eccc.hpi-web.de/report/1999/023/.

[SW01] A. Shpilka and A. Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Computational Complexity*, 10(1):1–27, 2001.

[SY10] Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3-4):207–388, 2010.

[Tav13] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *MFCS*, pages 813–824, 2013.

[Val79] L. G. Valiant. Completeness Classes in Algebra. In *STOC '79: Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 249–261, New York, NY, USA, 1979. ACM Press.

[VSBR83] L. G. Valiant, S. Skyum, S. Berkowitz, and C. Rackoff. Fast parallel computation of polynomials using few processors. *SIAM Journal on Computing*, 12(4):641–644, 1983.

A Proof of Lemma 27

Lemma 27 (restated). For any $g \in \mathbb{F}[\mathbf{y}, \mathbf{x}]$, $h \in \mathbb{F}[\mathbf{x}]$ and $k, \ell \in \mathbb{N}$, $\mathrm{SP}_{k,\ell}(h \cdot g) = \mathrm{SP}_{k,\ell}(g)$.

Proof.

$$\begin{aligned}
\mathrm{SP}_{k,\ell}(h \cdot g) &= \dim(\mathbf{x}^{\leq \ell} \cdot \sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k}(h \cdot g))) \\
&= \dim(\mathbf{x}^{\leq \ell} \cdot \sigma_{\mathbf{y}}(h \cdot \boldsymbol{\partial}_{\mathbf{y}}^{=k} g)) \\
&= \dim\left(h \cdot \left(\mathbf{x}^{\leq \ell} \cdot \sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k} g)\right)\right) \\
&= \dim(\mathbf{x}^{\leq \ell} \cdot \sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k} g)) \\
&= \mathrm{SP}_{k,\ell}(g).
\end{aligned}$$

□

B Proof of Lemma 29

Lemma 29 (restated). Let $0 < \delta \leq 1/5$ be a constant and $w \geq 3$.
1) Then

$$\mathrm{SP}_{k,\ell}(F_{w,k,r,\alpha}) \geq M \cdot \binom{|\mathbf{x}| + \ell}{|\mathbf{x}|} - \frac{M^2}{2} \cdot \binom{|\mathbf{x}| + \ell - \lceil \delta k \rceil \cdot \alpha r}{|\mathbf{x}|}, \tag{13}$$

where $M = \left(\left\lfloor \frac{w^{2-\delta}}{2} \right\rfloor\right)^k$.

2) Moreover, if $\ell \geq |\mathbf{x}|$ and $2 \cdot |\mathbf{x}| \cdot \alpha r \geq \ell \beta \cdot \ln w$ where $\beta \geq 4(2-\delta)/\delta$ is a constant then we can also conclude that (13) is lower bounded by $M \cdot \binom{|\mathbf{x}|+\ell}{|\mathbf{x}|}/2$.

Proof. Let us first prove Equation (13). For any two k-uplets $(\mathbf{a} = (a_1, \ldots, a_k), \mathbf{b} = (b_1, \ldots, b_k))$ in $\left([w]^k\right)^2$, let us define

$$\mathbf{y}_{\mathbf{a},\mathbf{b}} \stackrel{\text{def}}{=} (y_{1,a_1,b_1}, \ldots, y_{k,a_k,b_k})$$

and by denoting $F_{w,k,r,\alpha}$ by F,

$$\boldsymbol{\partial}_{\mathbf{a},\mathbf{b}}(F) \stackrel{\text{def}}{=} \frac{\boldsymbol{\partial}^k F}{\boldsymbol{\partial} \mathbf{y}_{\mathbf{a},\mathbf{b}}} = \prod_{i=1}^{k} \prod_{c \in [\alpha]} x_{i,c,a_i}^r \cdot x_{i,c+\alpha,b_i}^r.$$

Notice that $\{\boldsymbol{\partial}_{\mathbf{a},\mathbf{b}}(F)\}$ is a subset of w^{2k} monomials belonging to the set $\sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k} F)$. Hence,

$$\begin{aligned}
\mathrm{SP}_{k,\ell}(F) &= \dim(\mathbf{x}^{\leq \ell} \cdot \sigma_{\mathbf{y}}(\boldsymbol{\partial}_{\mathbf{y}}^{=k} F)) \\
&\geq \dim(\mathbf{x}^{\leq \ell} \cdot \{\boldsymbol{\partial}_{\mathbf{a},\mathbf{b}}(F)\}) \\
&= \left| \mathbf{x}^{\leq \ell} \cdot \{\boldsymbol{\partial}_{\mathbf{a},\mathbf{b}}(F)\} \right|.
\end{aligned} \tag{14}$$

The third step is due to the fact that the dimension of the vector space generated by a set of monomials is exactly the cardinal of this set.

In the following, we will consider a subset of $\{\boldsymbol{\partial}_{\mathbf{a},\mathbf{b}}(F)\}$ made of monomials which are pairwise sufficiently far away. For that, let us define some distances. If $\mathbf{u}$ and $\mathbf{v}$ are two k-vectors,

$$\Delta(\mathbf{u},\mathbf{v}) \stackrel{\text{def}}{=} |\{i \mid u_i \neq v_i\}|.$$

And then

$$\Delta\left(\boldsymbol{\partial}_{\mathbf{a}_1,\mathbf{b}_1}(F), \boldsymbol{\partial}_{\mathbf{a}_2,\mathbf{b}_2}(F)\right) \stackrel{\text{def}}{=} \Delta(\mathbf{a}_1,\mathbf{a}_2) + \Delta(\mathbf{b}_1,\mathbf{b}_2).$$

Claim 31. *There exists $\mathcal{P}_{M,\delta}$ a subset of $\{\boldsymbol{\partial}_{\mathbf{a},\mathbf{b}}(F)\}$ of cardinal M such that if $\boldsymbol{\partial}_{\mathbf{a}_1,\mathbf{b}_1}(F)$ and $\boldsymbol{\partial}_{\mathbf{a}_2,\mathbf{b}_2}(F)$ are two distinct elements of $\mathcal{P}_{M,\delta}$, then*

$$\Delta\left(\boldsymbol{\partial}_{\mathbf{a}_1,\mathbf{b}_1}(F), \boldsymbol{\partial}_{\mathbf{a}_2,\mathbf{b}_2}(F)\right) \geq \lceil \delta k \rceil.$$

Proof. For any monomial m in $\{\boldsymbol{\partial}_{\mathbf{a},\mathbf{b}}(F)\}$, there are at most $\binom{2k}{\lceil \delta k\rceil} \cdot w^{\lceil \delta k \rceil}$ monomials from $\{\boldsymbol{\partial}_{\mathbf{a},\mathbf{b}}(F)\}$ which are at distance at most $\lceil \delta k \rceil$ (for the distance Δ). In particular such a $\mathcal{P}_{M,\delta}$ can be obtained by a greedy algorithm since, for $0 < \delta \leq 1/5$ and sufficiently large k

$$M \cdot \binom{2k}{\lceil \delta k \rceil} \cdot w^{\lceil \delta k \rceil} \leq \frac{w^{2k}}{2^k} \left(\frac{2ek}{\lceil \delta k \rceil}\right)^{\lceil \delta k \rceil} < w^{2k} = |(\boldsymbol{\partial}_{\mathbf{a},\mathbf{b}}(F))|.$$

□

Then, with Equation (14),

$$\begin{aligned} \mathrm{SP}_{k,\ell}(F) &\geq |\mathbf{x}^{\leq \ell} \cdot \mathcal{P}_{M,\delta}| \\ &= \left| \bigcup_{m \in \mathcal{P}_{M,\delta}} \left(\mathbf{x}^{\leq \ell} \cdot m\right) \right| \\ &\geq \sum_{m \in \mathcal{P}_{M,\delta}} |\mathbf{x}^{\leq \ell} \cdot m| - \frac{1}{2} \sum_{m_1 \neq m_2 \in \mathcal{P}_{M,\delta}} |(\mathbf{x}^{\leq \ell} \cdot m_1) \cap (\mathbf{x}^{\leq \ell} \cdot m_2)|. \end{aligned} \tag{15}$$

Let us upperbound the cardinal of $|(\mathbf{x}^{\leq \ell} \cdot m_1) \cap (\mathbf{x}^{\leq \ell} \cdot m_2)|$ for any $m_1 \neq m_2$. For any $\tilde{m}$ in $|(\mathbf{x}^{\leq \ell} \cdot m_1) \cap (\mathbf{x}^{\leq \ell} \cdot m_2)|$, we have $\tilde{m} = m_1 \cdot \tilde{m}_1$ where $\tilde{m}_1$ is a $\mathbf{x}$-monomial of degree at most ℓ. As $\Delta(m_1, m_2) \geq \lceil \delta k \rceil$, it implies there are at least $\lceil \delta k \rceil \cdot \alpha$ many $\mathbf{x}$-variables $\{t_1, \ldots, t_{\lceil \delta k \rceil \cdot \alpha}\}$ which appear (with degree r) in m_2 and not in m_1. So, these variables have to appear in $\tilde{m}_1$. In particular, $\tilde{m} = m_1 \cdot t_1^r \cdot \ldots \cdot t_{\lceil \delta k \rceil \cdot \alpha}^r \cdot \tilde{m}_2$ where $\tilde{m}_2$ is a $\mathbf{x}$-monomial of degree at most $\ell - \lceil \delta k \rceil \cdot \alpha r$. Consequently, for any pair of distinct monomials m_1, m_2 of $\mathcal{P}_{M,\delta}$,

$$|(\mathbf{x}^{\leq \ell} \cdot m_1) \cap (\mathbf{x}^{\leq \ell} \cdot m_2)| \leq \binom{|\mathbf{x}| + \ell - \lceil \delta k \rceil \cdot \alpha r}{|\mathbf{x}|}.$$

Plugging this bound in Equation (15) directly implies Equation (13).

In the case where $\ell \geq |\mathbf{x}|$, $2 \cdot |\mathbf{x}| \cdot \alpha r \geq \ell\beta \cdot \ln w$ and $\beta \geq 4(2-\delta)/\delta$, let us prove that

$$M \cdot \binom{|\mathbf{x}|+\ell}{|\mathbf{x}|} - \frac{M^2}{2} \cdot \binom{|\mathbf{x}|+\ell-\lceil \delta k \rceil \cdot \alpha r}{|\mathbf{x}|} \geq \frac{M}{2} \cdot \binom{|\mathbf{x}|+\ell}{|\mathbf{x}|}.$$

It is sufficient to prove that

$$\frac{M^2}{2} \cdot \binom{|\mathbf{x}|+\ell-\lceil \delta k \rceil \cdot \alpha r}{|\mathbf{x}|} \leq \frac{M}{2} \cdot \binom{|\mathbf{x}|+\ell}{|\mathbf{x}|}.$$

We have

$$\begin{aligned} M \cdot \frac{\binom{|\mathbf{x}|+\ell-\lceil \delta k \rceil \cdot \alpha r}{|\mathbf{x}|}}{\binom{|\mathbf{x}|+\ell}{|\mathbf{x}|}} &= M \cdot \frac{(|\mathbf{x}|+\ell-\lceil \delta k \rceil \cdot \alpha r)! \cdot \ell!}{(\ell-\lceil \delta k \rceil \cdot \alpha r)! \cdot (|\mathbf{x}|+\ell)!} \\ &= M \cdot \frac{(\ell-\lceil \delta k \rceil \cdot \alpha r+1) \cdot \ldots \cdot (\ell)}{(|\mathbf{x}|+\ell-\lceil \delta k \rceil \cdot \alpha r+1) \cdot \ldots \cdot (|\mathbf{x}|+\ell)} \\ &\leq M \cdot \left(1 - \frac{|\mathbf{x}|}{|\mathbf{x}|+\ell}\right)^{\lceil \delta k \rceil \cdot \alpha r} \\ &\leq M \cdot e^{-\frac{|\mathbf{x}| \cdot \lceil \delta k \rceil \cdot \alpha r}{|\mathbf{x}|+\ell}} \\ &\leq \left(\frac{w^{2-\delta}}{2}\right)^k \cdot e^{-\frac{|\mathbf{x}| \cdot \delta \alpha r k}{2\ell}} \qquad (\text{as } \ell \geq |\mathbf{x}|) \\ &\leq \left(w^{2-\delta-\frac{\delta\beta}{4}}\right)^k \qquad (\text{as } 2 \cdot |\mathbf{x}| \cdot \alpha r \geq \ell\beta \cdot \ln w) \\ &\leq 1, \end{aligned}$$

where the last inequality is true since $2 - \delta - \frac{\delta\beta}{4} \leq 0$. $\square$