Superpolynomial lower bounds for general homogeneous depth 4 arithmetic circuits

Mrinal Kumar* Shubhangi Saraf†

Abstract

In this paper, we prove superpolynomial lower bounds for the class of homogeneous depth 4 arithmetic circuits. We give an explicit polynomial in VNP of degree n in n^2 variables such that any homogeneous depth 4 arithmetic circuit computing it must have size $n^{\Omega(\log \log n)}$.

Our results extend the works of Nisan-Wigderson [NW95] (which showed superpolynomial lower bounds for homogeneous depth 3 circuits), Gupta-Kamath-Kayal-Saptharishi and Kayal-Saha-Saptharishi [GKKS13, KSS13] (which showed superpolynomial lower bounds for homogeneous depth 4 circuits with bounded bottom fan-in), Kumar-Saraf [KS13a] (which showed superpolynomial lower bounds for homogeneous depth 4 circuits with bounded top fan-in) and Raz-Yehudayoff and Fournier-Limaye-Malod-Srinivasan [RY08, FLMS13] (which showed superpolynomial lower bounds for multilinear depth 4 circuits). Several of these results in fact showed exponential lower bounds.

The main ingredient in our proof is a new complexity measure of *bounded support* shifted partial derivatives. This measure allows us to prove exponential lower bounds for homogeneous depth 4 circuits where all the monomials computed at the bottom layer have *bounded support* (but possibly unbounded degree/fan-in), strengthening the results of Gupta et al and Kayal et al [GKKS13, KSS13]. This new lower bound combined with a careful "random restriction" procedure (that transforms general depth 4 homogeneous circuits to depth 4 circuits with bounded support) gives us our final result.

1 Introduction

Proving lower bounds for explicit polynomials is one of the most important open problems in the area of algebraic complexity theory. Valiant [Val79] defined the classes VP and VNP as the algebraic analog of the classes P and NP, and showed that proving superpolynomial lower bounds for the Permanent would suffice in separating VP from VNP. Despite the amount of attention received by the problem, we still do not know any superpolynomial (or even *quadratic*) lower bounds for general arithmetic circuits. This absence of progress on the general problem has led to a lot of attention on the problem of proving lower bounds for restricted classes of arithmetic circuits. The hope is that an understanding of restricted classes might lead to a better understanding of the nature of the more general problem, and the techniques developed in this process could possibly be adapted to understand general circuits better. Among the many restricted classes of arithmetic circuits that have been studied with this motivation, *bounded depth* circuits have received a lot of attention.

*Department of Computer Science, Rutgers University. Email: mrinal.kumar@rutgers.edu.

†Department of Computer Science and Department of Mathematics, Rutgers University. Email: shubhangi.saraf@gmail.com.

In a striking result, Valiant et al [VSBR83] showed that any n variate polynomial of degree poly(n) which can be computed by a polynomial sized arithmetic circuit of arbitrary depth can also be computed by an arithmetic circuit of depth $O(\log^2 n)$ and size poly(n). Hence, proving superpolynomial lower bounds for circuits of depth $\log^2 n$ is as hard as proving lower bounds for general arithmetic circuits. In a series of recent works, Agrawal-Vinay [AV08], Koiran [Koi12] and Tavenas [Tav13] showed that the depth reduction techniques of Valiant et al [VSBR83] can in fact be extended much further. They essentially showed that in order to prove superpolynomial lower bounds for general arithmetic circuits, it suffices to prove strong enough lower bounds for just *homogeneous depth 4* circuits. In particular, to separate VNP from VP, it would suffice to focus our attention on proving strong enough lower bounds for homogeneous depth 4 circuits.

The first superpolynomial lower bounds for homogeneous circuits of depth 3 were proved by Nisan and Wigderson [NW95]. Their main technical tool was the use of the *dimension of partial derivatives* of the underlying polynomials as a complexity measure. For many years thereafter, progress on the question of improved lower bounds stalled. In a recent breakthrough result on this problem, Gupta, Kamath, Kayal and Saptharishi [GKKS13] proved the first superpolynomial $(2^{\Omega(\sqrt{n})})$ lower bounds for homogeneous depth 4 circuits when the fan-in of the product gates at the bottom level is bounded (by $\sqrt{n}$). This result was all the more remarkable in light of the results by Koiran [Koi12] and Tavenas [Tav13] which showed that $2^{\omega(\sqrt{n}\log n)}$ lower bounds for this model would suffice in separating VP from VNP. The results of Gupta et al were further improved upon by Kayal Saha and Sapthrashi [KSS13] who showed $2^{\Omega(\sqrt{n}\log n)}$ lower bounds for the model of homogeneous depth 4 circuits when the fan-in of the product gates at the bottom level is bounded (by $\sqrt{n}$). Thus even a slight asymptotic improvement in the exponent of either of these bounds would imply lower bounds for general arithmetic circuits!

The main tool used in both the papers [GKKS13] and [KSS13] was the notion of the dimension of *shifted partial derivatives* as a complexity measure, a refinement of the Nisan-Wigderson complexity measure of dimension of partial derivatives.

In spite of all this exciting progress on homogeneous depth 4 circuits with bounded bottom fanin (which suggests that possibly we might be within reach of lower bounds for much more general classes of circuits) these results give almost no non trivial (not even super linear) lower bounds for general homogeneous depth 4 circuits (with no bound on bottom fanin). Indeed the only lower bounds we know for general homogeneous depth 4 circuits are the slightly superlinear lower bounds by Raz using the notion of elusive functions [Raz10].

Thus nontrivial lower bounds for the class of general depth 4 homogeneous circuits seems like a natural and basic question left open by these works, and strong enough lower bounds for this model seems to be an important barrier to overcome before proving lower bounds for more general classes of circuits.

In this direction, building upon the work in [GKKS13, KSS13], Kumar and Saraf [KS13b, KS13a] proved superpolynomial lower bounds for depth 4 circuits with unbounded bottom fan-in but *bounded top fan-in*. For the case of *multilinear* depth 4 circuits, superpolynomial lower bounds were first proved by Raz and Yehudayoff [RY08]. These lower bounds were recently improved in a paper by Fournier, Limaye, Malod and Srinivasan [FLMS13]. The main technical tool in the work of Fournier et al was the use of the technique of *random restrictions* before using shifted partial derivatives as a complexity measure. By setting a large collection of variables at random to zero, all the product gates with high bottom fan-in got set to zero. Thus the resulting circuit had bounded bottom fanin and then known techniques of shifted partial derivatives could be applied. This idea of random restrictions crucially uses the multilinearity of the circuits, since in multilinear circuits high bottom fanin means *many* distinct variables feeding in to a gate, and thus if a large collection of variables is set at random to zero, then with high probability that gate is also set to zero.

Our Results: In this paper, we prove the first superpolynomial lower bounds for general

homogeneous depth 4 circuits with no restriction on the fan-in, either top or bottom. The main ingredient in our proof is a new complexity measure of *bounded support* shifted partial derivatives. This measure allows us to prove exponential lower bounds for homogeneous depth 4 circuits where all the monomials computed at the bottom layer have only few variables (but possibly large degree/fan-in). This exponential lower bound combined with a careful "random restriction" procedure that allows us to transform general depth 4 homogeneous circuits to this form gives us our final result. We will now formally state our results.

Our main theorem is stated below.

Theorem 1.1 (Lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits)**.** *There is an explicit family of homogeneous polynomials of degree n in n^2 variables in* VNP *which requires homogeneous $\Sigma\Pi\Sigma\Pi$ circuits of size $n^{\Omega(\log\log n)}$ to compute it.*

We prove our lower bound for the family of Nisan-Wigderson polynomials NW_d which is based upon the idea of Nisan-Wigderson designs. We give the formal definition in Section 3.

As a first step in the proof of Theorem 1.1, we prove an exponential lower bound on the top fan-in of any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit where every product gate at the bottom level has at most $O(\log n)$ distinct variables feeding into it. Let homogeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuits denote the class of homogeneous $\Sigma\Pi\Sigma\Pi$ circuits where every product gate at the bottom level has at most s distinct variables feeding into it (i.e. has support at most s).

Theorem 1.2 (Lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits with bounded bottom support)**.** *There exists a constant $\beta > 0$, and an explicit family of homogeneous polynomials of degree n in n^2 variables in* VNP *such that any homogeneous $\Sigma\Pi\Sigma\Pi^{\{\beta\log n\}}$ circuit computing it must have top fan-in at least $2^{\Omega(n)}$.*

Observe that since homogeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuits are a more general class of circuits than homogeneous $\Sigma\Pi\Sigma\Pi$ circuits with bottom fan-in at most s, our result strengthens the results of of Gupta et al and Kayal et al [GKKS13, KSS13] when $s = O(\log n)$.

We prove Theorem 1.1 by applying carefully chosen random restrictions to both the polynomial family and to any arbitrary homogeneous $\Sigma\Pi\Sigma\Pi$ circuit and showing that with high probability the circuit simplifies into a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit with bounded bottom support while the polynomial (even after the restriction) is still rich enough for Theorem 1.2 to hold. Our results hold over every field.

Organization of the paper : The rest of the paper is organized as follows. In Section 2, we provide a high level overview of the proof. In Section 3, we introduce some notations and preliminary notions used in the paper. In Section 4, we give a proof of Theorem 1.2. In Section 5, we describe the random restriction procedure and analyze its effect on the circuit and the polynomial. In Section 6, we prove Theorem 1.1. We conclude with some open problems in Section 7.

2 Proof Overview

Our proof is divided into two parts. In the first part we show a $2^{\Omega(n)}$ lower bound for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits whose *bottom support* is at most $O(\log n)$. To the best of our knowledge, even when the bottom support is 1, none of the earlier lower bound techniques sufficed for showing nontrivial lower bounds for this model. Thus a new complexity measure was needed. We consider the measure of *bounded support* shifted partial derivatives, a refinement of the measure of shifted partial derivatives used in several recent works [GKKS13, KSS13, KS13b, KS13a, FLMS13]. For this measure, we show that the complexity of the NW_d polynomial (an explicit polynomial in VNP) is *high* whereas any subexponential sized homogeneous depth 4 circuit with bounded bottom support has a much smaller complexity measure. Thus for any depth 4 circuit

to compute the NW_d polynomial, it must be large – we show that it must have exponential top fan-in. Thus we get an exponential lower bound for bounded bottom support homogeneous $\Sigma\Pi\Sigma\Pi$ circuits. We believe this result might be of independent interest.

In the second part we show how to "reduce" any $\Sigma\Pi\Sigma\Pi$ circuit that is not too large to a $\Sigma\Pi\Sigma\Pi$ circuit with bounded bottom support. This reduction basically follows from a random restriction procedure that sets some of the variables feeding into the circuit to zero. At the same time we ensure that when this random restriction procedure is applied to NW_d, the polynomial does not get affected very much, and still has large complexity.

We could have set variables to zero by picking the variables to set to zero independently at random. For instance consider the following process: Independently keep each variable alive (i.e. nonzero) with probability $1/n^{\epsilon}$. Then any monomial with $\Omega(\log n)$ distinct variables is set to the zero polynomial with probability at least $1 - 1/n^{\Omega(\log n)}$. Since any circuit of size $n^{o(\log n)}$ will have only $n^{o(\log n)}$ monomials computed at the bottom layer, hence by the union bound, each such monomial with $\Omega(\log n)$ distinct variables will be set to zero. Thus the resulting circuit will have bounded bottom support. The problem with this approach is that we do not know how to analyze the effect of this simple randomized procedure on NW_d. Thus we define a slightly more refined random restriction procedure which keeps the NW_d polynomial hard and at the same time makes the $\Sigma\Pi\Sigma\Pi$ circuit one of bounded bottom support. We describe the details of this procedure in Section 5.1

3 Preliminaries and Notations

Arithmetic Circuits: An arithmetic circuit over a field $\mathbb{F}$ and a set of variables $x_1, x_2, \ldots, x_N$ is an directed acyclic graph whose internal nodes are labelled by the field operations and the leaf nodes are labelled by the variables or field elements. The nodes with fan-out zero are called the output gates and the nodes with fan-in zero are called the leaves. In this paper, we will always assume that there is a unique output gate in the circuit. The *size* of the circuit is the number of nodes in the underlying graph and the *depth* of the circuit is the length of the longest path from the root to a leaf. We will call a circuit *homogeneous* if the polynomial computed at every node is a homogeneous polynomial. By a $\Sigma\Pi\Sigma\Pi$ circuit or a depth 4 circuit, we mean a circuit of depth 4 with the top layer and the third layer only have sum gates and the second and the bottom layer have only product gates. In this paper, we will confine ourselves to working with homogeneous depth 4 circuits. A homogeneous polynomial P of degree n in N variables, which is computed by a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit can be written as

$$P(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{T} \prod_{j=1}^{d_i} Q_{i,j}(x_1, x_2, \ldots, x_N) \tag{1}$$

Here, T is the top fan-in of the circuit. Since the circuit is homogeneous, we know that for every $i \in \{1, 2, 3, \ldots, T\}$,

$$\sum_{j=i}^{d_i} \deg(Q_{i,j}) = n$$

By the support of a monomial α, we will refer to the set of variables which have a positive degree in α. In this paper, we will also study the class of homogeneous $\Sigma\Pi\Sigma\Pi$ circuits such that for every i, j, every monomial in $Q_{i,j}$ has bounded support. We will now formally define this class.

Homogeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ Circuits: A homogeneous $\Sigma\Pi\Sigma\Pi$ circuit in Equation 1, is said to be a $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuit if every product gate at the bottom level has support at most s. Observe that there is no restriction on the bottom fan-in except that implied by the restriction of homogeneity.

Shifted Partial Derivatives: In this paper will use a variant of the notion of *shifted partial derivatives* which was introduced in [Kay12] and has subsequently been the complexity measure used to to prove lower bounds for various restricted classes of depth four circuits and formulas [FLMS13, GKKS13, KSS13, KS13b, KS13a]. For a field $\mathbb{F}$, an N variate polynomial $P \in \mathbb{F}[x_1, \ldots, x_N]$ and a positive integer r, we denote by $\partial^r P$, the set of all partial derivatives of order equal to r of P. For a polynomial P and a monomial γ, we denote by $\partial_\gamma(P)$ the partial derivative of P with respect to γ. We now reproduce the formal definition from [GKKS13].

Definition 3.1 (Order-r ℓ-shifted partial derivatives)**.** *For an N variate polynomial $P \in \mathbb{F}[x_1, x_2, \ldots, x_N]$ and positive integers $r, \ell \geq 0$, the space of order-r ℓ-shifted partial derivatives of P is defined as*

$$\langle \partial^r P \rangle_\ell \stackrel{def}{=} \mathbb{F}\text{-}span\{ \prod_{i \in [N]} {x_i}^{j_i} \cdot g : \sum_{i \in [N]} j_i = \ell, g \in \partial^r P\} \tag{2}$$

In this paper, we introduce the variation of *bounded support* shifted partial derivatives as a complexity measure. The basic difference is that instead of shifting the partial derivatives by all monomials of degree ℓ, we will shift the partial derivatives only by only those monomials of degree ℓ which have support(the number of distinct variables which have non-zero degree in the monomial) exactly equal to m. We now formally define the notion.

Definition 3.2 (Support-m degree-ℓ shifted partial derivatives of order-r)**.** *For an N variate polynomial $P \in \mathbb{F}[x_1, x_2, \ldots, x_N]$ and positive integers $r, \ell, m \geq 0$, the space of support-m degree-ℓ shifted partial derivatives of order-r of P is defined as*

$$\langle \partial^r P \rangle_{(\ell,m)} \stackrel{def}{=} \mathbb{F}\text{-}span\{ \prod_{\substack{i \in S \\ S \subseteq [N] \\ |S|=m}} {x_i}^{j_i} \cdot g : \sum_{i \in S} j_i = \ell, j_i \geq 1, g \in \partial^r P\} \tag{3}$$

The following property follows from the definition above.

Lemma 3.3. *For any two multivariate polynomials P and Q in $\mathbb{F}[x_1, x_2, \ldots, x_N]$ and any positive integers r, ℓ, m, and scalars α and β*

$$\mathsf{Dim}(\langle \partial^r(\alpha P + \beta Q) \rangle_{(\ell,m)}) \leq \mathsf{Dim}(\langle \partial^r P \rangle_{(\ell,m)}) + \mathsf{Dim}(\langle \partial^r Q \rangle_{(\ell,m)})$$

In the rest of the paper, we will use the term (m, ℓ, r)-shifted partial derivatives to refer to support-m degree-ℓ shifted partial derivatives of order-r of a polynomial. For any linear or affine space V over a field $\mathbb{F}$, we will use $\mathsf{Dim}(V)$ to represent the dimension of V over $\mathbb{F}$. We will use the dimension of the space $\langle \partial^r P \rangle_{(\ell,m)}$ which we denote by $\mathsf{Dim}(\langle \partial^r P \rangle_{(\ell,m)})$ as the measure of complexity of a polynomial.

Nisan-Wigderson Polynomials: We will show our lower bounds for a family of polynomials in VNP which were used for the first time in the context of lower bounds in [KSS13]. The construction is based upon the intuition that over any finite field, any two distinct low degree polynomials do not agree at too many points. For the rest of this paper, we will assume n to be of the form 2^k for some positive integer k. Let $\mathbb{F}_n$ be a field of size n. For the set of $N = n^2$ variables $\{x_{i,j} : i, j \in [n]\}$ and $d < n$, we define the degree n homogeneous polynomial NW_d as

$$NW_d = \sum_{\substack{f(z) \in \mathbb{F}_n[z] \\ deg(f) \leq d-1}} \prod_{i \in [n]} x_{i,f(i)}$$

From the definition, we can observe the following properties of NW_d.

1. The number of monomials in NW_d is exactly n^d.

2. Each of the monomials in NW_d is multilinear.
3. Each monomial corresponds to evaluations of a univariate polynomial of degree at most $d-1$ at all points of $\mathbb{F}_n$. Thus, any two distinct monomials agree in at most $d-1$ variables in their support.

For any $S \subseteq [n]$ and each $f \in \mathbb{F}_n[z]$, we define the monomial

$$m_f^S = \prod_{i \in S} x_{i,f(i)}$$

and

$$m_f = \prod_{i \in [n]} x_{i,f(i)}$$

We also define the set $\mathcal{M}^S$ to represent the set $\{\prod_{i \in S} \prod_{j \in [n]} x_{i,j}\}$. Clearly,

$$NW_d = \sum_{\substack{f(z) \in \mathbb{F}_n[z] \\ deg(f) \leq d-1}} m_f$$

Monomial Ordering and Distance: We will also use the notion of a monomial being an extension of another as defined below.

Definition 3.4. *A monomial θ is said to be an extension of a monomial $\tilde{\theta}$, if θ divides $\tilde{\theta}$.*

In this paper, we will imagine our variables to be coming from a $n \times n$ matrix $\{x_{i,j}\}_{i,j \in [n]}$. We will also consider the following total order on the variables. $x_{i_1,j_1} > x_{i_2,j_2}$ if either $i_1 < i_2$ or $i_1 = i_2$ and $j_1 < j_2$. This total order induces a lexicographic order on the monomials. For a polynomial P, we will use the notation $\mathsf{Lead\text{-}Mon}(P)$ to indicate the leading monomial of P under this monomial ordering.

We will use the following notion of distance between two monomials which was also used in [CM13].

Definition 3.5 (Monomial distance)**.** *Let m_1 and m_2 be two monomials over a set of variables. Let S_1 and S_2 be the multiset of variables in m_1 and m_2 respectively, then the distance $\Delta(m_1, m_2)$ between m_1 and m_2 is the $min\{|S_1| - |S_1 \cap S_2|, |S_2| - |S_1 \cap S_2|\}$ where the cardinalities are the order of the multisets.*

In this paper, we will invoke this definition only for multilinear monomials of the same degree. In this special case, we have the following crucial observation.

Observation 3.6. *Let α and β be two multilinear monomials of the same degree which are at a distance Δ from each other. If $Supp(\alpha)$ and $Supp(\beta)$ are the supports of α and β respectively, then*

$$|Supp(\alpha)| - |Supp(\alpha) \cap Supp(\beta)| = |Supp(\beta)| - |Supp(\alpha) \cap Supp(\beta)| = \Delta$$

Approximations: We will repeatedly refer to the following lemma to approximate expressions during our calculations.

Lemma 3.7 ([GKKS13])**.** *Let $a(n), f(n), g(n) : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be integer valued functions such that $(f+g) = o(a)$. Then,*

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

In our setup, very often $(f+g)^2$ will be $\theta(a)$. In this case, the error term will be an absolute constant. Hence, up to multiplication by constants, $\frac{(a+f)!}{(a-g)!} = a^{(f+g)}$.

We will also use the following basic fact in our proof.

Fact 3.8. *The number of positive integral solutions of the equation*

$$\sum_{i=1}^{t} y_i = k$$

equals $\binom{k-1}{t-1}$.

As a last piece of notation, for any $i \times j$ matrix H over $\mathbb{F}_2$ and a vector $\alpha \in \mathbb{F}_2^i$, we denote by $H||\alpha$ to be the $i \times (j+1)$ matrix which when restricted to the first j columns is equal to H and whose last column is α. Similarly, for any vector $\alpha \in \mathbb{F}_2^i$ and any $b \in \mathbb{F}_2$, $\alpha||b$ is the $i+1$ dimensional vector where b is appended to α.

4 Lower bounds for $\Sigma\Pi\Sigma\Pi^{\{O(\log n)\}}$ circuits

In this section, we will prove Theorem 1.2. We will prove an exponential lower bound on the top fan-in for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits such that every product gate at the bottom has a bounded number of variables feeding into it. We will use the dimension of the span of (m, ℓ, r)-shifted partial derivatives as the complexity measure. We will prove our lower bound for the NW_d polynomial. The proof will be in two parts. In the first part, we will prove an upper bounded on the complexity of the circuit. Then, we will prove a lower bound on the complexity of the NW_d polynomial. Comparing the two will then imply our lower bound. The bound holds for NW_d for any $d = \delta n$, where δ is a constant such that $0 < \delta < 1$.

4.1 Complexity of homogeneous depth 4 $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuits

Let C be a homogeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuit computing the NW_d polynomial. We will now prove an upper bound on the complexity of a product gate in such a circuit. The bound on the complexity of the circuit follows from the subadditivity of the complexity measure.

Lemma 4.1. *Let $Q = \prod_{i=1}^{n} Q_i$ be a product gate at the second layer from the top in a homogeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuit computing a homogeneous degree n polynomial in N variables. For any positive integers m, r, s, ℓ satisfying $m + rs \leq \frac{N}{2}$ and $m + rs \leq \frac{\ell}{2}$,*

$$\mathsf{Dim}(\langle \partial^r Q \rangle_{(\ell,m)}) \leq \mathit{poly}(nrs) \binom{n+r}{r} \binom{N}{m+rs} \binom{\ell+n-r}{m+rs}$$

Proof. By the application of chain rule, any partial derivative of order r of Q is a linear combination of a number of product terms. Each of these product terms is of the form $\prod_{i \in S} \partial_{\gamma_i}(Q_i) \prod_{j \in [n] \setminus S} Q_j$, where S is a subset of $\{1, 2, \ldots, n\}$ of size at most r and γ_i are monomials such that $\sum_{i \in S} \deg(\gamma_i) = r$. Also, observe that $\prod_{i \in S} \partial_{\gamma_i}(Q_i)$ is of degree at most $n - r$. In this particular special case all Q_i have support at most s, so every monomial in $\prod_{i \in S} \partial_{\gamma_i}(Q_i)$ has support at most rs. Shifting these derivatives is the same as multiplying them with monomials of degree ℓ and support equal to m. So, (m, ℓ, r)-shifted partial derivative of order r can be expressed as sum of the product of $\prod_{j \in [n] \setminus S} Q_j$ for $S \subseteq [n]$ of size at most r, and a monomial of support between m and $m + rs$ and degree between ℓ and $\ell + n - r$.

We can choose the set S in $\binom{n+r}{r}$ ways. The second part in each term is a monomial of degree between l and $\ell + n - r$ and support between m and $m + rs$. The number of monomials

over N variables of support between m and $m+rs$ and degree between ℓ and $\ell+n-r$ equals

$$\sum_{i=0}^{n-r}\sum_{j=0}^{rs}\binom{N}{m+j}\binom{\ell+i-1}{m+j-1}$$

Now, in the range of choice of our parameters m, r, s, ℓ, the binomial coefficients increase monotonically with i and j. Hence, we can upper bound the dimension by $\text{poly}(nrs)\binom{n+r}{r}\binom{N}{m+rs}\binom{\ell+n-r-1}{m+rs-1}$. □

For a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit where each of the bottom level product gates is of support at most s, Lemma 4.1 immediately implies the following upper bound on the complexity of the circuit due to subadditivity from Lemma 3.3.

Corollary 4.2 (Upper bound on circuit complexity)**.** *Let $C = \sum_{j=1}^{T}\prod_{i=1}^{n} Q_{i,j}$ be a a homogeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuit computing a homogeneous degree n polynomial in N variables. For any m, r, s, ℓ satisfying $m+rs \leq \frac{N}{2}$ and $m+rs \leq \frac{\ell}{2}$,*

$$\mathsf{Dim}(\langle \partial^r C\rangle_{(\ell,m)}) \leq T \times \textit{poly}(nrs)\binom{n+r}{r}\binom{N}{m+rs}\binom{\ell+n-r-1}{m+rs-1}$$

4.2 Lower bound on the complexity of the NW_d polynomial

We will now prove a lower bound on the complexity of the NW_d polynomial. For this, we will first observe that distinct partial derivatives of the NW_d polynomial are *far* from each other in some sense and then show that shifting such partial derivatives gives us a lot of distinct shifted partial derivatives. Recall that we defined the set $\mathcal{M}^S$ to represent the set $\{\prod_{i\in S}\prod_{j\in[n]} x_{i,j}\}$. We start with the following observation.

Lemma 4.3. *For any positive integer r such that $n-r > d$ and $r < d-1$, the set $\{\partial_\alpha(NW_d) : \alpha \in \mathcal{M}^{[r]}\}$ consists of $|\mathcal{M}^{[r]}| = n^r$ nonzero distinct polynomials.*

Proof. We need to show the following two statements.

- $\forall \alpha \in \mathcal{M}^{[r]}$, $\partial_\alpha(NW_d)$ is a non zero polynomial.
- $\forall \alpha \neq \beta \in \mathcal{M}^{[r]}$, $\partial_\alpha(NW_d) \neq \partial_\beta(NW_d)$.

For the first item, observe that, since $r < d-1$, for every $\alpha \in \mathcal{M}^{[r]}$, there is a polynomial f of degree at most $d-1$ in $\mathbb{F}_n[z]$ such that $\alpha = \prod_{i=1}^{r} x_{i,f(i)}$. So, $\partial_\alpha(m_f) \neq 0$ since m_f is an extension of α, in fact, there are many such extensions. Also, observe for any two extensions m_f and m_g, $\partial_\alpha(m_f)$ and $\partial_\alpha(m_g)$ are multilinear monomials at a distance at least $n-r-d > 0$ from each other. Hence, $\partial_\alpha(NW_d) = \sum_g \partial_\alpha(m_g)$ is a non zero polynomial, where the sum is over all $g \in \mathbb{F}_n[z]$ of degree $\leq d-1$ such that m_g is an extension of α.

For the second item, let us now consider the leading monomials of $\partial_\alpha(NW_d)$ and $\partial_\beta(NW_d)$. These leading monomials each come from some distinct polynomials $f, g \in \mathbb{F}_n[z]$ of degree at most $d-1$. Also, since $\alpha \neq \beta$ and $n-r > d$, $\partial_\alpha(m_f) \neq \partial_\beta(m_g)$. In fact, $\partial_\alpha(NW_d)$ and $\partial_\beta(NW_d)$ do not have a common monomial. Therefore, $\partial_\alpha(NW_d) \neq \partial_\beta(NW_d)$. □

Remark 4.4. *Observe that there is nothing special about the set $\mathcal{M}^{[r]}$ and the Lemma 4.3 holds for $\{\mathcal{M}\}^S$ for any set S, such that $S \subseteq [n]$ and $|S| < d-1$.*

In the proof above, we observed that for any $\alpha \neq \beta \in \mathcal{M}^{[r]}$, the leading monomials of $\partial_\alpha(NW_d)$ and $\partial_\beta(NW_d)$ are multilinear monomials of at a distance at least $n-r-d$ from each other. We will exploit this structure to show that shifting the polynomials in the set $\{\partial_\alpha(NW_d) : \alpha \in \mathcal{M}^{[r]}\}$ by monomials of support m and degree ℓ results in many linearly independent shifted partial derivatives. We will first prove the following lemma.

Lemma 4.5. *Let α and β be two distinct multilinear monomials of equal degree such that the distance between them is Δ. Let S_α and S_β be the set of all monomials obtained by shifting α and β respectively with monomials of degree ℓ and support exactly m over N variables. Then $|S_\alpha \cap S_\beta| \leq \binom{N-\Delta}{m-\Delta}\binom{\ell-1}{m-1}$.*

Proof. From the distance property, we know that there is a unique monomial γ of degree Δ and support Δ such that $\alpha\gamma$ is the lowest degree extension of α which is divisible by β. Therefore, any extension of α which is also an extension of β must have the support of $\alpha\gamma$ as a subset. In particular, for a shift of α to lie in S_β, α must be shifted by monomial of degree ℓ and support m which is an extension of γ. Hence, the freedom in picking the support is restricted to picking some $m - \Delta$ variables from the remaining $N - \Delta$ variables. Once the support is chosen, the number of possible degree ℓ shifts on this support equals $\binom{\ell-1}{m-1}$ by Fact 3.8. Hence, the number of shifts of degree equal to ℓ and support equal to m of α which equals some degree ℓ and support m shift of β is exactly $\binom{N-\Delta}{m-\Delta}\binom{\ell-1}{m-1}$. □

We will now prove the following lemma, which is essentially an application of Claim 4.5 to the NW_d polynomial. For any monomial α and positive integers ℓ, m, we will denote by $S_{\ell,m}(\alpha)$ the set of all shifts of $\partial_\alpha NW_d$ by monomials of degree ℓ and support m. More formally,

$$S_{\ell,m}(\alpha) = \{\gamma \cdot \partial_\alpha(NW_d) : \gamma = \prod_{\substack{i\in U \\ U\subseteq[N] \\ |U|=m}} {x_i}^{j_i}, \sum_{i\in U} j_i = \ell, j_i \geq 1\}$$

also, let

$$LM_{\ell,m}(\alpha) = \{\mathsf{Lead\text{-}Mon}(f) : f \in S_{\ell,m}(\alpha)\}$$

Lemma 4.6. *For any positive integers r, m and ℓ such that $n - r > d$ and $r < d - 1$, let α and β be two distinct monomials in $\mathcal{M}^{[r]}$. Then $|S_{\ell,m}(\alpha) \cap S_{\ell,m}(\beta)| \leq \binom{N-(n-d-r)}{m-(n-d-r)}\binom{\ell-1}{m-1}$.*

Proof. In the proof of Lemma 4.3, we have observed that the leading monomials of $\partial_\alpha(NW_d)$ and $\partial_\beta(NW_d)$ are equal to $\partial_\alpha(m_f)$ and $\partial_\beta(m_g)$ for two distinct polynomials $f, g \in \mathbb{F}_n[z]$ of degree at most $d - 1$. Hence, $\partial_\alpha(m_f)$ and $\partial_\beta(m_g)$ are multilinear monomials at a distance at least $\Delta = n - r - d$ from each other.

Since monomial orderings respect multiplication by the same polynomial, we know that the leading monomial of a shift equals the shift of the leading monomial. Therefore, if γ_α and γ_β are two monomials of degree ℓ and support equal to m such that $\gamma_\alpha\partial_\alpha(NW_d) = \gamma_\beta\partial_\beta(NW_d)$, then $\gamma_\alpha\partial_\alpha(m_f) = \gamma_\beta\partial_\beta(m_g)$. Hence, the $|S_{\ell,m}(\alpha) \cap S_{\ell,m}(\beta)|$ is at most the number of shifts of $\partial_\alpha(m_f)$ which is also a shift of $\partial_\beta(m_g)$. By Lemma 4.5, this is at most $\binom{N-(n-d-r)}{m-(n-d-r)}\binom{\ell-1}{m-1}$. □

We will now prove a lower bound on the dimension of the span of (m, ℓ, r)-shifted partial derivatives of the NW_d polynomial. For this, we will use the following proposition from [GKKS13], the proof of which is a simple application of Gaussian elimination.

Proposition 4.7 ([GKKS13])**.** *For any field $\mathbb{F}$, let $\mathcal{P} \subseteq \mathbb{F}[z]$ be any finite set of polynomials. Then,*

$$\mathsf{Dim}(\mathbb{F}\text{-}span(\mathcal{P})) = |\{\mathsf{Lead\text{-}Mon}(f) : f \in \mathbb{F}\text{-}span(\mathcal{P})\}|$$

Therefore, in order to lower bound $\mathsf{Dim}(\langle \partial^r NW_d\rangle_{(\ell,m)})$, it would suffice to obtain a lower bound on the size of the set $\bigcup_\alpha LM_{\ell,m}(\alpha)$, where the union is over all monomials α of degree equal to r. To obtain this lower bound, we will show a lower bound on the size of the set $\bigcup_{\alpha\in\mathcal{M}^{[r]}} LM_{\ell,m}(\alpha)$.

Lemma 4.8. *Let $d = \delta n$ for any constant $0 < \delta < 1$. Let ℓ, m, r be positive integers such that $n - r > d$, $r < d - 1$, $m \leq N$, $m = \theta(N)$ and for $\phi = \frac{N}{m}$, r satisfies $r \leq \frac{(n-d)\log\phi \pm O(\phi\frac{(n-d-r)^2}{N})}{\log n + \log\phi}$. Then,*

$$\mathsf{Dim}(\langle \partial^r NW_d \rangle_{(\ell,m)}) \geq 0.5n^r \binom{N}{m}\binom{\ell-1}{m-1}$$

Proof. Recall that $\mathcal{M}^{[r]} = \{\prod_{i=1}^{r}\prod_{j\in[n]} x_{i,j}\}$. We have argued in Lemma 4.3 that for each $\alpha, \beta \in \mathcal{M}^{[r]}$, such that $\alpha \neq \beta$, $\partial_\alpha(NW_d) \neq \partial_\beta(NW_d)$ and both of these are non zero polynomials. As discussed above, we will prove a lower bound on the size of the set $\bigcup_{\alpha\in\mathcal{M}^{[r]}} LM_{\ell,m}(\alpha)$. From the principle of inclusion-exclusion, we know

$$|\bigcup_{\alpha\in\mathcal{M}^{[r]}} LM_{\ell,m}(\alpha)| \geq \sum_{\alpha\in\mathcal{M}^{[r]}} |LM_{\ell,m}(\alpha)| - \sum_{\alpha\neq\beta\in\mathcal{M}^{[r]}} |LM_{\ell,m}(\alpha)\cap LM_{\ell,m}(\beta)|$$

Let us now bound both these terms separately.

- Since shifting preserves monomial orderings, therefore for any $\gamma \neq \tilde{\gamma}$ of degree ℓ and support m, and for any $\alpha \in \mathcal{M}^{[r]}$, $\mathsf{Lead\text{-}Mon}(\gamma\partial_\alpha(NW_d)) \neq \mathsf{Lead\text{-}Mon}(\tilde{\gamma}\partial_\alpha(NW_d))$. Hence, for each $\alpha \in \mathcal{M}^{[r]}$, $|LM_{\ell,m}(\alpha)|$ is the number of different shifts possible, which is equal to the number of distinct monomials of degree ℓ and support m over N variables. Hence,

  $$|LM_{\ell,m}(\alpha)| = \binom{N}{m}\binom{\ell-1}{m-1}$$

  .
- For any two distinct $\alpha, \beta \in \mathcal{M}^{[r]}$, from Lemma 4.6,

  $$|LM_{\ell,m}(\alpha)\cap LM_{\ell,m}(\beta)| \leq \binom{N-(n-d-r)}{m-(n-d-r)}\binom{\ell-1}{m-1}$$

  Therefore,

  $$|\bigcup_{\alpha\in\mathcal{M}^{[r]}} LM_{\ell,m}(\alpha)| \geq |\mathcal{M}^{[r]}|\binom{N}{m}\binom{\ell-1}{m-1} - \binom{|\mathcal{M}^{[r]}|}{2}\binom{N-(n-d-r)}{m-(n-d-r)}\binom{\ell-1}{m-1}$$

To simplify this bound, we will show that for the choice of our parameters, the second term is at most the half the first term. In this case, we have

$$|\bigcup_{\alpha\in\mathcal{M}^{[r]}} LM_{\ell,m}(\alpha)| \geq 0.5|\mathcal{M}^{[r]}|\binom{N}{m}\binom{\ell-1}{m-1}$$

We need to ensure,

$$\frac{\binom{|\mathcal{M}^{[r]}|}{2}\binom{N-(n-d-r)}{m-(n-d-r)}\binom{\ell-1}{m-1}}{|\mathcal{M}^{[r]}|\binom{N}{m}\binom{\ell-1}{m-1}} \leq 0.5$$

It suffices to ensure

$$\frac{|\mathcal{M}^{[r]}|\binom{N-(n-d-r)}{m-(n-d-r)}}{\binom{N}{m}} \leq 1$$

which is the same as ensuring that

$$|\mathcal{M}^{[r]}| \times \frac{(N-(n-d-r))!}{N!} \times \frac{m!}{(m-(n-d-r))!} \leq 1$$

Now, using the approximation from Lemma 3.7,

$$\begin{aligned}\log \frac{N!}{(N-(n-d-r))!} &= (n-d-r)\log N \pm O\left(\frac{(n-d-r)^2}{N}\right) \text{ and}\\ \log \frac{m!}{(m-(n-d-r))!} &= (n-d-r)\log m \pm O\left(\frac{(n-d-r)^2}{m}\right)\end{aligned}$$

Thus we need to ensure that

$$\log |\mathcal{M}^{[r]}| \leq \log \left(\frac{N}{m}\right)^{n-d-r} \pm O\left(\frac{(n-d-r)^2}{N}\right) \pm O\left(\frac{(n-d-r)^2}{m}\right)$$

Substituting $|\mathcal{M}^{[r]}| = n^r$, we need

$$r \log n \leq \log \left(\frac{N}{m}\right)^{n-d-r} \pm O\left(\frac{(n-d-r)^2}{N} + \frac{(n-d-r)^2}{m}\right)$$

Substituting $m = \frac{N}{\phi}$ (and noting that $\phi > 1$), we require

$$r \log n \leq (n-d-r)\log \phi \pm O\left(\phi \frac{(n-d-r)^2}{N}\right).$$

Thus we require

$$r \leq \frac{(n-d)\log \phi \pm O(\phi^{\frac{(n-d-r)^2}{N}})}{\log n + \log \phi}$$

Observe that for any constant $0 < \delta < 1$ such that $d = \delta n$, r can be chosen any constant times $\frac{n}{\log n}$ by choosing ϕ to be an appropriately large constant. So, for such a choice of r,

$$\mathsf{Dim}(\langle \partial^r NW_d \rangle_{(\ell,m)}) \geq 0.5|\mathcal{M}^{[r]}|\binom{N}{m}\binom{\ell-1}{m-1}$$

For $|\mathcal{M}^{[r]}| = n^r$, we have

$$\mathsf{Dim}(\langle \partial^r NW_d \rangle_{(\ell,m)}) \geq 0.5n^r\binom{N}{m}\binom{\ell-1}{m-1}$$

□

Remark 4.9. *The proof above shows something slightly more general than a lower bound on just the complexity of the NW_d polynomial. The only property of the NW_d polynomial that we used here was that the leading monomials of any two distinct partial derivatives of it were far from each other. We will crucially use this observation in the proof of our main theorem. Also, there is nothing special about using the set $\mathcal{M}^{[r]}$. The proof works for any set of monomials $\mathcal{M}^S = \{\prod_{i\in S}\prod_{j\in[n]} x_{i,j}\}$, where S is a subset of $\{1,2,3,\ldots,n\}$ of size exactly r.*

4.3 Top fan-in lower bound

We are now ready to prove our lower bound on the top fan-in of any homogeneous $\Sigma\Pi\Sigma\Pi^{\{\beta \log n\}}$ (for some constant β) and computes the NW_d polynomial, where $d = \delta n$ for some constant δ between 0 and 1.

Theorem 4.10. *Let $d = \delta n$ for any constant $0 < \delta < 1$. There exists a constant β such that all homogeneous $\Sigma\Pi\Sigma\Pi^{\{\beta \log n\}}$ circuits which compute the NW_d polynomial have top fan-in at least $2^{\Omega(n)}$.*

Proof. By comparing the complexities of the circuit and the polynomial as given by Corollary 4.2 and Lemma 4.8, the top fan-in of the circuit must be at least

$$\frac{0.5n^r\binom{N}{m}\binom{\ell-1}{m-1}}{\text{poly}(nrs)\binom{n+r}{r}\binom{N}{m+rs}\binom{\ell+n-r}{m+rs}} \tag{4}$$

This bound holds for any choice of positive integers ℓ, m, r, a constant β such that $s = \beta \log n$ which satisfy the constraints in the hypothesis of Corollary 4.2 and Lemma 4.8. In other words, we want these parameters to satisfy

- $m + rs \leq \frac{N}{2}$
- $m + rs \leq \frac{\ell}{2}$
- $m = \theta(N)$
- $n - r > d$
- $r < d - 1$
- For $\phi = \frac{N}{m}$, $r \leq \frac{(n-d)\log\phi \pm O\left(\phi\frac{(n-d-r)^2}{N}\right)}{\log n + \log\phi}$

In the rest of the proof, we will show that there exists a choice of these parameters such that we get a bound of $2^{\Omega(n)}$ from Expression 4. We will show the existence of such parameters satisfying the asymptotics $\ell = \theta(N)$, $r = \theta\left(\frac{n}{\log n}\right)$ and $s = \theta(\log n)$. In the rest of the proof, we will crucially use these asymptotic bounds for various approximations.

For this, we will group together and approximate the terms in the ratio $\frac{0.5n^r\binom{N}{m}\binom{\ell-1}{m-1}}{\text{poly}(nrs)\binom{n+r}{r}\binom{N}{m+rs}\binom{\ell+n-r}{m+rs}}$

- $\frac{\binom{N}{m}}{\binom{N}{m+rs}} = \frac{(N-m-rs)!(m+rs)!}{(N-m)!m!} = (\frac{m}{N-m})^{rs}$ upto some constant factors, as long as $(rs)^2 = \theta(N) = \theta(m)$.
- $\frac{\binom{\ell-1}{m-1}}{\binom{\ell+n-r}{m+rs}} = \frac{(\ell-1)!}{(m-1)!(\ell-m)!} \times \frac{(m+rs)!(\ell-m+n-r-rs)!}{(\ell+n-r)!}$. We now pair up things we know how to approximate within constant factors. $\frac{\binom{\ell-1}{m-1}}{\binom{\ell+n-r}{m+rs}} = \frac{(\ell-1)!}{(\ell+n-r)} \times \frac{(m+rs)!}{(m-1)!} \times \frac{(\ell-m+n-r-rs)!}{(\ell-m)!} =$ poly(n) $\times \frac{1}{\ell^{n-r}} \times m^{rs} \times \frac{(\ell-m)^{n-r}}{(\ell-m)^{rs}}$. This simplifies to poly(n) $\times \left(\frac{m}{\ell-m}\right)^{rs} \times \left(\frac{\ell-m}{\ell}\right)^{n-r}$.
- $\frac{n^r}{\binom{n+r}{r}} \geq \frac{n^r}{\left(\frac{2(n+r)}{r}\right)^r}$. We just used Stirling's approximation here.

In the range of our parameters, the approximations above imply that the top fan-in, up to polynomial factors is at least

$$\left(\frac{r}{3}\right)^r \times \left(\frac{m}{\ell-m}\right)^{rs} \times \left(\frac{\ell-m}{\ell}\right)^{n-r} \times \left(\frac{m}{N-m}\right)^{rs}$$

Simplifying further, this is at least

$$2^{\Omega(r\log r - rs\log\frac{\ell-m}{m} - (n-r)\log\frac{\ell}{\ell-m} - rs\log\frac{N-m}{m})}$$

Recall that we will set m and ℓ to be $\theta(N)$ and r to be $\theta(\frac{n}{\log n})$. The constants have to be chosen carefully in order to satisfy the constraints. We will choose constants α, β and η such that $s = \beta\log n$, $r = \alpha \cdot n/\log n$ and $m = \eta\ell$. First choose η to be any small constant > 0 (for instance $\eta = 1/4$). Now, choose α to be a constant much larger than $\log\frac{1}{1-\eta}$. This makes sure that $r\log r$ dominates $(n-r)\log\frac{\ell}{\ell-m}$. Recall that α can be chosen to be any large constant by choosing ϕ to be an appropriately large constant (by the constraint between r and ϕ in the fifth

bullet). Notice that this sets m to be a small constant factor of N. Fix these choices of η and α. Now, we choose the term β to be a small positive constant such that $rs \log \frac{1-\eta}{\eta}$ and $rs \log \frac{N-m}{m}$ are much less than $r \log r$. Observe that this choice of parameters satisfies all the constraints imposed in the calculations above, and the top fan-in is at least $2^{\Omega(r \log r)} = 2^{\Omega(n)}$. □

5 Random Restrictions

In this section, we will describe our random restriction algorithm and analyze the effect of random restrictions on $\Sigma\Pi\Sigma\Pi$ circuits as well as the NW_d polynomial.

Let $n = 2^k$. We identify elements of $[n]$ with elements of $\mathbb{F}_{2^k}$. We view $\mathbb{F}_{2^k}$ as a k-dimensional vector space over $\mathbb{F}_2$. Let $\phi : \mathbb{F}_{2^k} \to \mathbb{F}_2^k$ be an $\mathbb{F}_2$-linear isomorphism between $\mathbb{F}_{2^k}$ and $\mathbb{F}_2^k$. Thus $\phi(\alpha + \beta) = \phi(\alpha) + \phi(\beta)$. Let $M : \mathbb{F}_{2^k} \to \mathbb{F}_2^{k \times k}$, map $\alpha \in \mathbb{F}_{2^k}$ to the matrix $M(\alpha)$, which represents the linear transformation over $\mathbb{F}_2^k$ that is given by multiplication by α in $\mathbb{F}_{2^k}$. Thus it follows that $M(\alpha \times \beta) = M(\alpha) \times M(\beta)$, and $M(\alpha + \beta) = M(\alpha) + M(\beta)$. Moreover it is not hard to see that $\phi(\alpha \times \beta) = M(\alpha) \times \phi(\beta)$.

Since $n = 2^k$, thus $\mathbb{F}_n \equiv \mathbb{F}_{2^k}$. Let $\mathbb{F}_n[Z]$ denote the space of univariate polynomials over $\mathbb{F}_n$. For $f \in \mathbb{F}_n[Z]$ of degree $\leq d-1$, f is of the form $\sum_{i=0}^{d-1} a_i Z^i$, for $a_i \in \mathbb{F}_n$. Thus we can represent f as a vector of coefficients $(a_0, a_1, \ldots a_{d-1})$, and hence view f as an element of $\mathbb{F}_n^d$. For ease of notation, for $\alpha \in \mathbb{F}_n$ we will let $[\alpha]$ represent $\phi(\alpha)$. Also, for $f \in \mathbb{F}_n[Z]$ of degree at most $d-1$, we let $[f] \in \mathbb{F}_2^{kd}$ represent the concatenation of ϕ applied to each of the coefficients of f.

Let Eval_α be the $dk \times k$ matrix obtained by stacking the matrices $M(\alpha^0)$, $M(\alpha^1)$, ..., $M(\alpha^{d-1})$ one below the other. In other words, the first k rows are the rows of $M(\alpha^0)$, the second k rows are the rows of $M(\alpha^1)$ and so on. The following claim follows easily from the definitions.

Claim 5.1. *Let $f \in \mathbb{F}_n[Z]$ be of degree at most $d-1$, and let $\alpha \in \mathbb{F}_n$. Then*

$$[f(\alpha)] = [f] \times \mathsf{Eval}_\alpha.$$

In the rest of the discussion we will identify the elements of $\mathbb{F}_n$ with $\{1, 2, \ldots, n\}$. Let $\overline{\mathsf{Eval}_i}$ be the $dk \times 2^k$ matrix obtained by adding a column for each of the 2^k linear combinations of the columns of Eval_i. Let Eval be the $dk \times nk$ matrix obtained by concatenating Eval_i for all $i \in [n]$. Let $\overline{\mathsf{Eval}}$ be the $dk \times n2^k$ matrix obtained by concatenating $\overline{\mathsf{Eval}_i}$ for all $i \in [n]$.

In order to restrict the variables in the circuit, we will first "randomly restrict" the space of polynomials in $\mathbb{F}_n[Z]$ of degree at most $d-1$. We present the random restriction procedure in the next section.

5.1 Random Restriction Algorithm

Let $\epsilon > 0$ be any constant. We will define a randomized procedure R_ϵ which selects a subset of the variables $\{x_{i,j} \mid i, j \in [n]\}$ to set to zero.

The restriction proceeds by first restricting the space of polynomials $f \in \mathbb{F}_n[Z]$ of degree at most $d-1$. This restriction then naturally induces a restriction on the space of variables by selecting only those variables $x_{i,j}$ such that there is some polynomial f in the restricted space for which $f(i) = j$.

We restrict the space of polynomials by iteratively restricting the values the polynomials can take at points in $\mathbb{F}_{2^k}$. For each $i \in F_{2^k}$, we restrict the values f can take at i to a random affine subspace of codimension ϵk (when we view $\mathbb{F}_{2^k}$ as a k dimensional vector space over $\mathbb{F}_2$). We do this by sampling ϵk random and independent columns from $\overline{\mathsf{Eval}_i}$ and restricting the inner product of $[f]$ with these columns to be randomly chosen values. Each column that we pick in this manner imposes an $\mathbb{F}_2$-affine constraint on $[f]$, and restricts $[f]$ to vary in an affine

subspace of codimension 1. Since these random constraints for the various values of i might not be linearly independent, it is possible that at the end of the process no polynomial f satisfies the constraints. Thus we need to be more careful. We iteratively impose these random constraints for various values of i, but at the same time ensure that each new constraint that is imposed on f is linearly independent of the old constraints. We do this by making sure that each new column that is sampled is linearly independent of the old columns.

Random restriction procedure R_ϵ
Output: The set of variables that are set to zero.

1. Initialize $A_0 = \mathbb{F}_2^{kd}$, $\mathcal{B}$ to be a 0 dimensional vector, $\mathcal{M}$ to be an empty matrix over $\mathbb{F}_2$.
2. **Outer Loop :** For i from 1 to n, do the following:
   - **Inner Loop :** For j going from 1 to ϵk, do the following:
     (a) If all the columns of $\overline{\mathsf{Eval}_i}$ have been spanned by the columns in $\mathcal{M}$, then do nothing
     (b) Else pick a uniformly random column C of $\overline{\mathsf{Eval}_i}$ that has not been spanned by the columns of $\mathcal{M}$, and pick a uniformly random element b of $\mathbb{F}_2$.
     (c) Set $\mathcal{M} = \mathcal{M} \| C$ (appending C as a new column of M) and set $\mathcal{B} = \mathcal{B} \| b$ (appending b to the vector $\mathcal{B}$.
   - Set $A_i = \{[f] \mid [f] \times \mathcal{M} = \mathcal{B}; [f] \in \mathbb{F}_2^{kd}\}$
3. Let $S_0 = \{x_{i,j} \mid j \neq f(i) \; \forall \; [f] \in A_n\}$. Set all the variables $x_{i,j} \in S_0$ to 0.

The above random restriction procedure imposes at most $\epsilon k \times n$ independent $\mathbb{F}_2$-affine constraints on $[f]$. Each constraint restricts the space of possible $[f]$ by codimension 1. Thus in the end A_n is an affine subspace of $\mathbb{F}_2^{kd}$ of codimension at most $\epsilon k \times n$. This immediately implies the claim below which shows that the size of A_n is large. This in turn will imply that many of the monomials in NW_d will survive after the random restriction.

Claim 5.2. $|A_n| \geq n^d/2^{\epsilon kn} = n^{d-\epsilon n}$.

Proof. The main observation is that each time we are in Step (b) of the inner loop, we impose an *independent* $\mathbb{F}_2$-affine constraint on the possible choices of $[f]$. Thus the space of possible $[f]$ reduces by codimension exactly 1. Thus we never impose conflicting constraints on $[f]$ and we ensure that at each step the number of $[f]$ satisfying all constraints is large. □

5.2 Effect of random restriction on NW_d

Let S_0 be the set of variables output by the random restriction procedure R_ϵ. Let $R_\epsilon(NW_d)$ be the polynomial obtained from NW_d after setting the variables in S_0 to 0. In this section we will show that $R_\epsilon(NW_d)$ continues to remain hard in some sense. More precisely, we will show that for any S_0 output by the R_ϵ, and for $r < d$, a lot of distinct r^{th} order partial derivatives of $R_\epsilon(NW_d)$ are non zero.

Let $r < d-1$. Let $S \subset [n]$ be a set of size r. Let $T_S = \{\prod_{i \in S} x_{i,j_i} \mid (j_i)_{i \in S} \in [n]^r\}$ be a set of n^r monomials. We will consider partial derivatives of NW_d with respect to monomials in T_S for some choice of S.

Lemma 5.3 (Random restriction on NW_d)**.** *For every $\epsilon > 0$, and every set S_0 output by the random restriction procedure R_ϵ, there is a set $S \subset [n]$ of size r such that at least $n^{r(1-\epsilon n/d)}$ monomials in T_S are such that the partial derivative of $R_\epsilon(NW_d)$ with respect to each of these monomials is nonzero and distinct.*

Proof. Observe that for any polynomial of degree at most $d-1$, its evaluation at some d distinct points uniquely determines it. Let $S_i \in [n]$ be the set $\{(i-1)r+1, (i-1)r+2, \ldots, ir\}$. We will consider the set of evaluations of f such that $[f] \in A_n$ at points of the set S_i for various i. We will show that for some choice of i, the number of distinct sets of evaluations in S_i as $[f]$ ranges in A_n is large. Let m_i be the number of distinct r-tuples of evaluations on S_i as $[f]$ varies in A_n. Thus the total number of distinct d-tuples of evaluations on $[d]$ as $[f]$ varies in A_n is at most $\prod_{i=1}^{d/r} m_i$. However each d-tuple of evaluations on $[d]$ uniquely identifies $[f] \in A_n$. Thus $|A_n| \leq \prod_{i=1}^{d/r} m_i$. However by Claim 5.2 we know that $|A_n| \geq n^d/2^{\epsilon kn} = n^{d-\epsilon n}$. Thus there exists $i \leq d/r$ such that $m_i \geq n^{r(1-\epsilon n/d)}$. Thus there are $n^{r(1-\epsilon n/d)}$ monomials in T_{S_i} each of which is consistent with some polynomial f such that $[f] \in A_n$. Thus for each such monomial, there exists a monomial in $R_\epsilon(NW_d)$ extending it, and hence the corresponding partial derivative is nonzero. From Remark 4.4 it follows that each of these partial derivative is distinct. □

5.3 Effect of random restriction on $\Sigma\Pi\Sigma\Pi$ circuit

Let C be a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit of size at most $n^{\rho \log\log n}$ for some very small constant ρ that we will choose later. We will use $R_\epsilon(C)$ to refer to the $\Sigma\Pi\Sigma\Pi$ circuit obtained from C after setting the variables in S_0 to 0. This operation simply eliminates those monomials computed at the bottom later of C which contain at least one variable which is set to 0. Observe that homogeneity is preserved in this process. We will now show that with very high probability over the random restrictions, no product gate in C at the bottom layer which takes more than $\Omega(\log n)$ distinct variables as input survives.

Lemma 5.4 (Random restriction on $\Sigma\Pi\Sigma\Pi$ circuit)**.** *Let $\epsilon > 0$ and $\beta > 0$ be constants. Then there exists $\rho > 0$ such that if C is a $\Sigma\Pi\Sigma\Pi$ circuit of size at most $n^{\rho \log\log n}$, then with probability $> 9/10$, all the monomials computed at the bottom layer which have support at least $\beta \log n$ have some variable set to 0 by R_ϵ.*

Before we prove this lemma, we will first prove some simple results about affine subspaces and the probabilities of variables surviving the random restriction process.

Proposition 5.5. *Let V and W be fixed subspaces of $\mathbb{F}_2^k$ such that W is a subspace of V. Let U be a subspace of V which is chosen uniformly at random among all subspaces of V of dimension $\mathsf{Dim}(U)$. Then, the probability that W is a subspace of U is at most $\prod_{j=0}^{\mathsf{Dim}(W)-1} \frac{2^{\mathsf{Dim}(U)}-2^j}{2^{\mathsf{Dim}(V)}-2^j} \leq 2^{-(\mathsf{Dim}(V)-\mathsf{Dim}(U))\mathsf{Dim}(W)}$.*

Proof. Let us consider Y to be a fixed subspace of dimension $\mathsf{Dim}(U)$ of V. Now, let A_U be an invertible linear transformation from U to Y. Since, U is chosen uniformly at random, so A_U is also a uniformly random invertible matrix. Now, W was a subspace of U if and only if $A_U W$ is a subspace of Y. But since A_U is chosen uniformly at random, so $A_U W$ is a uniformly random subspace of $\mathbb{F}_2^k$ of dimension $\mathsf{Dim}(W)$. So, the desired probability is the same as the probability that for a fixed subspace Y of dimension $\mathsf{Dim}(U)$, a uniformly at random chosen subspace W of dimension $\mathsf{Dim}(W)$ lies in Y. Observe that sampling a uniformly random subspace can be done by greedily and uniformly at random sampling independent basis vectors for the subspace. Thus W is contained in Y if and only if all of the $\mathsf{Dim}(W)$ linearly independent basis vectors chosen while randomly sampling W lie in Y. This quantity is at most $\prod_{j=0}^{\mathsf{Dim}(W)-1} \frac{2^{\mathsf{Dim}(U)}-2^j}{2^{\mathsf{Dim}(V)}-2^j}$. Since, $\mathsf{Dim}(U) \leq \mathsf{Dim}(V)$, this probability is upper bounded by $2^{-(\mathsf{Dim}(V)-\mathsf{Dim}(U))\mathsf{Dim}(W)}$. □

We will now visualize our variables to be arranged in an $n \times n$ *variable matrix*, where the $(i,j)^{th}$ entry of this matrix is the variable $x_{i,j}$. We say that a monomial *survives* the random restriction procedure given by R_ϵ if no variable in the monomial is set to zero.

Definition 5.6 (Compact row)**.** *We say that the i^{th} row in the variable matrix is compact if the columns of $\mathcal{M}$ sampled by the random restriction algorithm span every column of Eval_i. Thus $\mathcal{M}$ and $\mathcal{B}$ uniquely determine the value of $f(\alpha_i)$. We say a row is non-compact otherwise.*

Proposition 5.7. *Suppose that the i^{th} row of the variable matrix is compact. Then, for every $j \in \mathbb{F}_n$, the probability that a variable $x_{i,j}$ survives R_ϵ is at most $\frac{1}{n}$.*

Proof. The columns of $\mathcal{M}$ sampled by the random restriction algorithm span every column of Eval_i, so the value of $\mathcal{B}$ uniquely determines the value of $[f] \times \mathsf{Eval}_i$. Moreover, since the columns of Eval_i are linearly independent (since for every $j \in [n]$, there exists an f such that $f(i) = j$) and $\mathcal{B}$ is chosen uniformly at random, so the value of $[f] \times \mathsf{Eval}_i$ is a uniformly random element of $\mathbb{F}_2^k$. This implies that the value of $f(i)$ is uniquely determined and is a uniformly random element of $\mathbb{F}_n$. Thus the probability that $f(i) = j$ equals $1/n$, and the result follows. □

Proposition 5.8. *Suppose that the i^{th} row of the variable matrix is non-compact. Then, for every $j \in \{1, 2, \ldots, n\}$, the probability that $x_{i,j}$ survives is at most $\frac{1}{n^\epsilon}$. In fact this holds even after conditioning on any choice of A_{i-1}, which is the affine subspace $[f]$ is allowed to vary in after $i-1$ stages on the random restriction algorithm.*

Proof. In the random restriction algorithm, since i is a non-compact row, in stage i, we picked ϵk independent columns of $\overline{\mathsf{Eval}_i}$. At the end of stage $i-1$, $[f]$ was restricted to vary in some affine subspace A_{i-1}. Thus the possible values of $f(i)$ also varied in some affine subspace V. At the end of stage i, $[f]$ was restricted to vary in some affine subspace of codimension ϵk of A_{i-1}. This affine subspace was chosen by restricting the values of f at i. Thus $[f(i)]$ was allowed to vary in a random affine subspace of codimension ϵk in V. Call this subspace U. Thus the probability that $x_{i,j}$ survives is at most the probability that j lies in the subspace U, which is at most $|U|/|V| = \frac{1}{n^\epsilon}$.

□

We will now prove that any monomial which has a large support in any row of the variable matrix survives the random restriction procedure with only a very small probability.

Lemma 5.9. *Any monomial which has a support larger than t in a row in the variable matrix survives R_ϵ with probability at most $\frac{1}{n^{\epsilon \log t}}$.*

Proof. Let α be a monomial which has support $\geq t$ in row i of the variable matrix. Let $S = \{x_{i,j_1}, x_{i,j_2}, \ldots, x_{i,j_t}\}$ be any subset of the variables in this support of size t. For $t = 1$, the lemma trivially holds. Now, if $t > 1$, then if the row i is compact then this monomial survives with probability 0. So, now we will assume that row i is non-compact. Since we identified $\mathbb{F}_n$ with $\mathbb{F}_2^k$, $\{j_i, j_2, \ldots, j_t\} \subset \mathbb{F}_2^k$. There must be $\log t$ of these elements that are linearly independent. Let this set of independent elements be $\beta_1, \beta_2, \ldots, \beta_{\log t}$. Thus α survives only if for each j, there is an f such that $[f] \in A_n$ and $f(i) = \beta_j$.

Recall that in the random restriction algorithm, in stage i, we picked ϵk independent columns of $\overline{\mathsf{Eval}_i}$. At the end of stage $i-1$, $[f]$ was restricted to vary in some affine subspace A_{i-1}. Thus the possible values of $[f(i)]$ also varied in some affine subspace V. If each of $\beta_1, \beta_2, \ldots, \beta_{\log t}$ were not contained in V then α does not survive. Thus let us assume that $\beta_1, \beta_2, \ldots, \beta_{\log t} \in V$.

At the end of stage i, $[f]$ was restricted to vary in some affine subspace of codimension ϵk of A_{i-1}. This affine subspace was chosen by restricting the values of f at i. Thus $[f(i)]$ was allowed to vary in a random affine subspace of codimension ϵk in V. Call this subspace U. Let W be the subspace given by the span of $\beta_1, \beta_2, \ldots, \beta_{\log t}$. Then $\beta_1, \beta_2, \ldots, \beta_{\log t} \in U$ if and only if $W \subseteq U$. By Lemma 5.5, the probability of this happening is at most $\frac{1}{n^{\epsilon \log t}}$.

□

Now, let us consider a monomial which has a large number of variables from different rows. We will now estimate the probability that this monomial survives.

Lemma 5.10. *Let $t < d-1$. Any monomial which has support in t non-compact rows survives R_ϵ with probability at most $\frac{1}{n^{\epsilon t}}$.*

Proof. Let α be a monomial which has at least one variable in each of t distinct non compact rows, say $i_1, i_2, i_3, \ldots, i_t$. From Lemma 5.8, we know that a variable in row i_j, $j \in [t]$, survives with probability at most $\frac{1}{n^\epsilon}$. In fact, conditioned on the variables in $i_1, i_2, \ldots, i_j$ surviving for any rows $i_1, i_2, \ldots, i_j$, the probability that the variable in row i_{j+1} survives is at most $\frac{1}{n^\epsilon}$. Hence, all of them survive with probability at most $\frac{1}{n^{\epsilon t}}$. □

We will now show that monomials which have nonzero support in many compact rows survive with very low probability.

Lemma 5.11. *Let $t < d-1$. Any monomial which has nonzero support in t compact rows survives R_ϵ with probability at most $\frac{1}{n^t}$.*

Proof. Let $i_1, i_2, \ldots, i_t$ be some t distinct compact rows. It is easy to see that the columns of the matrices $\mathsf{Eval}_{i_1}, \mathsf{Eval}_{i_2}, \ldots, \mathsf{Eval}_{i_t}$ are all linearly independent, since f can take all possible values at the points $i_1, i_2, \ldots, i_t$. Therefore, the probability that some variable survives in one of these rows is independent of the probability that some variable in another row survives. From Lemma 5.7, we know that any variable in any of these rows survives with probability at most $\frac{1}{n}$. From the above two observations, the probability that any monomial with support in these rows survives is at most $\frac{1}{n^t}$. □

Together, Lemma 5.9, Lemma 5.10 and Lemma 5.11 show that any monomial with large support survives only with a very small probability, which completes the proof of Lemma 5.4. We formally prove this below.

Proof of Lemma 5.4: From Lemma 5.9, we know that any monomial which has at least $\frac{\frac{\beta}{100}\log n}{\log\log n}$ variables in any row survives with probability at most $\frac{1}{n^{\epsilon(\log\frac{\beta}{100}+0.9\log\log n)}}$ (for n large enough). Hence, for any circuit of size at most $n^{\rho\log\log n}$, where $\rho < \epsilon/2$, by the union bound, with high probability none of the monomials which has at least $\frac{\frac{\beta}{100}\log n}{\log\log n}$ variables in any row survives.

Similarly, by Lemma 5.10, a monomial with nonzero support in at least $\log\log n$ non-compact rows survives with probability at most $\frac{1}{n^{\epsilon\log\log n}}$. Hence, for circuits of size $n^{\rho\log\log n}$, where $\rho < \epsilon/2$, with high probability none of these monomials survive.

Similarly, monomials with nonzero support in $\log\log n$ compact rows are eliminated with a very high probability if $\rho < 1/2$. Hence, at the end of any such random restriction process, with probability very close to 1, none of the surviving monomials has support larger than $\beta\log n$ if $\rho < \epsilon/2$. □

6 Lower Bounds for NW_d

In this section, we give a proof of our main theorem. We will heavily borrow from the proof of Theorem 4.10 in Section 4. The following lemma provides a lower bound on the complexity of the NW_d polynomial after restricting it via R_ϵ.

Lemma 6.1. *Let δ and ϵ be any constants such that $0 < \epsilon, \delta < 1$. Let $d = \delta n$. Let ℓ, m, r be positive integers such that $n - r > d$, $r < d-1$, $m \leq N$, $m = \theta(N)$ and for $\phi = \frac{N}{m}$, r satisfies*

$r \leq \frac{(n-d)\log\phi \pm O\left(\phi\frac{(n-d-r)^2}{N}\right)}{(1-\epsilon n/d)\log n + \log\phi}$. *Then, for every random restriction* R_ϵ,

$$\mathsf{Dim}(\langle \partial^r R_\epsilon(NW_d)\rangle_{(\ell,m)}) \geq 0.5 n^{(1-\epsilon n/d)r}\binom{N}{m}\binom{\ell-1}{m-1}$$

Proof. The proof is analogous to the proof of Lemma 4.8 till the point we substitute the value of $\mathcal{M}^{[r]}$ in the calculations in the proof of Lemma 4.8. For $R_\epsilon(NW_d)$, the value to be substituted is now $n^{r(1-\epsilon n/d)}$ as shown in Lemma 5.3. So, we know that

$$\mathsf{Dim}(\langle \partial^r R_\epsilon(NW_d)\rangle_{(\ell,m)}) \geq 0.5 n^{(1-\epsilon n/d)r}\binom{N}{m}\binom{\ell-1}{m-1}$$

as long the parameters satisfy

$$n^{r(1-\epsilon n/d)} \times \frac{(N-(n-d-r))!}{N!} \times \frac{m!}{(m-(n-d-r))!} \leq 1 \tag{5}$$

Now, using the approximation from Lemma 3.7,

$$\begin{aligned}
\log \frac{N!}{(N-(n-d-r))!} &= (n-d-r)\log N \pm O\left(\frac{(n-d-r)^2}{N}\right) \text{ and} \\
\log \frac{m!}{(m-(n-d-r))!} &= (n-d-r)\log m \pm O\left(\frac{(n-d-r)^2}{m}\right)
\end{aligned}$$

Now, taking logarithms on both sides in Equation 5 and substituting these approximations, we get

$$(1-\epsilon n/d) r \log n \leq \log\left(\frac{N}{m}\right)^{n-d-r} \pm O\left(\frac{(n-d-r)^2}{N} + \frac{(n-d-r)^2}{m}\right)$$

Substituting $m = \frac{N}{\phi}$ and noting that $\phi > 1$, we require

$$(1-\epsilon n/d) r \log n \leq (n-d-r)\log\frac{N}{m} \pm O\left(\phi\frac{(n-d-r)^2}{N}\right)$$

and

$$r \leq \frac{(n-d)\log\phi \pm O(\phi\frac{(n-d-r)^2}{N})}{(1-\epsilon n/d)\log n + \log\phi}$$

Observe that for any constant $0 < \delta < 1$ such that $d = \delta n$, r can be chosen any constant times $\frac{n}{\log n}$ by choosing ϕ to be an appropriately large constant. So, for such a choice of r, we get

$$\mathsf{Dim}(\langle \partial^r NW_d\rangle_{(\ell,m)}) \geq 0.5 n^{(1-\epsilon n/d)r}\binom{N}{m}\binom{\ell-1}{m-1}$$

□

The following lemma proves a lower bound on the top fan-in of any homogeneous $\Sigma\Pi\Sigma\Pi^{\{\beta\log n\}}$ circuit for the $R_\epsilon(NW_d)$ polynomial for a constant β. The proof of the lemma is essentially the same as the proof of Theorem 4.10.

Lemma 6.2. *Let* $d = \delta n$ *for any constant* δ *such that* $0 < \delta < 1$. *Then, there exist constants* ϵ, β *such that any homogeneous* $\Sigma\Pi\Sigma\Pi^{\{\beta\log n\}}$ *circuit computing the* $R_\epsilon(NW_d)$ *polynomial for any random restriction* R_ϵ *has top fan-in is at least* $2^{\Omega(n)}$.

Proof. By comparing the complexities of the circuit and the polynomial as given by Corollary 4.2 and Lemma 4.8, the top fan-in of the circuit must be at least

$$\frac{0.5n^{(1-\epsilon n/d)r}\binom{N}{m}\binom{\ell-1}{m-1}}{\text{poly}(nrs)\binom{n+r}{r}\binom{N}{m+rs}\binom{\ell+n-r}{m+rs}}$$

This bound holds for any choice of positive integers ℓ, m, r, a constant β such that $s = \beta \log n$ which satisfy the constraints in the hypothesis of Corollary 4.2 and Lemma 6.1. In other words, we want these parameters to satisfy

- $m + rs \leq \frac{N}{2}$
- $m + rs \leq \frac{\ell}{2}$
- $n - r > d$
- $r < d - 1$
- For $\phi = \frac{N}{m}$, $r \leq \frac{(n-d)\log\phi \pm O\left(\phi^{\frac{(n-d-r)^2}{N}}\right)}{(1-\epsilon n/d)\log n + \log\phi}$

In the rest of the proof, we will show that there exists a choice of these parameters such that we get a bound of $2^{\Omega(n)}$ from expression above. We will show the existence of such parameters satisfying the asymptotics $\ell = \theta(N)$, $r = \theta\left(\frac{n}{\log n}\right)$ and $s = \theta(\log n)$. In the rest of the proof, we will crucially use these asymptotic bounds for various approximations.

Let us now estimate this ratio term by term. We will invoke Lemma 3.7 for approximations.

- $\frac{\binom{N}{m}}{\binom{N}{m+rs}} = \frac{(N-m-rs)!(m+rs)!}{(N-m)!m!} = (\frac{m}{N-m})^{rs}$ upto some constant factors, as long as $(rs)^2 = \theta(N) = \theta(m)$.
- $\frac{\binom{\ell-1}{m-1}}{\binom{\ell+n-r}{m+rs}} = \frac{(\ell-1)!}{(m-1)!(\ell-m)!} \times \frac{(m+rs)!(\ell-m+n-r-rs)!}{(\ell+n-r)!}$. Lets now pair up things we know how to approximate within constant factors. $\frac{\binom{\ell-1}{m-1}}{\binom{\ell+n-r}{m+rs}} = \frac{(\ell-1)!}{(\ell+n-r)} \times \frac{(m+rs)!}{(m-1)!} \times \frac{(\ell-m+n-r-rs)!}{(\ell-m)!} =$ poly(n) $\times \frac{1}{\ell^{n-r}} \times m^{rs} \times \frac{(\ell-m)^{n-r}}{(\ell-m)^{rs}}$. This simplifies to poly(n) $\times \left(\frac{m}{\ell-m}\right)^{rs} \times \left(\frac{\ell-m}{\ell}\right)^{n-r}$.
- $\frac{n^{(1-\epsilon n/d)r}}{\binom{n+r}{r}} \geq \frac{n^{(1-\epsilon n/d)r}}{\left(\frac{2(n+r)}{r}\right)^r}$. We just used Stirling's approximation here.

In the asymptotic range of our parameters, the approximations above imply that the top fan-in, up to polynomial factors is at least

$$\left(\frac{r}{3}\right)^r \times \left(\frac{m}{\ell-m}\right)^{rs} \times \left(\frac{\ell-m}{\ell}\right)^{n-r} \times \frac{1}{n^{(\epsilon n/d)r}} \times \left(\frac{m}{N-m}\right)^{rs}$$

Simplifying further, this is at least

$$2^{\Omega(r\log r - rs\log\frac{\ell-m}{m} - (n-r)\log\frac{\ell}{\ell-m} - (\epsilon n/d)r\log n - rs\log\frac{N-m}{m})}$$

We will set m and ℓ to be $\theta(N)$ and r to be $\theta(\frac{n}{\log n})$. The constants have to be chosen carefully in order to satisfy the constraints. We will choose constants α, β and η such that $s = \beta \log n$, $r = \alpha \cdot n/\log n$ and $m = \eta\ell$. First let us choose ϵ to be a very small positive constant such that $\epsilon n/d = \epsilon/\delta << 0.1$ First choose η to be any small constant > 0 (for instance $\eta = 1/4$). Now, choose α to be a constant much much larger than $\log\frac{1}{1-\eta}$ and ϵ/δ. This makes sure that $r\log r$ dominates $(n-r)\log\frac{\ell}{\ell-m}$ and $(\epsilon n/d)r\log n$. Recall that α can be chosen to be any large constant by choosing ϕ to be appropriately large constant (by the constraint between r and ϕ in the fifth bullet). Notice that this sets m to be a small constant factor of N. Fix these choices of η and α.

Now, we choose the term β to be a small constant such that $rs \log \frac{1-\eta}{\eta}$ and $rs \log \frac{N-m}{m}$is much less than $r \log r$. Observe that this choice of parameters satisfies all the constraints imposed in the calculations above. Hence, the top fan-in must be at least $2^{\Omega(r \log r)} = 2^{\Omega(n)}$.

□

We now have all the ingredients to prove our main theorem.

Theorem 6.3. *Let $d = \delta n$ for any constant δ such that $0 < \delta < 1$. Any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing the NW_d must have size at least $n^{\Omega(\log \log n)}$.*

Proof. For every value of δ, such that $0 < \delta < 1$, choose the parameters $\epsilon = \tilde{\epsilon}, \beta = \tilde{\beta}$ such that Lemma 6.2 is true for $\tilde{d} = \delta n$. Now, let us choose a constant $\rho = \tilde{\rho}$ such that Lemma 5.4 holds. Now, let C be a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing the $NW_{\tilde{d}}$ polynomial. If the number of bottom product gates of C was at least $n^{\tilde{\rho} \log \log n}$, then C has large size and we are done. Else, let us now apply a random restriction R_ϵ to the circuit. By the choice of parameters, Lemma 5.4 holds and so with probability 0.9 every bottom product gate in C with support larger than $\tilde{\beta} \log n$ is set to zero. After a restriction, the circuit computes $R_{\tilde{\epsilon}}(NW_{\tilde{d}})$. So, now we are in the case when we have a small support homogeneous circuit of depth four computing some random restriction of the $NW_{\tilde{d}}$ polynomial and then, by Lemma 6.2 above, the top fan-in of $R_{\tilde{\epsilon}}(C)$ must be at least $2^{\Omega(n)}$. Hence, any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing $NW_{\tilde{d}}$ must have size at least $n^{\Omega(\log \log n)}$. □

7 Open Problems

The main question left open by this work is to prove much stronger, possibly exponential lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits. Given the earlier related works and the results of this paper, this question might be well within reach. It would be also very interesting to understand the limits of the new complexity measure of bounded support shifted partial derivatives that is introduced in this paper (as well as other variants) and investigate if they can be used to prove lower bounds for other interesting classes of circuits.

Acknowledgments

We would like to thank Mike Saks and Avi Wigderson for many helpful discussions and much encouragement. We are also thankful to Amey Bhangale, Ben Lund and Nitin Saurabh for carefully sitting through a presentation on an earlier draft of the proof.

References

[AV08] M. Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *Proceedings of the 49th Annual FOCS*, pages 67–75, 2008.

[CM13] Suryajith Chillara and Partha Mukhopadhyay. Depth-4 lower bounds, determinantal complexity: A unified approach. *CoRR*, abs/1308.1640v3, 2013.

[FLMS13] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:100, 2013.

[GKKS13] A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi. Approaching the chasm at depth four. *In Proceedings of CCC*, 2013.

[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.

[Koi12] P. Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theoretical Computer Science*, 448:56–65, 2012.

[KS13a] Mrinal Kumar and Shubhangi Saraf. The limits of depth reduction for arithmetic formulas: It's all about the top fan-in. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:153, 2013.

[KS13b] Mrinal Kumar and Shubhangi Saraf. Lower bounds for depth 4 homogenous circuits with bounded top fanin. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:68, 2013.

[KSS13] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:91, 2013.

[NW95] N. Nisan and A. Wigderson. Lower bounds on arithmetic circuits via partial derivatives. In *Proceedings of the 36th Annual FOCS*, pages 16–25, 1995.

[Raz10] Ran Raz. Elusive functions and lower bounds for arithmetic circuits. *Theory of Computing*, 6(1):135–177, 2010.

[RY08] R. Raz and A. Yehudayoff. Lower bounds and separations for constant depth multilinear circuits. In *Conference on Computational Complexity, 2008.*, pages 128–139, june 2008.

[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 *Proceedings of the 11th Annual STOC*, STOC '79, pages 249–261, New York, NY, USA, 1979. ACM.

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