LOWER BOUNDS FOR BOUNDED DEPTH ARITHMETIC CIRCUITS

BY MRINAL KUMAR

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

Swastik Kopparty and Shubhangi Saraf

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2017

ABSTRACT OF THE DISSERTATION

Lower bounds for bounded depth arithmetic circuits

by Mrinal Kumar

Dissertation Director: Swastik Kopparty and Shubhangi Saraf

Proving lower bounds for arithmetic circuits is a problem of fundamental importance in theoretical computer science. In recent years, an approach to this problem has emerged via the *depth reduction* results of Agrawal and Vinay [AV08], which show that strong enough lower bounds for extremely structured bounded depth circuits (even homogeneous depth-4 circuits) suffice for general arithmetic circuits lower bounds. In this dissertation, we study homogeneous depth-4 and homogeneous depth-5 arithmetic circuits with a view towards proving strong lower bounds, and understanding the optimality of the depth reduction results. Some of our main results are as follows.

- We show a hierarchy theorem for bottom fan-in for homogeneous depth-4 circuits with bounded bottom fan-in. More formally, we show that there for a wide range of choice of parameter t, there is a homogeneous polynomial in n variables of degree $d = n^{\Theta(1)}$ which can be computed by a homogeneous depth-4 circuit of bottom fan-in t, but any homogeneous depth-4 circuit of bottom fan-in at most $t/20$ must have top fan-in $n^{\Omega(d/t)}$.

- We show that there is an explicit polynomial family such that any homogeneous depth-4 arithmetic circuit computing it must have super-polynomial size. These were the first superpolynomial lower bounds for homogeneous depth-4 circuits

with no restriction on top or bottom fan-in. Simultaneously and independently, a similar lower bound was also proved by Kayal et al [KLSS14b].

- We show that any homogeneous depth-4 circuit computing the iterated matrix multiplication polynomial in n variables and degree $d = n^{\Theta(1)}$ must have size at least $n^{\Omega(\sqrt{d})}$. This shows that the upper bounds of depth reduction from general arithmetic circuits to homogeneous depth-4 circuits are almost optimal, up to a constant in the exponent. Moreover, these were the first $n^{\Omega(\sqrt{d})}$ lower bounds for homogeneous depth-4 circuits over all fields. Prior to our work, Kayal et al. [KLSS14a] had shown such a lower bound over the fields of characteristic zero.

- We show that there is a family of polynomials in n variables and degree $d = O(\log^2 n)$ which can be computed by linear size homogeneous depth-5 circuits and polynomial size non-homogeneous depth-3 circuits but require homogeneous depth-4 circuits of size $n^{\Omega(\sqrt{d})}$. In addition to indicating the power of increased depth, and non-homogeneity, these results also show that for the range of parameters considered here, the upper bounds for the depth reduction results [AV08, Koi12, Tav15] are close to optimal in a very strong sense : a general depth reduction to homogeneous depth-4 circuits of size $n^{o(\sqrt{d})}$ is not possible even for homogeneous depth-5 circuits of linear size.

- We show an exponential lower bound for homogeneous depth-5 circuits computing an explicit polynomial over all finite fields of constant size. For any non-binary field, these were the first such super-polynomial lower bounds, and prior to our work, even cubic lower bounds were not known for homogeneous depth-5 circuits.

On the way to our proofs, we study the complexity of some natural polynomial families (for instance, homogeneous depth-4, depth-5 circuits, iterated matrix multiplication) with respect to many existing partial derivative based complexity measures, and also define and analyze some new variants of these measures [KS14, KS15b].

Acknowledgements

Heartfelt thanks to my advisers Swastik and Shubhangi, for their time, their patience and the freedom to explore problems at my own pace. Their guidance has been invaluable. Apart from their interest and obvious involvement in my research, I am also grateful to them for the delightful classes they have taught over the last five years.

A part of this thesis is based on joint work with Ramprasad, most of which was done over email. Working with him has been pure joy, and I have learned a great deal about arithmetic circuits, complexity theory (and also the best ice-cream places in Tel Aviv!) from him. I am deeply thankful for all of it.

Among the most cherished memories of my time in grad school would be the conversations with Mike Saks. Mike has been generous with his time and with his ideas, and his visible enthusiasm during meetings has often been a source of much encouragement. His ability to interpret, rationalize and explain clever and complicated ideas to an extent that they seemed an almost inevitable step in the chain of thought, is the one superpower I would really like to have! I am also thankful to everyone in the theory reading group. I saw, learnt and read about things which I would not have otherwise.

I am grateful to Eric Allender and Ran Raz, for agreeing to be on my committee and for their guidance and to our graduate director William Steiger, graduate secretary Carol DiFrancesco, Maryann Holtsclaw, Ginger Olszewski and the rest of the department staff for their help in navigating the bureaucratic labyrinth.

Since my undergraduate days, I have often found great joy in Roman Smolensky's papers. From his choice of problems to his often elegant solutions, his work has perhaps been the most significant influence in shaping my own research interests. I am grateful to Jayalal for introducing me to them. I am also deeply thankful to Sounaka and to Saket for my first steps in theory.

During my time here, I have had the chance to interact with and work with some amazing people, and I am thankful to Pranjal Awasthi, Susmita Biswas, Arkadev Chattopadhyay, Suryajith Chillara, Michael Forbes, Ankit Garg, Josh Grochow, Ankit Gupta, Prahladh Harsha, Pooya Hatami, Neeraj Kayal, Vineet Nair, Rafael Oliveira, Ran Raz, Shadab Romani, Chandan Saha, Nitin Saurabh, Amir Shpilka, Madhu Sudan, Sébastien Tavenas, Ben Lee Volk, Avi Wigderson, Noga Ron-Zewi and others for many insightful conversations. I have also had the pleasure of being on multiple enjoyable academic visits. I am grateful to Neeraj and Chandan for hosting me in Bangalore on multiple occasions, to Madhu for hosting me at MSR-New England, to Amir for the visit to Tel-Aviv, and to Arkadev for the visit to TIFR.

My stay here would have been much less enjoyable if not for the company I had. Many thanks to Abdul, Abhishek, Aditya, Amey, Ben and Fei, and to Amruta and Simao for the many delightful conversations and for dragging me out of my home and office on many occasions. Cheers to my friends Abhishek, Adarsh, Amit, Apurva, KK, Nidhi, Nishant, Swati and Venkata for the many discussions about life, cricket, food, politics and Calvin & Hobbes.

I certainly owe one to Swati for her affection and her unwavering support. Her presence through sunshine and through rain, has been one of the invariants of life through the last few years.

I feel deeply indebted to my parents, my grandparents, my sister and rest of my family. They have been a strong source of motivation and support for most of my endeavors.

This thesis is dedicated to Dr. J. K. Singh, my math teacher in high school and a dear friend and a mentor since then. His skill and his enthusiasm during his classes, which were often interspersed with delightful anecdotes and stories brought about a major shift in how I felt about math. His optimism in life and his continued interest in my well being have been an incessant source of encouragement.

Dedication

To Dr. J. K. Singh - my math teacher in high school.

Table of Contents

Chapter 1

Introduction

1.1 Arithmetic circuits

Multivariate polynomials are perhaps the one of the most well studied mathematical objects in computer science. In addition to being interesting objects on their own, they have found surprising applications in a variety of areas of computer science like algorithm design, computational complexity, pseudorandomness and design of error correcting codes (for instance, see [AKS04, Bjö14, MVV87, RS60, Smo87]). Algebraic complexity theory is the study of computational questions arising in the formal computation of such polynomials using field operations $\{+, \times\}$ over an underlying field $\mathbb{F}$. A very natural computational model for computing multivariate polynomials, and the subject of study of this thesis is the model of arithmetic circuits, which we now define.

Definition 1.1 (Arithmetic circuits)**.** *An arithmetic circuit over a field $\mathbb{F}$ and variables $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ is a directed acyclic graph with nodes labelled by $+$ and $\times$ operations over $\mathbb{F}$ and leaves (nodes of in-degree 0) labelled by elements of $\mathbb{F}$ and $\mathbf{x}$. The circuit computes an n variate polynomial in $\mathbb{F}[\mathbf{x}]$ in the following natural way. A leaf node computes the polynomial which is equal to its label. A $+$ node computes the sum of polynomials computed at the nodes feeding into it, and a $\times$ node computes a product of polynomials computed at the nodes feeding into it. The size of the circuit is the number of edges in the circuit, and the depth of the circuit is the length of the longest path from an output (a node with out-degree 0) to a leaf.* ◊

Remark 1.2. *Sometimes, arithmetic circuits are defined so that the edges feeding into a sum gate have weights given by elements of the underlying field. A directed edge from*

u to v with an edge weight w just amounts to multiplying the polynomial computed at u by the scalar w before it is fed into v. The results in this thesis continue to hold for this general definition. However, for the rest of this thesis, we will stay with Definition 1.1.

$\Diamond$

The nodes feeding into a node v in the circuit are referred to as the children of v. Unless otherwise stated, the circuits considered in the thesis will have a single output gate.

An important parameter associated to a circuit, is the notion of the *formal degree* of C, denoted by Formal degee(C). It is defined inductively as follows.

- The formal degree of an input node which is labelled by a variable x_i is defined to be 1.
- The formal degree of an input node which is labelled by a field element is defined to be 0.
- The formal degree of a + gate equals the maximum of the formal degrees of its children.
- The formal degree of a $\times$ gate equals the sum of the formal degrees of its children.

Note that the formal degree of a circuit C which computes a polynomial P, must be at least as large as the degree of the polynomial P. However, the formal degree of C could be much larger than the degree of P. A circuit C is said to be homogeneous if the polynomial computed at every gate in the circuit is a homogeneous polynomial. Else, C is said to be non-homogeneous.

1.2 VP, VNP and arithmetic circuit lower bounds

In a seminal paper in 1979, Valiant [Val79] laid out a formal theoretical framework for the study of resource bounded algebraic computation and defined the complexity classes VP and VNP[1], which can be viewed as non-uniform algebraic analogs of P and

[1]See Definition 1.3 and Definition 1.4 respectively.

NP respectively. Since then, the problem of understanding whether VNP is different from VP has been a problem of fundamental significance in Algebraic complexity theory.

In spite of the fundamental nature of this problem, not much progress has been made on the question of proving strong arithmetic circuit lower bounds for explicit polynomials. The best known lower bound for general arithmetic circuits is a bound of $\Omega(n \log d)$ for the polynomial $x_1^d + x_2^d + x_3^d + \ldots + x_n^d$, which was proved by Strassen [Str73] and Baur-Strassen [BS83] more than three decades ago. The absence of much progress on the general question has led to much interest on the question of proving lower bounds for restricted and more structured classes of arithmetic circuits. Arithmetic formula, non-commutative circuits, bounded depth circuits, multilinear formulas and monotone arithmetic circuits are some restricted classes of arithmetic circuits which have been studied from this point of view, and for many of these classes substantial progress has been made on the question of proving lower bounds. For instance, some notable results in this direction are the quadratic lower bounds for arithmetic formula by Kalorkoti [Kal85], superpolynomial lower bound for monotone arithmetic circuits by Jerrum and Snir [JS82], superpolynomial lower bounds for multilinear formulas by Raz [Raz09, Raz06]. We refer the reader to the surveys of Shpilka-Yehudayoff [SY10] and Saptharishi [Sap15] and the references therein for the formal definition of these models, and for an overview of these results.

One interesting class of arithmetic circuits, studied from this perspective are the class of bounded depth arithmetic circuits. Even for such circuits, we really only understand lower bounds for depth 2 circuits and some classes of depth 3 and depth 4 circuits [NW97, SW01, GK98, GKKS14, KSS14]. We formally define arithmetic circuits in the following section. We end this section by defining the complexity classes VP and VNP.

Definition 1.3 (VP)**.** *A sequence* $\{P_n(x_1, x_2, \ldots, x_n)\}$ *of polynomials is said to be in* VP *if there exists a sequence* $\{C_n(x_1, x_2, \ldots, x_n)\}$ *of arithmetic circuits of size and formal degree bounded by* $\mathrm{poly}(n)$ *such that*

$$C_n(x_1, x_2, \ldots, x_n) \equiv P_n(x_1, x_2, \ldots, x_n).$$

Here, $\equiv$ refers to equality as formal polynomials. $\Diamond$

Definition 1.4 (VNP)**.** *A sequence $\{P_n(x_1, x_2, \ldots, x_n)\}$ of polynomials is said to be in* VNP *if there exists a univariate polynomial h and a sequence $\{Q_n(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_{h(n)})\}$ of polynomials in* VP*, such that for every n*

$$P_n(x_1, x_2, \ldots, x_n) \equiv \sum_{(a_1, a_2, \ldots, a_{h(n)}) \in \{0,1\}^{h(n)}} Q_n(x_1, x_2, \ldots, x_n, a_1, a_2, \ldots, a_{h(n)}).$$

$\Diamond$

Throughout this thesis, we say that a family of polynomials is explicit if the family lies in VNP. However, a more tangible sufficient condition for explicitness which is also easier to work with in many cases is given by the following lemma of Valiant [Val79].

Lemma 1.5 (Valiant's criterion)**.** *Let $\{P_n\}$ be a polynomial family, where for every n, P_n is a multilinear polynomial in n variables with $0, 1$ coefficients. Morever, let there be a function Φ which takes as input n and a vector $(a_1, a_2, \ldots, a_n) \in \{0,1\}^n$ and outputs the coefficient of the monomial $x_1^{a_1} \cdot x_2^{a_2} \cdots x_n^{a_n}$ in P_n. If Φ can be computed in deterministic polynomial time, then the polynomial family $\{P_n\}$ is in* VNP*.*

Remark 1.6. *For the rest of this thesis, for brevity we will sometime just say a polynomial, when we mean a family of polynomials.* $\Diamond$

1.3 Bounded depth arithmetic circuits

The depth of an arithmetic circuit is the length of the longest path from an output gate to an input gate. An arithmetic circuit of depth Δ is defined as follows.

Definition 1.7 (Bounded depth circuits)**.** *An arithmetic circuit C of depth Δ is an arithmetic circuit with Δ layers of alternating $+$ and $\times$ gates of unbounded fan-in, with the output gate being a $+$ gate. The output gate is said to be the first layer of C, and a node in the i^{th} layer has all its children in $i + 1^{th}$ layer.* $\Diamond$

A layer with $+$ gates is denoted by Σ and a layer with $\times$ gates is denoted by Π. Thus, a depth-2 circuit is a $\Sigma\Pi$ circuit, a depth-3 circuit is a $\Sigma\Pi\Sigma$ circuit, a depth-4 circuit is a $\Sigma\Pi\Sigma\Pi$ circuit and so on. We remark that for bounded depth circuits,

we must allow the gates in the circuit to have unbounded fan-in, since otherwise, the output cannot even depend on all the inputs.

Observe that lower bounds for depth-2 arithmetic circuits is obvious since such circuits compute a polynomial by separately computing every monomial in the polynomial and taking a sum. Thus, any polynomial with superpolynomially many monomials with non-zero coefficients requires a *large* depth-2 arithmetic circuit. The question of lower bounds for depth-3 arithmetic circuits already turns out to be non-trivial. In fact, even as of today, this problem is not very well understood. For the case of homogeneous depth-3 arithmetic circuits, Nisan and Wigderson [NW97] proved the first exponential lower bounds. This much celebrated paper used the notion of partial derivatives of a polynomial as a measure of its complexity, which found many further applications. For the case of non-homogeneous depth-3 circuits over constant sized finite fields, exponential lower bounds were proved by Grigoriev and Karpinski [GK98] and Grigoriev and Razborov [GR00]. Over large fields, the best known lower bounds continue to be cubic [KST16a].

Perhaps the most natural question following the work of Nisan and Wigderson [NW97] is to prove super-polynomial lower bounds for homogeneous depth-4 arithmetic circuits. However, this problem remained open for more than a decade. In a beautiful structural result, Agrawal and Vinay [AV08] (and subsequent optimisations by Koiran [Koi12] and Tavenas [Tav15]) showed that strong enough lower bounds for homogeneous depth-4 arithmetic circuits, even with bottom fan-in 2 (denoted by $\Sigma\Pi\Sigma\Pi^{[2]}$) would imply $\mathsf{VP} \neq \mathsf{VNP}$. More formally, they showed the following.

Theorem 1.8 ([AV08, Koi12, Tav15])**.** *Let $\mathbb{F}$ be any field. Let P be a homogeneous degree d polynomial in n variables over $\mathbb{F}$ which can be computed an arithmetic circuit of size s. Then, for every t such that $0 \leq t \leq d$, P can also be computed by a homogeneous $\Sigma\Pi\Sigma\Pi$ arithmetic circuit of bottom fan-in t and size $s^{O(t+d/t)}$.*

In fact, the top fan-in of the homogeneous depth-4 circuit with bottom fan-in t (referred to as a homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit) is at most $s^{O(d/t)}$. Thus, proving strong enough size (or top fan-in) lower bound for homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits computing P

is sufficient to imply super-polynomial lower bound on the size of *any* arithmetic circuit which computes P. Thus, Theorem 1.8 establishes the question of lower bounds for bounded depth arithmetic circuits as a plausible approach to the VP vs VNP question!

Remark 1.9. *Even though Theorem 1.8 as stated here is for depth reduction to homogeneous depth four arithmetic circuits, the result is more general and shows that any homogeneous polynomial P of degree d in n variables in* VP *can be computed by an arithmetic circuit of depth 2Δ of size at most $n^{O(\frac{d}{\Delta})}$.* $\Diamond$

1.4 Contributions of this thesis

The work in this thesis is primarily motivated by the the question of proving super-polynomial lower bounds for arithmetic circuits of small depth. In particular, we study the question of proving strong lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits, homogeneous $\Sigma\Pi\Sigma\Pi$ circuits and homogeneous $\Sigma\Pi\Sigma\Pi\Sigma$ circuits, and attempt to understand if the parameters of depth reductions in Theorem 1.8 can be improved. The results included in this thesis are based on the results in the papers [KS15d, KS14, KS17, KS15b, KS16]. We now summarize the main contributions of each of the chapters, including the results, their context in the light of previous work, and key technical contributions.

Chapter 2 : Lower bounds for homogeneous depth-4 circuits with bounded bottom fan-in [KS15d]

In Chapter 2, we prove the following results.

1. We show a superpolynomial lower bound on the size of homogeneous depth-4 arithmetic circuits with top fan-in $o(\log n)$. We show this by essentially showing a *non-trivial* depth reduction from homogeneous depth-4 arithmetic circuits with top fan-in $o(\log n)$ to homogeneous depth-4 circuits with bounded bottom fan-in.

2. We complement this strategy for the lower bound by showing that if the top fan-in is allowed to be $\log n$, then there is no depth reduction to homogeneous depth-4 arithmetic circuits of bottom fan-in t of size $n^{o(d/t)}$; thereby showing that the bounds obtained by Tavenas [Tav15] are tight for this model.

3. As a consequence of our results, it follows that there is a degree d n-variate polynomial which can be computed by a $poly(n)$ sized arithmetic formula, but any *regular* arithmetic formula computing it must have size at least $n^{\Omega(\log d)}$.

In the breakthrough results of Gupta, Kamath, Kayal and Saptharishi [GKKS14], and Kayal, Saha and Saptharishi [KSS14] an $n^{\Omega(\sqrt{d})}$ lower bound for homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuits was shown. These results, together with the depth reduction results [AV08, Koi12, Tav15], imply that superpolynomial lower bounds would follow for any class of arithmetic circuits which can be reduced to homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuits of size $n^{o(\sqrt{d})}$. Thus, it is extremely interesting to ask if certain important circuit classes, for instance homogeneous formulas, constant depth arithmetic circuits, or even homogeneous $\Sigma\Pi\Sigma\Pi$ circuits (with unbounded bottom fan-in) have this property.

The results in this chapter make concrete progress towards answering these questions. They show that even for very simple circuit classes like homogeneous depth-4 circuits with unbounded bottom fan-in, improved depth reduction to homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuits of size $n^{o(\sqrt{d})}$ is not possible for a wide range of parameters.

Chapter 3 : Superpolynomial lower bounds for homogeneous depth-4 circuits [KS14]

In Chapter 3, we prove the following theorem.

Theorem 1.10. *There is a family $\{P_n\}$ of polynomials in VP, where P_n is a homogeneous polynomial of degree $d = n^{O(1)}$ on n variables, such that any homogeneous depth-4 arithmetic circuit computing P_n has size at least $n^{\Omega(\log\log n)}$.*

Theorem 1.10 provided the first superpolynomial lower bound for homogeneous depth-4 arithmetic circuits, with no restriction on bottom or top fan-in. Independently and almost simultaneously a stronger lower bound of $n^{\Omega(\log d)}$ was also shown by Kayal et al [KLSS14a]. This result extends the work of Nisan-Wigderson [NW97] (which showed superpolynomial lower bounds for homogeneous depth 3 circuits), Gupta-Kamath-Kayal-Saptharishi and Kayal-Saha-Saptharishi [GKKS14, KSS14] (which showed superpolynomial lower bounds for homogeneous depth 4 circuits with bounded bottom

fan-in), Kumar-Saraf [KS15d] (which showed superpolynomial lower bounds for homogeneous depth 4 circuits with bounded top fan-in) and Raz-Yehudayoff and Fournier-Limaye-Malod-Srinivasan [RY09, FLMS14] (which showed superpolynomial lower bounds for multilinear depth-4 circuits).

The key tool here is the use of random restrictions to reduce the homogeneous depth-4 circuit to a homogeneous depth-4 circuit with bounded bottom support. Finally, we prove a superpolynomial lower bound for homogeneous depth-4 arithmetic circuits with bounded bottom support. This step needed an appropriate variant of the method of shifted partials, which we call shifted partials with bounded support.

Chapter 4 : Exponential lower bounds for homogeneous depth-4 circuits [KS17]

In Chapter 4, we prove the following theorem.

Theorem 1.11. *There is a family $\{P_n\}$ of polynomials in VP, where P_n is a homogeneous polynomial of degree $d = n^{O(1)}$ on n variables, such that any homogeneous depth-4 arithmetic circuit computing P_n has size at least $n^{\Omega(\sqrt{d})}$.*

Since the family of hard polynomials here is in VP, it follows that the parameters of depth reduction results to homogeneous depth-4 arithmetic circuits, as given by Tavenas [Tav15] are optimal up to constants in the exponent.

Following the results in [KS14], Kayal et al. [KLSS14a] had shown a lower bound of $n^{\Omega(\sqrt{d})}$ for homogeneous depth-4 arithmetic circuits over fields of characteristic zero. Their hard polynomial was a family of homogeneous degree d polynomials in n-variables in VNP. Thus, any asymptotic improvement in the exponent, in either the lower bound or the depth reduction results would have sufficed to separate VP and VNP. The results in this chapter show that VP cannot be depth reduced to a homogeneous depth-4 circuits of size $n^{o(\sqrt{d})}$, there by settling an important research direction. Moreover, the lower bounds in this paper hold over all fields, as opposed to the lower bounds of Kayal et al. [KLSS14a], which do not hold over fields of small characteristic.

The main technical content of this chapter is an argument which shows a lower bound on the rank of the matrix of projected shifted partial derivatives of the iterated matrix

multiplication polynomial under random restrictions. Our argument is combinatorial (hence, is field independent), and works by showing that there is a large upper triangular submatrix in this matrix.

Chapter 5 : Finer separations between shallow circuits [KS16]

In Chapter 5, we prove the following theorem.

Theorem 1.12. *There is a family of polynomials* $\{P_n\}$*, where* P_n *is of degree* $d = \Theta(\log^2 n)$ *in* n *variables such that*

1. P_n *can be computed by a homogeneous depth-*5 *arithmetic circuit of size* $O(n)$*.*
2. P_n *can be computed by a non-homogeneous depth-*3 *arithmetic circuit of size* $\text{poly}(n)$*.*
3. *Any homogeneous depth-*4 *arithmetic circuit computing* P_n *must have size* $n^{\Omega(\sqrt{d})}$*.*

Even though such a sharp separation between depth-5 and depth-4 circuits was expected to be true, prior to Theorem 1.12, it wasn't even known if arithmetic formulas could be depth reduced to homogeneous depth-4 arithmetic circuits of size $n^{o(\sqrt{d})}$. In the low degree regime, this result shows that even simple classes homogeneous depth-5 arithmetic circuits of linear size can not be depth reduced to homogeneous depth-4 circuits of size $n^{o(\sqrt{d})}$. This improves the following results.

1. For homogeneous depth-4 circuits with *bounded bottom fan-in*, such results were already shown in [KS15d]. Here, we get rid of the bounded bottom fan-in restriction.
2. For homogeneous depth-4 circuits, it was shown in [KS17] that depth reductions from algebraic branching programs to homogeneous depth-4 circuits was optimal up to constants in the exponent. Thus, the results in this chapter are a refinement of the results in [KS17].

The proofs in this chapter also give an alternative and much simpler proof of an $n^{\Omega(\sqrt{d})}$ lower bound for homogeneous depth-4 circuits, albeit in a low degree regime.

The key idea for the proofs in this chapter is to design a complexity measure which tries to use the fact that the hard polynomial family for many of the known homogeneous depth-4 lower bounds are set multilinear. This appears to be a much much stricter restriction than just multilinearity. We exploit this intuition by working with shifted partials modulo an appropriate ideal as a complexity measure. For this measure, we are able to show that the read once regular depth-5 arithmetic circuit [2] has a large measure. In the course of our proof, we also recover a hierarchy theorem for bounded bottom support (as opposed to fan-in) for homogeneous depth-4 circuits.

Chapter 6 : Exponential lower bound for homogeneous depth-5 circuits over small fields [KS15b]

In Chapter 6, we prove the following theorem.

Theorem 1.13. *There is an explicit family of polynomials* $\{P_d : d \in \mathbb{N}\}$*, with* $\mathsf{Deg}(P_d) = d$*, in the class* VNP *such that for any finite field* $\mathbb{F}_q$ *such that* $q = O(1)$*, any homogeneous depth-*5 *circuit computing* P_d *must have size* $\exp(\Omega(\sqrt{d}))$.

Prior to Theorem 1.13, no superpolynomial lower bounds were known for homogeneous depth-5 arithmetic circuits over any field apart from $\mathbb{F}_2$. Over non-binary fields, the best known lower bounds for depth-5 circuits were the superlinear lower bounds due to Raz [Raz10a] and quadratic lower bounds due to Kalorkoti [Kal85] [3].

The main technical contribution of this chapter is the idea of looking at the space of shifted partial derivatives of a polynomial as a space of functions over $\mathbb{F}^n$, where $\mathbb{F}$ is the underlying finite field. We show that the complexity of the circuit with respect to this *functional* shifted partials measure is non-trivially small. Another key idea is a much tighter and cleaner analysis of the dimension of projected shifted partials of an explicit polynomial over all fields. This refines and provides a much more modular proof of such a bound as shown in [KS17]. Since our work, these two ideas have found further applications in new results on functional lower bounds for arithmetic circuits [FKS16].

---

[2] The algebraic analog of the Sipser function.

[3] Kalorkoti's lower bound requires the circuit to be a formula.

We conclude with some open problems in Chapter 7.

Chapter 2

Limits of depth reduction for arithmetic formulas : it's all about the top fan-in[1]

2.1 Introduction

In the last few years, a very promising and exciting new framework for proving lower bounds for arithmetic circuits has emerged. The framework consists of two major components. Let $\mathcal{C}$ be the class of circuits one wants to prove lower bounds for. The first step is to show that any circuit in $\mathcal{C}$ can be efficiently *depth reduced* to a depth-4 circuit with bounded bottom fan-in ($\Sigma\Pi\Sigma\Pi^{[t]}$ circuit). This depth reduction procedure was introduced and developed in the works of Agrawal-Vinay [AV08], Koiran [Koi12] and Tavenas [Tav15], building upon the initial depth reduction procedure of Valiant et al. [VSBR83]. The second step is to prove strong lower bounds for $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits using the *shifted partial derivative* complexity measure, which was developed in the works of Kayal [Kay12] and Gupta et al. [GKKS14]. This framework was used successfully to prove the first superpolynomial lower bounds for *regular formulas* [KSS14], and it seemed promising that such techniques could be used to prove lower bounds for more general classes such as *general* arithmetic formulas.

In this chapter, we successfully apply this framework to prove the first superpolynomial lower bounds for homogeneous depth-4 circuits with bounded top fan-in. We prove our results via an *improved depth reduction*[2]. We also show that if the bound on the top

[1] The results in this chapter appear in [KS15d].

[2] By depth reduction in this chapter, we really mean a reduction to homogeneous depth-4 circuits with bounded bottom fan-in. So, it makes sense to talk of depth reduction for depth 4 circuits.

fan-in is relaxed (even by a small amount), then *efficient depth reduction is unlikely.* In particular this suggests that the method of improved depth reduction + shifted partial derivatives may not be powerful enough to prove lower bounds for (even) homogeneous arithmetic formulas. This result strengthens the results in [KSS14, FLMS14] and answers some of the main open questions posed in them.

We now outline the major results and the sequence of events that build up to the results of this chapter.

In the discussion in the rest of this section, we will refer to the class of circuits of depth 4 ($\Sigma\Pi\Sigma\Pi$ circuits) with bottom (product) fan-in bounded by t as $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits.

Depth-4 Lower Bounds and VNP vs VP

In light of the results of Agrawal-Vinay [AV08], Koiran [Koi12] and Tavenas [Tav15], proving lower bounds for homogeneous *depth*-4 circuits seems like an extremely promising direction to pursue in order to separate VNP from VP. In a breakthrough result in this direction, Gupta, Kamath, Kayal and Saptharishi [GKKS14] proved that any homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuit computing the permanent of an $n \times n$ generic matrix must have size (and top fan-in) $\exp(\sqrt{n})$. This was strengthened in a follow up work of Kayal, Saha and Saptharishi [KSS14], where it was shown that there is an explicit family $\{P_n\}$ of polynomials in VNP, where P_n is a homogeneous polynomial of degree d in n variables, such that any homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuit computing it must have size (and top fan-in) at least $\exp(\Omega(\sqrt{d}\log n))$. More precisely,

Theorem 2.1 ([GKKS14, KSS14])**.** *There is a family $\{P_n\}$ of polynomials in* VNP*, where P_n is a homogeneous polynomial of degree d in $n = \Theta(d^2)$ variables, such that any homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit computing it must have top fan-in at least $\exp(\Omega(\frac{d}{t}\log n))$.*

The depth reduction results combined with the lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits is indeed a remarkable collection of results. As it stood at this point, in order to separate VP from VNP, any small asymptotic improvement in the exponent on either the lower bound front or on the depth reduction front would be sufficient! In

fact for any class of circuits $\mathcal{C}$ for which we can improve the parameters of Theorem 1.8 for homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits, we would get superpolynomial lower bounds for that class using Theorem 2.1.

Unfortunately, it seems that in general, we cannot hope for a better depth reduction. In a recent work, Fournier, Limaye, Malod and Srinivasan [FLMS14] gave an example of an explicit polynomial in VP (of degree d and in $n = d^{O(1)}$ variables) such that any homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit computing it must have top fan-in at least $\exp(\Omega(\frac{d}{t}\log n))$. This immediately implies that the depth reduction parameters in the result of Tavenas [Tav15] are *tight* [3] for general circuits. This observation, along with the fact that the hard polynomial used by Kayal et al. [KSS14] has a shifted partial derivative span only a polynomial factor away from the maximum possible value suggests that the technique of improving depth reduction and then using shifted partial derivatives may not be strong enough to separate VNP from VP[4]. In a recent result, Chillara and Mukhopadhyay [CM14a] gave a clean unified way of way of lower bounding the shifted partial derivative complexities of the polynomials considered by [KSS14, FLMS14].

Arithmetic formula lower bounds

Even though improved depth reduction to homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits along with shifted partials does not seem to be powerful enough to separate VNP from VP, it is conceivable that this approach could lead to superpolynomial lower bounds for other interesting classes, for instance homogeneous arithmetic formulas, or even general arithmetic formulas. This hope was further strengthened when Kayal et al. [KSS14] used these precise ideas to prove superpolynomial lower bounds for a restricted class of formulas which they called *regular* formulas. (Regular formulas are formulas which have alternating sum and product layers. Moreover, for every fixed layer, the fan-ins of the gates in that layer are the same and the formal degree of the formula is at most

[3] Up to constants in the exponent.

[4] The reason this statement is not completely formal is that we still do not know know if the upper bounds on the shifted partial derivative measure for $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits is tight for all choices of derivatives and shifts, though the results of [FLMS14] and those in this chapter show that they are indeed tight for many of the choices.

a constant times the formal degree of the polynomial being computed.) Kayal et al. proved their result by showing that one can reduce any polynomial size regular formula to a $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit (for a carefully chosen choice of t) of size asymptotically better in the exponent than the $\exp(\frac{d}{t}\log n)$ bound (which as we just discussed is known to be tight up to constants in the exponent for circuits). This improvement in depth reduction immediately leads to superpolynomial lower bounds for regular formulas by using Theorem 2.1.

Removing the restriction on regularity and proving superpolynomial lower bounds for general formulas or even general homogeneous formulas would be a huge step forward - it would be by far the strongest and most natural class of arithmetic circuits for which we would be able to prove lower bounds, and it would represent a real breakthrough. The authors of the two papers [KSS14, FLMS14] leave as a tantalizing open question whether formulas (or even homogeneous formulas) can have better depth reduction than circuits (such as is true for regular formulas). If true, this would imply superpolynomial lower bounds for (homogeneous) formulas. Perhaps it could also be true that every formula could be reduced to a regular formula with only a polynomial blow up in size. If so, the improved depth reduction for formulas (and hence the lower bounds) would follow from the improved depth reduction of regular formulas.

Thus to summarize, the main challenge that remained was to understand the limits of the techniques of depth reduction to homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits and shifted partial derivatives. In particular, are there any other interesting classes of circuits for which improved depth reduction is possible? Is improved depth reduction possible for arithmetic formulas?

2.2 Results

In this chapter, we study the power and limitations of depth reduction to homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits for arithmetic formulas. We do this via studying depth reduction for depth-4 arithmetic circuits[5]. Let homogeneous $\Sigma\Pi\Sigma\Pi(r)$ circuits be the class of

[5]Since depth-4 arithmetic circuits are also equivalent to depth-4 arithmetic formulas up to a polynomial blow up in size, we will use the term circuits and formulas interchangeably when referring to

homogeneous depth-4 circuits with *top fan-in* bounded by r, and with *no restriction on the bottom fan-in.* This is a very natural class of circuits and is quite different in nature from $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits.

Our results are divided into two parts. In the first part we show the first superpolynomial lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi(r)$ circuits when $r = o(\log d)$. The core of our result is an improved depth reduction result for these circuits. (As we pointed out, when we refer to 'depth reduction', we really mean a reduction to homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits. Thus it makes sense to talk about a depth reduction for $\Sigma\Pi\Sigma\Pi$ circuits as well.) $\Sigma\Pi\Sigma\Pi(r)$ circuits have received significant attention for the problems of polynomial identity testing and polynomial reconstruction [KMSV10, SV11, GKL12], however prior to this work there were no nontrivial lower bounds for this class of circuits for any value of $r \geq 2$.

In the second part we show that more efficient depth reduction to homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits *is not possible* for homogeneous arithmetic formulas. We show this result by studying the very simple class of formulas given by homogeneous $\Sigma\Pi\Sigma\Pi(\log d)$ circuits. We show that for this class of circuits, improved depth reduction is not possible. This shows that improved depth reduction is unfortunately not powerful enough to prove lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi(\log d)$ circuits, and in particular not strong enough to prove lower bounds for homogeneous arithmetic formulas, answering the main open questions of [KSS14, FLMS14].

Informally, our main results are the following:

Main Theorem 1 (Informal): *There is an explicit family of polynomials in* VNP *of degree d in $n = d^{O(1)}$ variables such that for $r = o(\log d)$, any polynomial size homogeneous $\Sigma\Pi\Sigma\Pi(r)$ circuit computing it must have superpolynomial size.*

At the core of the result is the following "depth reduction" result:

---

depth-4 circuits.

Improved Depth Reduction (Slightly wishful):[6] *For $r = o(\log d)$ and a small enough constant ε, any polynomial size homogeneous $\Sigma\Pi\Sigma\Pi(r)$ circuit computing a polynomial of degree d and in n variables is equivalent to a homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit of size $\exp\left(o(\frac{d}{t}\log n)\right)$ for some choice of t such that $\log^2 d \leq t \leq \varepsilon d$.*

Observe that the parameters of the depth reduction we hope to obtain above improve upon the parameters of depth reduction given by [Koi12, Tav15].

We also show that when $r = \Omega(\log(d))$, depth reduction as above is no longer true.

Main Theorem 2 (Informal) *For $r = \Omega(\log d)$, there exists an explicit family of polynomials $\{\mathcal{Q}_n\}$, where $\mathcal{Q}_n$ is a homogeneous polynomial of degree d in $n = d^{O(1)}$ variables which can be computed by a $\mathrm{poly}(n)$ size homogeneous $\Sigma\Pi\Sigma\Pi(r)$ circuit (and hence homogeneous formula), such that for every t satisfying $t_0 \leq t \leq \varepsilon d$,[7], any homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit computing $\mathcal{Q}_n$ must have top fan-in at least $\exp\left(\Omega(\frac{d}{t}\log n)\right)$.*

An immediate consequence of this result is that the depth reduction procedure of Tavenas [Tav15] gives optimal parameters up to constants, even for homogeneous arithmetic formulas (strengthening the results of [FLMS14]).

At the core of our result is a *hierarchy* theorem for homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits which shows that homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits are a much richer class than homogeneous $\Sigma\Pi\Sigma\Pi^{[t/20]}$ circuits. We state this result more formally in Theorem 2.5.

It was shown in [KSS14] that any ABP (even non homogeneous) can be converted to a regular formula with a quasipolynomial blow up in size. If one could improve this transformation even slightly for formulas or even for homogeneous formulas, this

[6] Indeed the above statement is not quite true, and our reduction turns out to be much more subtle. We do not depth reduce to a $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit, but to one in which the sum of degrees of any $\varepsilon d/t$ product gates at the bottom is at most εd. This is a more refined notion and a slightly more general class of circuits than $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits. We observe that the shifted partial derivative technique does not distinguish between these two kinds of circuits, and thus we are still able to obtain our lower bounds. Thus in spirit we still get depth reduction. In fact everywhere in this chapter we could replace $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits with this slightly more general class of circuits, and none of the results would be affected.

[7] *here t_0 and ε are constants*

would imply superpolynomial lower bounds for formulas/homogeneous formulas. Another consequence of our results is that such an improvement is not possible. We build upon the results of [KSS14] and show that the conversion of general formulas to regular formulas must incur a quasipolynomial blow up in size.

Theorem (Conversion to Regular Formulas is Tight) *For $r = \Omega(\log d)$, there exists an explicit family of polynomials $\{\mathcal{Q}_n\}$, where $\mathcal{Q}_n$ is a homogeneous polynomial of degree d in $n = d^{O(1)}$ variables which can be computed by a* $\mathrm{poly}(n)$ *size homogeneous $\Sigma\Pi\Sigma\Pi(r)$ circuit (and hence homogeneous formula), such that any regular formula computing $\mathcal{Q}_n$ must have size $n^{\Omega(\log d)}$.*

In the sections below we formally state our results and elaborate on them in greater detail, as well as highlight some of the interesting corollaries of our proof techniques.

2.2.1 Lower bounds for $\Sigma\Pi\Sigma\Pi(r)$ circuits, $r = o(\log d)$

In the first part of the chapter, we explore the limits of computation of depth-4 homogeneous circuits when the restriction for the bottom fan-in is removed. For the general model of (even homogeneous) $\Sigma\Pi\Sigma\Pi$ circuits, only extremely weak lower bounds seem to be known. Even PIT for $\Sigma\Pi\Sigma\Pi$ circuits is known only when the top fan-in is constant and the circuit is multilinear[8]. For (non-multilinear) depth-4 circuits of low formal degree, even when the top fan-in is 2, prior to this work there were no lower bounds known in general. Unlike the class of depth-3 circuits with bounded top fan-in which cannot even compute all polynomials irrespective of the size of the circuit, the class of $\Sigma\Pi\Sigma\Pi(r)$ circuits is complete (even for $r = 1$).

We consider homogeneous $\Sigma\Pi\Sigma\Pi(r)$ circuits, which are depth-4 homogeneous circuits whose *top fan-in* is bounded by r. When r is a constant we prove exponential lower bounds[9] for the class of $\Sigma\Pi\Sigma\Pi(r)$ circuits, and for any $r = o(\log d)$ we show

[8] In all the results of this chapter, the restriction of homogeneity can be replaced by the restriction that all gates in the circuit compute polynomials of degree at most d.

[9] In the rest of the chapter, by exponential lower bound we will mean a lower bound of the form $2^{n^{\varepsilon}}$

superpolynomial lower bounds for $\Sigma\Pi\Sigma\Pi(r)$ circuits[10]. In particular, we prove the following theorem:

Theorem 2.2. *There exists an explicit family of polynomials in* VNP*,* $\{NW_n\}$*, such that for each* n*,* NW_n *has degree* $d = \Theta(\sqrt{n})$*, and number of variables* $\Theta(n)$ *and such that the following holds: Let* C *be a homogeneous* $\Sigma\Pi\Sigma\Pi(r)$ *circuit that computes* NW_n*. Let* s *be the size of* C*. Then*

$$s \geq \exp\left(d^{\Omega(1/r)} \log n\right).$$

Lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi^*$ circuits:

Another class of circuits we are able to prove a lower bound for is the class of depth-4 circuits where each product at the second layer (from the top) has the same degree sequence of incoming polynomials, and there is no restriction on the top fan-in.

For any degree sequence $\mathcal{D} = D_1, D_2, \ldots, D_k$ of non-negative integers such that $\sum D_i = d$, we study the class of homogeneous $\Sigma\Pi^{\mathcal{D}}\Sigma\Pi$ circuits, which are homogeneous circuits where each Π gate at the second layer is restricted to having its inputs be polynomials whose sequence of degrees is precisely $\mathcal{D}$.

We show that for *every* degree sequence $\mathcal{D}$, any $\Sigma\Pi^{\mathcal{D}}\Sigma\Pi$ circuit computing NW_n (an explicit family of polynomials in VNP) must have size at least $\exp(d^{\varepsilon})$, for some fixed absolute constant ε independent of $\mathcal{D}$. In particular, let the class of $\Sigma\Pi\Sigma\Pi^*$ circuits be the union of the classes of $\Sigma\Pi^{\mathcal{D}}\Sigma\Pi$ for all $\mathcal{D}$. Then our lower bounds hold for homogeneous $\Sigma\Pi\Sigma\Pi^*$ circuits as well.

Theorem 2.3. *There exists an explicit family of polynomials in* VNP*,* $\{NW_n\}$*, such that for each* n*,* NW_n *has degree* $d = \Theta(\sqrt{n})$*, and number of variables* $\Theta(n)$ *and such that the following holds: Let* C *be a homogeneous* $\Sigma\Pi\Sigma\Pi^*$ *circuit that computes* NW_n*.*

---

for some constant ε.

[10]It is important to observe that the reduction of a polynomial sized homogeneous $\Sigma\Pi\Sigma\Pi$ circuit with arbitrary bottom fan-in to a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit with bounded bottom fan-in as given by the results of [AV08, Koi12] can lead to circuits of size $\exp(\Omega(d/t)\log n)$ and so Theorem 2.1 does not imply any nontrivial lower bounds for it.

Let s be the size of C. Then

$$s \geq \exp\left(d^{\varepsilon}\right),$$

for some fixed absolute constant $\varepsilon > 0$.

2.2.2 Depth reduction is tight for $\Sigma\Pi\Sigma\Pi(r)$ circuits, $r = \Omega(\log d)$

The main question that was left open by both the works of [KSS14] and [FLMS14] was to understand whether an improved depth reduction was possible for general (homogeneous) arithmetic formulas.

In particular, the following tantalizing questions naturally emerge and were left as open questions by the works of [KSS14] and [FLMS14].

- Can the depth reduction by Koiran and Tavenas [Koi12, Tav15] be improved for formulas: In other words, can one show that for every polynomial of degree d and in n variables which has a polynomial sized (homogeneous) formula, it can be reduced to a $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuit of size $n^{o(\sqrt{d})}$?
- Can every homogeneous arithmetic formula be converted to a regular formula with only a polynomial blow up in its size?

A positive answer to any of the above questions would suffice in proving superpolynomial lower bounds for general homogeneous arithmetic formulas. We settle both the questions and show that unfortunately neither is true.

We settle these questions by constructing an explicit family of polynomials $\{\mathcal{Q}_n\}$, where $\mathcal{Q}_n$ is a polynomial in $\Theta(n)$ variables and is of degree $d = \Theta(\sqrt{n})$, such that for each n, $\mathcal{Q}_n$ can be computed by a *polynomial sized homogeneous formula*, but any $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuit computing $\mathcal{Q}_n$ must have top fan-in at least $2^{\Omega(\sqrt{d}\log n)}$. Moreover $\mathcal{Q}_n$ is computed by a polynomial size homogeneous $\Sigma\Pi\Sigma\Pi(r)$ formula for $r = \Theta(\log d)$. More formally, we prove the following theorem.

Theorem 2.4 (Depth reduction is tight for formulas)**.** *There exists an explicit family of polynomials $\{\mathcal{Q}_n\}$ and an absolute constant $\varepsilon > 0$ such that $\mathcal{Q}_n$ is of degree $d = \Theta(\sqrt{n})$, in $\Theta(n)$ variables, and computed by a* $\mathrm{poly}(n)$ *size homogeneous $\Sigma\Pi\Sigma\Pi(\log d)$ circuit*

(in particular a homogeneous arithmetic formula); and for every t such that $t_0 \leq t \leq \varepsilon d$ for constants ε and t_0, any homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit computing $\mathcal{Q}_n$ must have top fan-in at least $\exp\left(\Omega(\frac{d}{t}\log n)\right)$.

The above theorem follows by an interpolation argument applied to a *hierarchy* theorem for $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits, which is the heart of our argument. The hierarchy theorem shows that by increasing the bound on the bottom fan-in of $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits even slightly, we get a much richer class of arithmetic circuits. We believe this is an interesting result in its own right.

Theorem 2.5 (Hierarchy theorem for $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits)**.** *There exists a constant $\varepsilon \in (0,1)$ such that for every t with $t_0 \leq t \leq \varepsilon d$ for a constant t_0, there exists an explicit family of polynomials $\{\mathcal{P}_{t,n}\}$, such that $\mathcal{P}_{t,n}$ is an n-variate polynomial of degree $d = \sqrt{n}$, and is computed by a* $\mathrm{poly}(n)$ *size homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit, and for every t' s.t. $t' < \delta t$ for any constant $\delta < 1$, any homogeneous $\Sigma\Pi\Sigma\Pi^{[t']}$ circuit computing $\mathcal{P}_{t,n}$ must have top fan-in at least* $\exp\left(\Omega(\frac{d}{t}\log n)\right)$.

These results immediately imply that Koiran's and Tavenas' depth reduction [Koi12, Tav15] is tight for formulas, for all but a small number of choices of the bottom fan-in. In particular, it is tight for the case where the bottom fan-in is bounded by $\sqrt{d}$. Interestingly enough, the polynomial size formulas computing $\mathcal{Q}_n$ are of depth-4. In fact, they are a sum of $O(\log d)$ *regular* homogeneous formulas of depth-4.

A corollary of our results is that any conversion of a general (homogeneous) formula to a regular formula must incur a quasipolynomial blow up in size. It was shown in [KSS14] that any algebraic branching program can be converted to a regular formula with a quasipolynomial blow up in size. Since it is widely believed that formulas are much weaker than ABPs, it was conjectured that formulas, or homogeneous formulas might have a more efficient conversion (which would suffice in proving superpolynomial lower bounds for homogeneous formulas!). We show however that this is not true. Combining our results with the result of [KSS14], we obtain the following (tight) lower bound for converting homogeneous formulas to regular formulas.

Theorem 2.6 (Lower bounds for reduction to Regular Formulas)**.** *There exists an explicit family of polynomials* $\{\mathcal{Q}_n\}$ *and an absolute constant* $\varepsilon > 0$ *such that* $\mathcal{Q}_n$ *is of degree* $d = \Theta(\sqrt{n})$*, in* $\Theta(n)$ *variables, and computed by a* $\mathrm{poly}(n)$ *size homogeneous* $\Sigma\Pi\Sigma\Pi(\log d)$ *circuit (in particular a homogeneous arithmetic formula); and any regular formula computing* $\mathcal{Q}_n$ *must have size at least* $n^{\Omega(\log d)}$*.*

Organization of the chapter

The rest of the chapter is organized as follows. In Section 2.3, we introduce some preliminary notions and notations. In Section 2.4 we prove our lower bound for homogeneous $\Sigma\Pi\Sigma\Pi(r)$ circuits when $r = o(\log d)$. In Section 2.5, we show that depth reduction is tight for homogeneous arithmetic formulas by showing it is tight for homogeneous $\Sigma\Pi\Sigma\Pi(\Omega(\log d))$ circuits.

2.3 Preliminaries

We start with some notations.

- Unless otherwise stated, we shall use bold-face letters such as $\mathbf{x}$ to denote a set of variables $\{x_1, \ldots, x_n\}$. Most of the time, the size of this set would be clear from context. We shall also abuse this notation to use $\mathbf{x}^{\mathbf{e}}$ to refer to the monomial $x_1^{e_1} \cdots x_n^{e_n}$.

- For an integer $m > 0$, we shall use $[m]$ to denote the set $\{1, \ldots, m\}$.

- We shall use the short-hand $\partial_{\mathbf{x}^{\mathbf{e}}}(P)$ to denote

$$\frac{\partial^{e_1}}{\partial x_1^{e_1}} \left(\frac{\partial^{e_2}}{\partial x_2^{e_2}} \left(\cdots (P) \cdots \right) \right).$$

- For a set of polynomials $\mathcal{P}$ shall use $\partial^{=k}\mathcal{P}$ to denote the set of all k-th order partial derivatives of polynomials in $\mathcal{P}$, and $\partial^{\leq k}\mathcal{P}$ similarly.

  Also, $\mathbf{x}^{=\ell}\mathcal{P}$ shall refer to the set of polynomials of the form $\mathbf{x}^{\mathbf{e}} \cdot P$ where $\mathsf{Deg}(\mathbf{x}^{\mathbf{e}}) = \ell$ and $P \in \mathcal{P}$. Similarly $\mathbf{x}^{\leq \ell}\mathcal{P}$.

- For a polynomial $P \in \mathbb{F}_q[\mathbf{x}]$ and for a set $S \subseteq \mathbb{F}_q^n$, we shall denote by $\mathsf{Eval}_S(P)$ the vector of the evaluation of P on points in S (in some natural predefined order like say the lexicographic order).
- For a set of vectors V, their span over the underlying field $\mathbb{F}$ will be denoted by $\mathsf{Span}(V)$ and their dimension by $\mathsf{Dim}(V)$.

Recall that a polynomial $P(\mathbf{x})$ computed by a depth-4 circuit can be expressed as

$$P(\mathbf{x}) = \sum_{i=1}^{r} \prod_{j=1}^{d_i} Q_{ij}(\mathbf{x}) \tag{2.7}$$

Based upon this definition, we will now define the specific restrictions of depth-4 circuits that we study in this chapter.

Homogeneous $\Sigma\Pi\Sigma\Pi^{[a]}$ circuits and homogeneous $\Sigma\Pi^{[b]}\Sigma\Pi^{[a]}$ circuits

The depth-4 $\Sigma\Pi\Sigma\Pi$ circuit in Equation 2.7 is said to be a $\Sigma\Pi^{[b]}\Sigma\Pi^{[a]}$ circuit if each $Q_{ij}(\mathbf{x})$ is a polynomial of degree at most a and each d_i is at most b. The depth-4 $\Sigma\Pi\Sigma\Pi$ circuit in Equation 2.7 is said to be a $\Sigma\Pi\Sigma\Pi^{[a]}$ circuit if each $Q_{ij}(\mathbf{x})$ is a polynomial of degree at most a. In this case we say that the *botton fan-in is bounded by a*. If the circuit is homogeneous, then we can assume without loss of generality that for each i, $\prod_{j=1}^{d_i} Q_{ij}(\mathbf{x})$ is a polynomial of degree exactly d (the degree of P).

Observe that, for each i, by grouping together and multiplying out some of the Q_{ij}, we can transform a homogeneous $\Sigma\Pi\Sigma\Pi^{[a]}$ circuit into a homogeneous $\Sigma\Pi^{[b]}\Sigma\Pi^{[a]}$ circuit, where $b = O(\frac{d}{a})$. This operation of grouping together and multiplying would increase the size of the resulting circuit, but notice that it does not affect the top fan-in of the circuit. Thus lower bounds on the top fan-in of homogeneous $\Sigma\Pi^{[O(b)]}\Sigma\Pi^{[a]}$ circuits imply the same lower bounds on the top fan-in of homogeneous $\Sigma\Pi\Sigma\Pi^{[a]}$ circuits.

Homogeneous $\Sigma\Pi\Sigma\Pi(r)$ Circuits

The depth-4 $\Sigma\Pi\Sigma\Pi$ circuit in Equation 2.7 is said to be a $\Sigma\Pi\Sigma\Pi(r)$ circuit if the fan-in of the summation(top fan-in) is bounded by r. Observe that there is no restriction on the bottom fan-in except that implied by the restriction of homogeneity.

For each $i \in [r]$, the product $P_i = \prod_{j=1}^{d_i} Q_{ij}$ is said to be computed by the product gate i. Therefore, $P = \sum_{i=1}^{r} P_i$. Here for every i and j, Q_{ij} is an n variate homogeneous polynomial being computed by a $\Sigma\Pi$ circuit. The homogeneity restriction on C implies that for every product gate i,

$$\deg(P) = d = \sum_{j=1}^{d_i} \deg(Q_{ij}) \tag{2.8}$$

With every product gate $i \in [r]$, we can associate a multiset (D_i, m_i), where

$$D_i = \{\deg(Q_{ij}) : j \in [d_i]\} \tag{2.9}$$

and m_i is a map from D_i to $\mathbb{N}$, which assigns to every element l in D_i, the number of $j \in [d_i]$ such that Q_{ij} has degree equal to l. For a homogeneous depth-4 circuit, computing a degree d polynomial, Equation 2.8 can be rewritten as

$$\deg(P) = d = \sum_{l \in D_i} l \times m_i(l) \tag{2.10}$$

for each i in $[r]$. $\Sigma\Pi\Sigma\Pi(r)$ circuits for which the multiset (D_i, m_i) is the same for every product gate $i \in [r]$, are said to be $\Sigma\Pi\Sigma\Pi^*$ circuits.

Regular Formula

The notion of regular formulas was introduced in [KSS14], where superpolynomial lower bounds for this model were proved.

Definition 2.11. *A formula computing a degree d polynomial in n variables is said to be regular, if it satisfies the following conditions:*

1. *It has alternating layers of sum and product gates.*
2. *All gates in a single layer have the same fan-in.*
3. *The formal degree of the formula is at most some constant multiple of the degree of the polynomial being computed.*

$\Diamond$

Shifted Partial Derivatives

A very useful notion related to polynomials is the notion of *shifted partial derivatives* of a polynomial. The precise notion was defined by Kayal [Kay12], and has been at the heart of the rapid progress on the question of bounded depth arithmetic circuit lower bounds in the last few years. In particular, for our results in this thesis, we will study the complexity of many polynomial families using their shifted partial derivatives themselves, and using many variants of this notion. Formally the dimension of the space of shifted partial derivatives of P is defined as follows.

Definition 2.12 (Shifted partial derivatives [GKKS14])**.** *Let k, ℓ be some parameters. For any polynomial P, define $\Gamma_{k,\ell}(P)$ as*

$$\Gamma_{k,\ell}(P) \quad := \quad \mathsf{Dim}\left\{\mathsf{Span}\left(\mathbf{x}^{=\ell}\partial^{=k}(P)\right)\right\}. \qquad \Diamond$$

Nisan-Wigderson Polynomials

We now define the specific variant of the Nisan-Wigderson design polynomials used in this chapter. For a prime power d, let $\mathbb{F}_d$ be a field of size d. For the set of $n = d^2$ variables $\{x_{i,j} : i, j \in [d]\}$ and $t \in [d]$, we define the n variate degree d homogeneous polynomial $NW_{t,n}$ as

$$NW_{t,n} = \sum_{\substack{f(z)\in\mathbb{F}_d[z] \\ deg(f)<\lfloor\frac{d}{2t}\rfloor}} \prod_{i\in[d]} x_{i,f(i)}$$

Clearly, for every prime power d and every t such that $\frac{d}{2t}$ is an integer, $NW_{t,n}$ is in VNP. The Nisan-Wigderson polynomial family $\{NW_n\}$ is a family of polynomials in VNP such that NW_n is a polynomial of degree $d+1$ in $n = d^2 + d$ variables $\{x_{i,j} : i, j \in [d]\} \cup \{y_i : i \in [d]\}$ defined as follows

$$NW_n = \sum_{i=1}^{d} y_i \cdot NW_{i,n}$$

Approximations

We use the following lemma to approximate expressions during our calculations.

Lemma 2.13 ([GKKS14]). *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,*

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

We use the symbol $\approx$ to indicate equality up to constant factors. For most of the applications of this lemma in this thesis, $\frac{(f+g)^2}{a} = O(1)$.

2.4 Lower bounds for $\Sigma\Pi\Sigma\Pi(r)$ circuits, $r = o(\log d)$

In earlier works by Gupta et al. [GKKS14] and Kayal et al. [KSS14], exponential lower bounds were shown for the class of homogeneous $\Sigma\Pi\Sigma\Pi$ circuits with bounded bottom fan-in. Without the restriction on the bottom fan-in, basically no lower bounds for $\Sigma\Pi\Sigma\Pi$ circuits are known. In this section we prove the first super-polynomial lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits with bounded top fan-in. The main technical core of our result is a *depth reduction* result, very similar in spirit to those by Koiran [Koi12] and Tavenas [Tav15]. By exploiting the structure of these circuits, we show how to get *improved depth reduction* for them. The proof of our depth reduction is quite different from that of [Koi12, Tav15] and is somewhat subtle. We don't reduce to a $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit, but to a related and slightly more general class of circuits, where instead of requiring an absolute bound on the bottom fan-in, we just require that, in some sort of average sense, the bottom fan-in is small. In particular we reduce to a $\Sigma\Pi\Sigma\Pi$ circuit in which the sum of degrees of any $\varepsilon d/t$ product gates at the bottom is at most εd. This is a more refined notion and a slightly more general class of circuits than $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits. We observe that the shifted partial derivative technique does not distinguish between these two kinds of circuits, and thus we are still able to use a variant of Theorem 2.1 to obtain our lower bounds. Thus in spirit we still get depth reduction. In fact everywhere in this chapter we could replace $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits with this slightly more general class of circuits, and none of the results would be affected.

There seem to be two main obstacles in extending the lower bounds of [GKKS14, KSS14] to lower bounds for general depth-4 homogeneous $\Sigma\Pi\Sigma\Pi$ circuits. The lower bounds in [GKKS14, KSS14] work only when the degrees of *all* polynomials feeding

into the product gate at the second layer are small (in other words, the bottom fan-in is small), say $\leq \sqrt{d}$. If the degrees of all polynomials feeding into the product gate at the second layer is large (i.e. the bottom fan-in of all the gates is large), say $\geq \sqrt{d}$, then for sparsity reasons and simple monomial counting, it is easy to obtain exponential lower bounds. The first obstacle is to handle the case when the degrees of some of the polynomials is small and for some of them it is large. For instance fix any arbitrary sequence $\mathcal{D}$ of degrees summing to d, and assume that the polynomials feeding into each product gate at the second from top layer have their degrees coming from this sequence. Is it still possible to obtain exponential lower bounds? The second obstacle to extending the results from [GKKS14] is to find a way to combine the lower bounds for all these various cases into a common lower bound for the case when the circuit is composed of product gates of different kinds. For instance we know lower bounds when all product gates at the second layer have small incoming degrees and when all product gates have large incoming degrees. However we do not know how to combine these lower bounds into a single lower bound when the circuit is the sum of two circuits, one of the low degree kind, and one of the high degree kind. In this chapter we show how to resolve the first obstacle. Moreover when the top fan-in is $o(\log d)$, the second obstacle turns out to not be a problem either.

Proof Overview

Most lower bounds for arithmetic circuits proceed by identifying some kind of "progress measure", and show that for any given circuit in a circuit class, the measure is small if the size of the circuit is small, whereas for the polynomial one is trying to compute (for instance the permanent), the measure is large. In the results by Gupta et al. [GKKS14] and Kayal et al. [KSS14], the progress measure used is the dimension of the ℓ shifted k^{th} order partial derivative $\Gamma_{k,\ell}(P)$, for a suitable choice of ℓ and k. It is shown that every small depth-4 circuit with bounded bottom fan-in has small $\Gamma_{k,\ell}(P)$ compared to that of an explicit polynomial in VNP, the NW_n polynomial. Thus if a depth-4 circuit with bounded bottom fan-in must compute NW_n, then it must be large. More precisely, it is shown that every product gate $Q_i = \prod_{j=1}^{d} Q_{ij}$ has $\Gamma_{k,\ell}(P)$ much smaller than that

of the NW_n polynomial, provided the degrees of the Q_{ij} are small. This is the core of the argument. Combined with the sub-additivity of $\Gamma_{k,\ell}(P)$, the result easily follows.

Our proof builds upon the results of [GKKS14] and [KSS14], and combines the use of the progress measure $\Gamma_{k,\ell}(P)$ with the notion of "sparsity" to prove our improved depth reduction and the lower bounds for the polynomial family $\{NW_n\}$. Suppose $C = \sum_{i=1}^{r} \prod_{j=1}^{d_i} Q_{ij}$ is a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing NW_n. If all the Q_{ij} had low degree, then the results of [GKKS14] and [KSS14] give exponential lower bounds for the top fan-in of C. Also in the extreme case where all the Q_{ij} have high degree, then since C is homogeneous, the number of Q_{ij} per product gate $Q_i = \prod_{j=1}^{d} Q_{ij}$ must be small, and hence their product cannot have too many monomials[11]. If the number of monomials is too few, we would not even be able to get all the monomials in NW_n. In general, of course there might be some high degree and some low degree polynomials, and we attempt to interpolate between the two settings to obtain our results.

For each product gate $Q_i = \prod_{j=1}^{d_i} Q_{ij}$, recall that each Q_{ij} is a homogeneous polynomial of degree d_{ij} (say), and $\sum_{j=1}^{d_i} d_{ij} = d$. If the size of the circuit is at most s, then each Q_{ij} has at most s monomials. We decompose each product gate into its inputs Q_{ij} of *high degree* (those of degree $\geq t$) and its inputs Q_{ij} of *low degree* (those of degree <t). Observe that there cannot be too many (greater than d/t) high degree polynomials Q_{ij} as otherwise their product would have degree exceeding d. Thus the product of all the high degree Q_{ij} cannot have more than $s^{d/t}$ monomials. Let H be the product of the the high degree Q_{ij}, and L be the product of the low degree Q_{ij}. Then, by writing out H as a sum of monomials ($H = \sum_k h_k$) and multiplying each monomial h_k with L, we can expand out Q_i as $\sum_k h_k \cdot L$. Note that L is a product of low degree polynomials. Also, each h_k is a monomial and hence a product of degree 1 polynomials. Thus we have expressed Q_i as a $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit, where now all the product gates multiply polynomials of degree at most t.

The hope at this point would be to apply this transformation to all the product gates and then possibly apply the result in [KSS14] to obtain a lower bound. The

[11]The number of monomials in each Q_{ij} is a most the size of the circuit.

trouble with this argument is that under the transformation described, the top fan-in of the original circuit might blow up by a factor equaling the number of monomials in H, which could be nearly as large as $s^{d/t}$. With this loss in parameters, the bound given by the [KSS14] result gives nothing nontrivial. Thus in general one cannot choose an absolute threshold t and for all product gates choose degrees greater than t to be the high degree polynomials and the ones below t to be the low degree polynomials.

What we show is that by examining the degrees of the polynomials feeding into the product gates, one can carefully choose a threshold t that works for each product gate individually, though it might not be the same threshold for all gates. It turns out that this threshold that we find is purely a function of the degree sequence $\mathcal{D}$ of the product gate. Thus if all product gates have the *same* degree sequence, i.e. we have a $\Sigma\Pi^{\mathcal{D}}\Sigma\Pi$ circuit, then we obtain exponential lower bounds. However, for general $\Sigma\Pi\Sigma\Pi$ circuits it can be a problem, since if the threshold is different for different gates, we do not have any one single progress measure that works for all gates and thus for the entire circuit. However we are still able to show that for each gate, only very few thresholds are "bad", and when the top fan-in is $o(\log d)$, then we show there is a single threshold that will work for all gates to give superpolynomial lower bounds.

2.4.1 Proof of Theorem 2.2

In this subsection, we will present the proof of Theorem 2.2. Let us consider a homogeneous $\Sigma\Pi\Sigma\Pi(r)$ circuit C of size s computing NW_n. From Equation 2.7, this implies that

$$NW_n = \sum_{i=1}^{r} \prod_{j=1}^{d_i} Q_{ij} \tag{2.14}$$

where for every value of i and j, Q_{ij} is a homogeneous polynomial being computed by a subcircuit of depth 2 of C. Observe that Q_{ij} is being computed by a $\Sigma\Pi$ circuit, hence the number of monomials with nonzero coefficients in a sum of products expansion of Q_{ij} will be at most the size of C. In other words, Q_{ij} is s sparse for each $i \in [r]$ and $j \in [d_i]$. Without loss of generality, we will assume that for every $i \in [r]$, $d_i = d$, since if $d_i < d$ for any i, we can always make it equal to d adding dummy polynomials that

are the constant 1.

Let us now consider the polynomial computed at a product gate near the top of C. It is of the form $Q = \prod_{i\in[d]} Q_i$. Let us also assume without loss of generality that the Q_i are arranged in non-increasing order of their degrees. The idea of the proof, as described in the overview, would be to decompose the Q_i into *high degree* and *low degree* parts and then multiply out all the *high degree* parts and count on their sparsity to show that the product does not blow up the dimension of the space of shifted partial derivatives by too much. We will then use the following lemma implicit in the work of [GKKS14], to obtain our bounds.

Lemma 2.15 (Implicit in [GKKS14])**.** *Let $P = \prod_{i=1}^{d} \tilde{P}_i$ be a polynomial in n variables such that the sum of the degrees of any k of these d polynomials $\tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_d$ is at most D. Then, for every integer $\ell \geq 0$,*

$$\Gamma_{k,\ell}(f) \leq \binom{d+k-1}{k}\binom{n+D-k+\ell}{n}.$$

Proof. The proof of the lemma is exactly the same calculation as in [GKKS14]. We replace their bound of tk (which for them was the sum of degrees of k polynomials of degree at most t), by our bound of D. □

The following lemma is the core of our argument.

Lemma 2.16. *Let $Q = \prod_{j\in[d]} Q_j$ be a depth 3 $\Pi\Sigma\Pi$ homogeneous circuit of degree d in n variables, where each Q_i has at most s monomials. Let $0 < \varepsilon < 1$ be any small constant and let m be $o(\log n)$. Consider $k = d^{i/m}$, for $1 \leq i \leq m$ and any integer $\ell \geq 0$. Then for all but $1/\varepsilon$ choices of i ,*

$$\Gamma_{k,\ell}(Q) \leq s^{k\cdot d^{-1/m}} \cdot \binom{k/\varepsilon+k-1}{k}\binom{n+4\varepsilon d-k+\ell}{n}.$$

Proof. Since the Q_j's are arranged in order of decreasing degree, Q_1 has highest degree and Q_d has the smallest degree.

For $1 \leq i \leq m$, let $S_i = \{Q_j | j \leq d^{i/m}\}$ be the set of the first $d^{i/m}$ of the Q_j's. For each i, we will sum the degrees of the Q_j's in $S_i \setminus S_{i-1}$. Let

$$D_i = \sum_{j \text{ s.t. } Q_j \in S_i \setminus S_{i-1}} \mathsf{Deg}(Q_j).$$

Then $\sum_{i=1}^{m} D_i = d$. Thus there are at most $1/\varepsilon$ choices of i for which $D_i \geq \varepsilon d$. We will show that for all other choices of i, for $k = d^{i/m}$ and any integer $\ell \geq 0$, $\Gamma_{k,\ell}(Q) \leq s^{k \cdot d^{-1/m}} \cdot \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n}$.

Let us fix i such that $D_i \leq \varepsilon d$. We will split up the various Q_j's into those that are in S_{i-1} and those that are not. For those Q_j in S_{i-1}, we will exploit the fact that there aren't too many of them and they each have at most s monomials, to show that they do not affect the dimension of shifted partial derivatives by too much. For the rest of the Q_j we will take advantage of the fact that their degrees are not too large, and hence the sum of degrees of any k of them is small, and thus we will be able to bound the span of shifted partial derivatives of their product using the argument presented in [GKKS14].

Let $H = \prod_{Q_j \in S_{i-1}} Q_j$, and let $Q_{\bar{H}} = Q/H$. Since each Q_j has at most s monomials, thus H has at most $s^{d^{(i-1)/m}}$ monomials. Hence we can express the polynomial Q as the sum of at most $s^{d^{(i-1)/m}}$ polynomials $P_1, P_2, \ldots P_u$, where each of the polynomials is the product of some monomial (from H), and the product of all the Q_j that are not in S_{i-1} (i.e. those in $Q_{\bar{H}}$) .

We will show that for each P_j, $1 \leq j \leq u$, for $k = d^{i/m}$,

$$\Gamma_{k,\ell}(P_j) \leq \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n}.$$

Since u is at most the number of monomials in H, thus $u \leq s^{d^{(i-1)/m}} = s^{k \cdot d^{-1/m}}$. Since $Q = \sum_{j \in [u]} P_j$, the sub-additivity of $\Gamma_{k,\ell}(\cdot)$ will imply that

$$\Gamma_{k,\ell}(Q) \leq s^{k \cdot d^{-1/m}} \cdot \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n}.$$

Let us focus our attention on any one of these polynomials P_j, and call it P.

Then $P = h \cdot Q_{\bar{H}} = h \cdot \prod_{j \geq d^{(i-1)/m}+1} Q_j$, where h is a monomial of H and can be thus written as a product of degree one homogeneous polynomials. Let us rename the degree 1 polynomials in h and the different Q_j dividing $Q_{\bar{H}}$, so that $P = \hat{P}_1 \hat{P}_2 \cdots \hat{P}_\ell$.

Consider all the polynomials $\hat{P}_i$ dividing P which have degree at most $\varepsilon d/k$. Let this set be G. We can partition the set G into subsets such that the sum of the degree of polynomials in any such partition is at least $\varepsilon d/k$ and at most $2\varepsilon d/k$, by greedily adding

polynomials into the first partition G_1 as long as the sum of degree of polynomials in it is at most $\varepsilon d/k$, and so on. We take the product of the polynomials in each partition, and call them *grouped* polynomials. Call the new set of polynomials (the grouped ones and the ones that had degree at least $\varepsilon d/k$ to start out with) $\tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_\tau$. Since the sum of their degrees is at most d, thus the total number τ of these polynomials is at most k/ε.

Proposition 2.17. *The sum of the degrees of any k of these τ polynomials $\tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_\tau$ is at most $4\varepsilon d$.*

Proof. Out of the k polynomials, we see what fraction lie among the "grouped" polynomials, and what lie among the original ungrouped polynomials. Recall that by the choice of i, and setting $k = d^{\frac{i}{m}}$, the sum of degrees of any $k - kd^{\frac{-1}{m}}$ of the $\hat{P}_j$ dividing P was at most εd. Since m is $o(\log d)$, the sum of the degrees of any k of them will be at most $2\varepsilon d$. Thus, the contribution from the original ungrouped polynomials is at most $2\varepsilon d$. Also, the contribution from the grouped polynomials can be at most $2\varepsilon d$ since there are at most k of them, and each has degree at most $2\varepsilon d/k$. Thus the total sum of degrees is at most $4\varepsilon d$. □

Thus, $P = \prod_{i=1}^{\tau} \tilde{P}_i$ is a polynomial in n variables such that the sum of the degrees of any k of the τ polynomials $\tilde{P}_1, \tilde{P}_2, \ldots, \tilde{P}_\tau$ is at most $D = 4\varepsilon d$. Recall also that $\tau \leq k/\varepsilon$. Hence, by Lemma 2.15, for any integer $\ell \geq 0$,

$$\Gamma_{k,\ell}(P) \leq \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n}.$$

□

Theorem 2.18. *Let C be a homogeneous $\Sigma\Pi\Sigma\Pi(r)$ circuit in n variables, of size s and of degree at most d. Then for all constants ε, with $0 < \varepsilon < 1$, there exists a choice of i, with $1 \leq i \leq 2r/\varepsilon$, such that for $k = d^{\varepsilon i/2r}$, and for all integers $\ell \geq 0$,*

$$\Gamma_{k,\ell}(C) \leq r \cdot s^{k \cdot d^{-\varepsilon/2r}} \cdot \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n}.$$

Proof. Let $m = 2r/\varepsilon$. Let $C = \sum_{j=1}^{r} Q_j$. Let $i \in [m]$.

Then for each Q_j, by Lemma 2.16, for all but $1/\varepsilon$ choices of i, for $k = d^{i/m}$,

$$\Gamma_{k,\ell}(Q_j) \leq s^{k \cdot d^{-1/m}} \cdot \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n}.$$

Hence for each Q_j we get at most $1/\varepsilon$ choices of i that may not work to get the bound above, and we call those choices "bad" for Q_j. We call the rest of the choices "good" for Q_j. Thus by the union bound there are at most r/ε choices of i that are bad for some Q_j. Since $m > r/\varepsilon$, thus there is a choice of $i \in [m]$ that is good for every Q_j.

Thus for any integer $\ell \geq 0$ and $k = d^{i/m}$, for all $j \in [r]$,

$$\Gamma_{k,\ell}(Q_j) \leq s^{k \cdot d^{-1/m}} \cdot \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n}.$$

Hence

$$\Gamma_{k,\ell}(C) \leq r \cdot s^{k \cdot d^{-1/m}} \cdot \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n}.$$

□

We can observe that the choice of the threshold and k for every product gate just depends upon the multiset of the degrees associated with the input feeding into it. In particular, if we start with a $\Sigma\Pi\Sigma\Pi^*$ circuit, then the value of the threshold and k that works for one product gate also works for the circuit in general. Hence, we have the following theorem which gives us an upper bound on the dimension of the shifted partial derivative space of a $\Sigma\Pi\Sigma\Pi^*$ circuit.

Theorem 2.19. *Let C be a homogeneous $\Sigma\Pi\Sigma\Pi^*$ circuit in n variables, of size s, top fan-in r and of degree at most d. Then for all constants ε, with $0 < \varepsilon < 1$, there exists a choice of i, with $1 \leq i \leq m$, where $m = 1/\varepsilon + 1$ such that for $k = d^{i/m}$, and for all integers $\ell \geq 0$,*

$$\Gamma_{k,\ell}(C) \leq r \cdot s^{k \cdot d^{-1/m}} \cdot \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n}.$$

It is important to note the difference between the bounds in Theorem 2.18 and Theorem 2.19. In Theorem 2.19, the exponent of s is independent of the top fan-in r as m is a constant.

In order to complete the proof now, we will look at the shifted partial derivative complexity of the circuit as well as of the polynomial NW_n under restrictions where all the variables $\{y_1, y_2, \ldots, y_d\}$ are set to constants. The partial derivatives as well the final shifts are just taken with respect to monomials in the d^2 variables $\{x_{1,1}, x_{1,2}, \ldots, x_{d,d}\}$. The following theorem tells us that under some restrictions of this type, NW_n has large complexity. This happens because under the restriction where $y_t = 1$ and $y_j = 0$ for every $j \neq t$, we obtain $NW_{t,n}$ from NW_n.

Theorem 2.20 ([KSS14]). *For any integers t, k, ℓ such that $\log^2 d \leq t \leq \frac{d}{100}$, $k = \lfloor \frac{d}{2t} \rfloor$, and $\ell = \lceil \frac{5d^2 t}{\log d} \rceil$,*

$$\Gamma_{k,\ell}(NW_{t,n}) \geq \frac{1}{d^3} \binom{d^2 + \ell + d - k}{d^2}$$

We will also use the following result, which is implicit in the calculations in the proof of Theorem 2 in [KSS14] in our calculations.

Theorem 2.21 ([KSS14]). *Let d be a prime power. For any fixed constant α and t, k, ℓ such that $\log^2 d \leq t \leq \frac{d}{100}$, $k = \lfloor \frac{d}{2t} \rfloor$ and $\ell = \lceil \frac{5d^2 t}{\log d} \rceil$, if*

$$E = \frac{\frac{1}{d^3}\binom{d^2+\ell+d-k}{d^2}}{\binom{\frac{\alpha d}{t}}{k}\binom{d^2+\ell+k(t-1)}{d^2}}$$

Then, $E \geq \exp(\Omega(\frac{d}{t} \log d))$.

For the range of values of t stated above, the value of k lies in the range $200 \leq k \leq \frac{d}{2 \log^2 d}$. To complete the proof, we will argue that after setting the y variables to a constant, there is a value of k in this range and an ℓ such that the dimension of the shifted partial derivative span of the circuit is small . Based on this value of k, we will then invoke a particular projection $NW_{t,n}$ of NW_n and then use the bound from Theorem 2.20.

Proof of Theorem 2.2. Let us consider a $\Sigma\Pi\Sigma\Pi(r)$ circuit of size s which computes the polynomial NW_n. As discussed, we will analyze the shifted partial derivative complexity of the circuit and the polynomial under the restriction that the $\{y_1, y_2, \ldots, y_d\}$ variables are set to constants. So, the degree of the polynomial computed is d and the number of alive variables is $n = d^2$.

Let $0 < \varepsilon < 1$ be a constant. We will now show that we can choose a value of k such that the conditions in Theorem 2.18 and Theorem 2.20 hold. From the proof of Theorem 2.18, we know that there are at most $\frac{r}{\varepsilon}$ many choices of integer $0 < i < \frac{2r}{\varepsilon}$ that are bad i.e that $k = d^{\frac{\varepsilon \cdot i}{2r}}$ does not give us the upper bound on the complexity of the shifted partial derivatives as stated in Theorem 2.18. Now, all we need to show is that there is one such "good" i such that $200 \leq k = d^{\frac{\varepsilon \cdot i}{2r}} \leq \frac{d}{2\log^2 d}$. For this to hold, we need to show a "good" i in the range $\frac{2r}{\varepsilon \log d} \log 200 < i < \frac{2r}{\varepsilon}(1 - \frac{1+2\log\log d}{\log d})$. The number of integers in this range is at least $\frac{2r}{\varepsilon}(1 - 3\frac{\log\log d}{\log d})$, while the number of bad i is at most $\frac{r}{\varepsilon}$. Hence, for d large enough, there is an i such that for the resulting k, the bound in Theorem 2.18 holds and $t = \frac{d}{2k}$ satisfies $\log^2 d \leq t \leq \frac{d}{100}$. Let us fix such a good k. Let us now fix $t = \frac{d}{2k}$, $\ell = \frac{5d^2 t}{\log d}$ and $\varepsilon = \frac{1}{8}$. Now, let us consider the restriction of C when just y_t is set to 1 and y_j is set to zero for every $j \neq t$. In this case, the circuit just computes $NW_{t,n}$. From Theorem 2.18, we get

$$\Gamma_{k,\ell}(C) \leq r \cdot s^{k \cdot d^{-\varepsilon/2r}} \cdot \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n}$$

and from Theorem 2.20, we get

$$\Gamma_{k,\ell}(NW_{t,n}) \geq \frac{1}{d^3} \binom{d^2 + \ell + d - k}{d^2}$$

So, if C computes NW_n, then $\Gamma_{k,\ell}(C) \geq \Gamma_{k,\ell}(NW_{t,n})$. Thus

$$r \cdot s^{k \cdot d^{-\varepsilon/2r}} \cdot \binom{k/\varepsilon + k - 1}{k} \binom{n + 4\varepsilon d - k + \ell}{n} \geq \frac{1}{d^3} \binom{d^2 + \ell + d - k}{d^2}.$$

Substituting the parameters and $n = d^2$, we get

$$r \cdot s^{k \cdot d^{-1/16r}} \geq \frac{\frac{1}{d^3}\binom{d^2+\ell+d-k}{d^2}}{\binom{\frac{\alpha d}{t}}{k}\binom{d^2+\ell+k(t-1)}{d^2}}$$

for some appropriate constant α dependent on ε. From Theorem 2.21, this implies that

$$r \cdot s^{k \cdot d^{-1/16r}} \geq \exp(\Omega(\frac{d}{t}\log d)) = \exp(\Omega(k \log d))$$

Using the fact that r is at most s (in fact it is much much smaller), we conclude that

$$k \cdot d^{-1/16r} \cdot \log s \geq \Omega(k \log d).$$

Thus

$$\log s \geq \Omega(d^{1/16r} \log d)$$

and hence

$$s \geq \exp\left(d^{\Omega(1/r)} \log d\right).$$

Since $\log d = \Theta(\log n)$, we get

$$s \geq \exp\left(d^{\Omega(1/r)} \log n\right).$$

□

A very similar calculation lets us prove Theorem 2.3.

Proof of Theorem 2.3. For a $\Sigma\Pi\Sigma\Pi^*$ circuit, the calculation will proceed exactly the same as above, and in the end, we will get

$$s \geq \exp\left(d^{\Omega(1/m)}\right),$$

which on substituting $m = 1/\varepsilon + 1$, completes the proof. Thus, we obtain exponential lower bounds for $\Sigma\Pi\Sigma\Pi^*$ circuits computing the polynomial NW_n regardless of their top fan-in. □

2.5 Depth reduction is tight for $\Sigma\Pi\Sigma\Pi(\Omega(\log d))$ circuits

In this section, we will show that the depth reduction procedure of Koiran and Tavenas [Koi12, Tav15] as given in Theorem 1.8 gives an almost optimal upper bound on the size of homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit obtained. On the way to this result, we will prove a *Hierarchy* theorem (Theorem 2.5) for formulas of depth-4 with bounded bottom fan-in. We will then build up on this proof, and prove Theorem 2.4 and Theorem 2.6. We will first provide an overview of the proof.

Proof Overview

We will construct an explicit family of polynomials $\{\mathcal{Q}_n\}$, such that $\mathcal{Q}_n$ is a homogeneous polynomial in $\Theta(n)$ variables of degree $d = \Theta(\sqrt{n})$ which can be computed by a

polynomial sized homogeneous $\Sigma\Pi\Sigma\Pi(O(\log d))$ circuit. We will show that there exists a constant a_0 such that for each $a \geq a_0$, $\mathcal{Q}_n$ requires homogeneous $\Sigma\Pi\Sigma\Pi^{[a]}$ circuits of top fan-in $2^{\Omega(\frac{d}{a}\log n)}$. In order to construct this polynomial family, we will construct for each $t \geq a_0$, a family of polynomials $\{\mathcal{P}_{t,n}(\mathbf{x})\}$, such that each $\mathcal{P}_{t,n}(\mathbf{x})$ is a homogeneous polynomial in d^2 variables and of degree d, and can be computed by a polynomial sized homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit. Moreover, we will show that for every constant $\delta < 1$, any homogeneous $\Sigma\Pi\Sigma\Pi^{[\delta t]}$ circuit computing it must have top fan-in at least $2^{\Omega(\frac{d}{t}\log n)}$. We will then apply the interpolation trick of [KSS14] to $\mathcal{P}_{t,n}$ for various t to obtain the $\mathcal{Q}_n$. The construction is heavily inspired by the idea of constructing hard polynomials using Nisan-Wigderson designs used in [KSS14]. To show the lower bound for each t, we will use ideas from [CM14a] and [FLMS14], and show that for suitable k, $\partial^{=k}P(\mathbf{x})$ has a large number of elements whose leading monomials are at a "large distance" from each other.

2.5.1 Proof of Theorem 2.5

For the rest of this section, we will assume that d is a prime power. For each such d, we will identify the elements of the field $\mathbb{F}_d$ with the elements of the set $[d] = \{1, 2, \ldots, d\}$. For a parameter t which is a positive integer less than d, let us now partition the set $[d]$ into $\lceil \frac{d}{t} \rceil$ parts which are roughly equal and each is of size about t. For brevity, we will indicate $\frac{d}{t}$ by $\tilde{t}$. We will let $C_i = \{t(i-1)+1, t(i-1)+2, \ldots, ti\}$ denote the i^{th} such partition. Also, for every $j \in [\tilde{t}]$ and $i \leq t$, let C_j^i be the set of the i smallest elements in C_j.

In the rest of the chapter, we will use $\mathbf{x}$ to denote the set of $n = d^2$ variables $\{x_{i,j} : i, j \in [d]\}$ and $\mathbf{y}$ to denote the set of variables $\{y_1, y_2, \ldots, y_d\}$. We will use the following notion of distance between two monomials as defined in [CM14a].

Definition 2.22 ([CM14a])**.** *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.* $\Diamond$

We will also use the following notion of distance between strings or ordered tuples. For any two strings s_1, s_2 of the same length, the distance between them $\Delta(s_1, s_2)$ is the number of coordinates at which s_1, s_2 disagree with each other. For brevity, we will use αm to refer to $\lfloor \alpha m \rfloor$ for any positive integer m and any real number α.

Based on the notations defined, we now define the polynomials $\mathcal{P}_{t,n}$ on $n = d^2$ variables $\mathbf{x} = \{x_{i,j} : i \in [d], j \in [d]\}$.

$$\mathcal{P}_{t,n}(\mathbf{x}) = \prod_{j \in [\tilde{t}]} \sum_{u \in [d]} \prod_{i \in C_j} x_{i,u}$$

From the expression above, it follows that for every d, t, $\mathcal{P}_{t,n}$ can be computed by a polynomial sized $\Pi\Sigma\Pi$ formula. Observe that in fact it can be computed by a *regular* formula[12]. We summarize this observation below.

Observation 2.23. *For every d, t and $n = d^2$, $\mathcal{P}_{t,n}$ can be computed by a $\Pi\Sigma\Pi$ regular formula of size polynomial in n.*

Our goal now is to try and show a lower bound on the dimension of the shifted partial derivatives of $\mathcal{P}_{t,n}$. To this end, we will first choose a structured set of monomials with respect to which we take partial derivatives, and use the resulting structure to show a lower bound.

From the definition of $\mathcal{P}_{t,n}$, every monomial in it can be identified by an ordered tuple of length $\tilde{t}$ over the set $[d]$ and vice-versa. So, for any $\overline{u} = (u_1, u_2, \ldots, u_{\tilde{t}}) \in [d]^{\tilde{t}}$, let

$$m_{\overline{u}} = \prod_{j \in [\tilde{t}]} \prod_{i \in C_j} x_{i,u_j(i)}$$

From the definitions above and that of $\mathcal{P}_{t,n}(\mathbf{x})$, it follows that

$$\mathcal{P}_{t,n}(\mathbf{x}) = \sum_{\overline{u} \in [d]^{\tilde{t}}} m_{\overline{u}}$$

Let

$$m'_{\overline{u}} = \prod_{j \in [\tilde{t}]} \prod_{i \in C_j^1} x_{i,u_j(i)}$$

[12]When t divides d the formula will be exactly regular, and if not a simple modification could make it regular, but the details are simple.

For $t > 1$

$$\frac{m_{\overline{u}}}{m'_{\overline{u}}} = \prod_{j \in [\tilde{t}]} \prod_{i \in C_j \setminus C_j^1} x_{i,u_j(i)}$$

Now, we set $k = \tilde{t}$ and look at the partial derivatives of $\mathcal{P}_{t,n}$ of order k. For each $\overline{u} \in [d]^{\tilde{t}}$, the degree of $m'_{\overline{u}}$ equals k. Hence, $\partial^{=k}\mathcal{P}_{t,n}$ includes the set of partial derivatives of $\mathcal{P}_{t,n}$ with respect to $m'_{\overline{u}}$ for each $\overline{u} \in [d]^{\tilde{t}}$. From the definition of $m_{\overline{u}}$ and $m'_{\overline{u}}$, for any $\overline{u} \in [d]^{\tilde{t}}$ and $\overline{v} \in [d]^{\tilde{t}}$,

$$\partial_{m'_{\overline{u}}} m_{\overline{v}} = \begin{cases} 0 & \overline{u} \neq \overline{v} \\ \frac{m_{\overline{u}}}{m'_{\overline{u}}} & \overline{u} = \overline{v} \end{cases}$$

From this discussion, the following lemma follows.

Lemma 2.24. *For every $\overline{u} \in [d]^{\tilde{t}}$, $\partial_{m'_{\overline{u}}} \mathcal{P}_{t,n}$ is a monomial and equals $\frac{m_{\overline{u}}}{m'_{\overline{u}}}$.*

At this point, we might hope to argue that for each $\overline{u} \in [d]^{\tilde{t}}$ and $\overline{v} \in [d]^{\tilde{t}}$ such that $\overline{u} \neq \overline{v}$, the distance between the monomials $\frac{m_{\overline{u}}}{m'_{\overline{u}}}$ and $\frac{m_{\overline{v}}}{m'_{\overline{v}}}$ is large. This statement in itself is not true, for if $\overline{u}$ and $\overline{v}$ differ in just one coordinate, then the distance between $\frac{m_{\overline{u}}}{m'_{\overline{u}}}$ and $\frac{m_{\overline{v}}}{m'_{\overline{v}}}$ could be as small as t, which as it turns out is insufficient for our proof. Observe that if $\overline{u}$ and $\overline{v}$ differ in i coordinates, then the distance between $\frac{m_{\overline{u}}}{m'_{\overline{u}}}$ and $\frac{m_{\overline{v}}}{m'_{\overline{v}}}$ is at least $i(t-1)$. (We prove this fact in Lemma 2.27).

To prove the lower bound, we will show that there is a "large" nice subset $\mathcal{N} \subseteq [d]^{\tilde{t}}$ such that any distinct $\overline{u}$ and $\overline{v}$ in $\mathcal{N}$ differ in a constant fraction of all coordinates. The following lemma, which just follows from the existence and properties of Reed-Solomon codes, guarantees the existence of such an $\mathcal{N}$.

Lemma 2.25. *Let $0 < \alpha < 1$ be any absolute constant and let q be a prime power. For any alphabet Σ of size q and positive integer m such that $m < q$, there is a set $\mathcal{C}$ of strings of length m over Σ of size $q^{(1-\alpha)m}$ such that any two strings in $\mathcal{C}$ are at a distance at least αm apart.*

Proof. Let $\mathcal{C}$ be the set of codewords obtained when the set $\Sigma^{(1-\alpha)m}$ is encoded using Reed-Solomon codes of message length $(1-\alpha)m$ and code length m. The distance of

the code is αm and the number of codewords is $q^{(1-\alpha)m}$. Hence the set satisfies the properties stated in the statement. □

Lemma 2.25 immediately implies the existence of a set $\mathcal{N}$, when invoked with parameters $\Sigma = [d]$, $m = \tilde{t}$. So, we have the following corollary.

Corollary 2.26. *For all α such that $0 < \alpha < 1$, there exists $\mathcal{N} \subseteq [d]^{\tilde{t}}$ of size equal to $d^{(1-\alpha)\tilde{t}}$ such that for any distinct pair $\overline{u}$ and $\overline{v}$ in $\mathcal{N}$, $\overline{u}$ and $\overline{v}$ differ in at least $\alpha\tilde{t}$ coordinates.*

Informally, the set $\mathcal{N}$ now gives us a large number of partial derivatives which are at a large distance from each other. We formalize this claim in the lemma below.

Lemma 2.27. *For $k = \tilde{t}$, the set $\partial^{=k}\mathcal{P}_{t,n}$ has a subset S of size at least $d^{(1-\alpha)\tilde{t}}$ such that every element in this subset is a monomial and any two such monomials are at a distance of at least $\alpha\tilde{t}(t-1)$ from each other.*

Proof. Let us pick any two $\overline{u}$ and $\overline{v}$ in $\mathcal{N}$. Let $i \in [\tilde{t}]$ be an index such that $u_i \neq v_i$. Then, the monomials $m_{\overline{u}}$ and $m_{\overline{v}}$ differ in at least t variables of the form $x_{h,j}$ for $h \in C_i$. Hence, $\frac{m_{\overline{u}}}{m_{\overline{u}}}$ and $\frac{m_{\overline{v}}}{m_{\overline{v}}}$ differ in at least $t-1$ variables of the form $x_{h,j}$ for $h \in C_i$. So, each coordinate i where $\overline{u}$ and $\overline{v}$ differ from each other contributes $t-1$ to the distance between $\frac{m_{\overline{u}}}{m_{\overline{v}}}$ and $\frac{m_{\overline{v}}}{m_{\overline{v}}}$. Hence, for every distinct $\overline{u}$ and $\overline{v} \in \mathcal{N}$, $\frac{m_{\overline{u}}}{m_{\overline{u}}}$ and $\frac{m_{\overline{v}}}{m_{\overline{v}}}$ are at a distance at least $\alpha\tilde{t}(t-1)$ apart. The lemma now follows from the fact that the size of $\mathcal{N}$ is at least $d^{(1-\alpha)\tilde{t}}$. □

Using Lemma 2.27, we now show that the dimension of the shifted partial derivative span of $\mathcal{P}_{t,n}$ is quite large for an appropriate choice of parameters.

Let ε be a small enough constant.[13] Let the parameter ℓ be such that

$$\frac{n+\ell}{\ell} = d^{\frac{\varepsilon}{t}}.$$

Recall that $n = d^2$.

[13]We will set the value of ε later in this discussion based on the value of t and α.

We use the following lemma in Chillara and Mukhopadhyay [CM14a]. The proof is a simple but clever application of the principle of inclusion-exclusion. We refer the reader to Lemma 3 in [CM14a] for the proof.

Lemma 2.28 ([CM14a])**.** *Let W be a subset of partial derivatives of a polynomial P of order k such that the distance between the leading monomials of any two distinct polynomials in W is at least Δ. Then, for every ℓ,*

$$\Gamma_{k,\ell}(P) \geq |W| \cdot \binom{n+\ell}{\ell} - |W|^2 \cdot \binom{n+\ell-\Delta}{\ell-\Delta}$$

For the choice of parameters in this chapter, we have the following corollary.

Corollary 2.29. *Let Δ be such that $\Delta = \alpha\tilde{t}(t-1)$. For every constant ε such that $\varepsilon < 1$ and $\varepsilon\alpha < (1-\alpha)$ and for ℓ such that $\frac{n+\ell}{\ell} = d^{\frac{\varepsilon}{t}}$,*

$$\Gamma_{k,\ell}\left(\mathcal{P}_{t,n}\right) \geq \frac{1}{2d} \cdot \left(\frac{n+\ell}{\ell}\right)^{\Delta} \cdot \binom{n+\ell}{\ell} \cdot \exp\left(-\frac{\Delta^2}{\ell}\right)$$

Proof. From Lemma 2.27, we know that there is a subset $\mathcal{S}$ of derivatives of $\mathcal{P}_{t,n}$ of order k such that for any two two polynomials in $\mathcal{S}$, the leading monomials are at a distance at least $\Delta = \alpha\tilde{t}(t-1)$ far from each other. We also know that the size of $\mathcal{S}$ is at least $d^{(1-\alpha)\tilde{t}}$. Now observe that

$$\left(\frac{n+\ell}{\ell}\right)^{\Delta} \leq d^{\varepsilon\alpha\tilde{t}}$$

for any constant $\varepsilon < 1$ such that $\varepsilon\alpha < (1-\alpha)$. Therefore, for such a choice of parameters,

$$|\mathcal{S}| \geq \left(\frac{n+\ell}{\ell}\right)^{\Delta}$$

Let W be a subset of $\mathcal{S}$ of size $1/d \cdot \left(\frac{n+\ell}{\ell}\right)^{\Delta} \cdot \exp\left(-\frac{\Delta^2}{\ell}\right)$. From Lemma 2.28, we know that

$$\Gamma_{k,\ell}\left(\mathcal{P}_{t,n}\right) \geq |W| \cdot \binom{n+\ell}{\ell} - |W|^2 \cdot \binom{n+\ell-\Delta}{\ell-\Delta}$$

We now show that for our choice of parameters, the following inequality is true.

$$\frac{1}{2} \cdot |W| \cdot \binom{n+\ell}{\ell} \geq |W|^2 \cdot \binom{n+\ell-\Delta}{\ell-\Delta}$$

which is the same as showing that

$$|W| \leq \frac{1}{2} \cdot \frac{\binom{n+\ell}{\ell}}{\binom{n+\ell-\Delta}{\ell-\Delta}}$$

Observe that

$$\frac{\binom{n+\ell}{\ell}}{\binom{n+\ell-\Delta}{\ell-\Delta}} = \frac{(n+\ell)!}{(n+\ell-\Delta)!} \cdot \frac{(\ell-\Delta)!}{\ell!}$$

From the choice of ℓ, we know that $\Delta = o(\ell)$ when $t > 1$ and $n = \Omega(\Delta^2)$. By Lemma 2.13, we have

$$\frac{\binom{n+\ell}{\ell}}{\binom{n+\ell-\Delta}{\ell-\Delta}} \approx \left(\frac{n+\ell}{\ell}\right)^{\Delta} \cdot \exp\left(-\frac{\Delta^2}{\ell}\right)$$

Therefore, from the choice of size of W, it follows that for every constant c,

$$|W| \leq c \cdot \frac{\binom{n+\ell}{\ell}}{\binom{n+\ell-\Delta}{\ell-\Delta}}$$

Hence,

$$\Gamma_{k,\ell}\left(\mathcal{P}_{t,n}\right) \geq \frac{1}{2} \cdot |W| \cdot \binom{n+\ell}{\ell} = \frac{1}{2d} \cdot \left(\frac{n+\ell}{\ell}\right)^{\Delta} \cdot \binom{n+\ell}{\ell} \cdot \exp\left(-\frac{\Delta^2}{\ell}\right)$$

□

We also need the following upper bound on the dimension of shifted partial derivatives of a homogeneous depth four circuit with bounded bottom fan-in.

Lemma 2.30 ([GKKS14])**.** *Let δ be any positive constant less than 1. Let C be a homogeneous $\sum\prod\sum\prod^{[\delta\cdot t]}$ circuit of top fan-in s computing a degree d polynomial. Then, for every k and ℓ*

$$\Gamma_{k,\ell}(C) \leq s \cdot \binom{O(d/t)+k}{k}\binom{n+\ell+\delta kt}{\ell+\delta kt}.$$

We now have all the ingredients we need for showing lower bounds for homogeneous $\sum\prod\sum\prod^{[\delta\cdot t]}$ circuits computing $\mathcal{P}_{t,n}$.

Theorem 2.31. *Let δ be any positive constant less than 1. Then, there exists a constant t_0 such that for any $t > t_0$, any homogeneous $\sum\prod\sum\prod^{[\delta\cdot t]}$ circuit computing $\mathcal{P}_{t,n}$ has top fan-in at least $2^{\Omega(\frac{d}{t}\log n)}$.*

Proof. Let C be a $\sum\prod\sum\prod^{[\delta\cdot t]}$ circuit of top fan-in s computing $\mathcal{P}_{t,n}$. We pick $k = \tilde{t}$ and ℓ according to Corollary 2.29. Let Δ be such that $\Delta = \alpha\tilde{t}(t-1)$. Then, by Corollary 2.29, we have

$$\Gamma_{k,\ell}\left(\mathcal{P}_{t,n}\right) \geq \frac{1}{2d} \cdot \left(\frac{n+\ell}{\ell}\right)^{\Delta} \cdot \binom{n+\ell}{\ell} \cdot \exp\left(-\frac{\Delta^2}{\ell}\right)$$

And by Lemma 2.30, we have

$$\Gamma_{k,\ell}(C) \leq s \cdot \binom{O(d/t)+k}{k}\binom{n+\ell+\delta kt}{\ell+\delta kt}.$$

Since C computes $\mathcal{P}_{t,n}$, it must be the case that

$$s \cdot \binom{O(d/t)+k}{k}\binom{n+\ell+\delta kt}{\ell+\delta kt} \geq \frac{1}{2d} \cdot \left(\frac{n+\ell}{\ell}\right)^{\Delta} \cdot \binom{n+\ell}{\ell} \cdot \exp\left(-\frac{\Delta^2}{\ell}\right)$$

In other words,

$$s \geq \frac{\frac{1}{2d} \cdot \left(\frac{n+\ell}{\ell}\right)^{\Delta} \cdot \binom{n+\ell}{\ell} \cdot \exp\left(-\frac{\Delta^2}{\ell}\right)}{\binom{O(d/t)+k}{k}\binom{n+\ell+\delta kt}{\ell+\delta kt}}$$

Simplifications as in the proof of Corollary 2.29 and using the Lemma 2.13, we get

$$s \geq \frac{1}{2d} \cdot 2^{-O(d/t)} \cdot \left(\frac{n+\ell}{\ell}\right)^{\Delta-\delta kt} \cdot \exp\left(-\frac{\Delta^2}{\ell} - \frac{k^2t^2}{400\ell}\right)$$

Simplifying further, and substituting the value of ℓ, we obtain

$$s \geq \frac{1}{d^2} \cdot \left(2^{-O(d/t)} \cdot d^{\frac{\varepsilon}{t}\cdot(\Delta-\delta kt)} \cdot \exp\left(-d^{\varepsilon/t}\right)\right)$$

Substituting the values of Δ, k, we get

$$s \geq \frac{1}{d^2} \cdot \left(2^{-O(d/t)} \cdot d^{\frac{\varepsilon d}{t}\cdot(\alpha-\delta)} \cdot \exp\left(-d^{\varepsilon/t}\right)\right)$$

Now, based on δ, we can choose the parameter α such that $\alpha > \delta$ and $\alpha > 1/2$. Then, we can pick an ε such that $\varepsilon\alpha < (1-\alpha)$. Now, for any fixed choice of α and ε, observe that $d^{\varepsilon/t}$ decreases much faster as a function of t when compared to $\varepsilon d/t$. So, it is clear that there is a constant t_0 (for example $t_0 = 10\varepsilon$) such that for any $t > t_0$,

$$s \geq d^{\Omega(\frac{d}{t})}$$

Now, using $n = d^2$, the result follows. □

This completes the proof of Theorem 2.5. We will now build upon this proof to obtain Theorem 2.4.

2.5.2 Proof of Theorem 2.4

So far, we have constructed a polynomial family $\mathcal{P}_{t,n}$ such that $\mathcal{P}_{t,n}$ requires homogeneous $\Sigma\Pi\Sigma\Pi^{[\frac{t}{20}]}$ circuits with top fan-in at least $n^{\Omega(\frac{d}{t})}$. We can now build upon the construction of $\mathcal{P}_{t,n}$ described so far to construct a single polynomial family which is hard for any homogeneous $\Sigma\Pi\Sigma\Pi^{[a]}$ circuit for every $a \geq t_0$. We will now use a variation of the interpolation trick described in Lemma 14 in [KSS14]. The idea now is just to take a linear combination of $\mathcal{P}_{b,n}$ for $O(\log d)$ many such values of b in a geometric progression with some constant common ratio $\gamma > 1$, with coefficients being the variables $\mathbf{y}$, so that for every a such that $a \geq t_0$, there is a b such that $\gamma a \leq b \leq \gamma^2 a$ and such that $\mathcal{P}_{b,n}$ is in the linear combination.

In particular let us define the following family of polynomials $\mathcal{Q}_n$:

$$\mathcal{Q}_n(\mathbf{x}, \mathbf{y}) = \sum_{i=0}^{O(\log d)} y_i \cdot \mathcal{P}_{\gamma^i,n}(\mathbf{x})$$

Observe that $\mathcal{Q}_n$ can be computed by a polynomial size homogeneous $\Sigma\Pi\Sigma\Pi(\log d)$ circuit.

If $\mathcal{Q}_n$ could be computed *efficiently* by a homogeneous $\Sigma\Pi\Sigma\Pi^{[a]}$ for some a, then so could any projection of the sum (i.e. we set all but one of the y_i to 0), i.e. so could $\mathcal{P}_{b,n}$. This contradicts Theorem 2.5. In particular we get that for every a such that a is large enough, any homogeneous $\Sigma\Pi\Sigma\Pi^{[a]}$ circuit computing $\mathcal{Q}_n$ must have top fan-in at least $2^{\Omega(\frac{d}{a}\log n)}$. This completes the proof of Theorem 2.4.

2.5.3 Proof of Theorem 2.6

The following theorem shown in [KSS14] provides a connection between lower bounds for homogeneous depth four circuits with bounded bottom fan-in and lower bounds for regular formulas.

Theorem 2.32 ([KSS14])**.** *Let P be a polynomial in n variables and degree d with the property that there exists a $\delta > 0$ such that for every $\log^2 d < t < d/100$, any homogeneous $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuit computing the polynomial P has top fan-in at least $\exp(\delta \cdot \frac{d}{t}\log n)$. Then any regular formula computing P must be of size $n^{\Omega(\log d)}$.*

The proof of Theorem 2.6 follows immediately from Theorem 2.4 and the above theorem.

Chapter 3

Superpolynomial lower bounds for general homogeneous depth-4 arithmetic circuits[1]

3.1 Introduction

In this chapter, we prove 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 superpolynomial 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 3.1 (Lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits)**.** *There is an explicit family of homogeneous polynomials of degree n in $N = n^2$ 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 a variant of the family of Nisan-Wigderson polynomials NW. We give the formal definition in Section 3.3.

As a first step in the proof of Theorem 3.1, 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

[1]The results in this chapter appear in [KS14].

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 3.2 (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 [GKKS14, KSS14] when $s = O(\log n)$.

We prove Theorem 3.1 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 3.2 to hold. Our results hold over every field.

Organization of the chapter : The rest of the chapter is organized as follows. In Section 3.2, we provide a high level overview of the proof. In Section 3.3, we introduce some notations and preliminary notions used in the chapter. In Section 3.4, we give a proof of Theorem 3.2. In Section 3.5, we describe the random restriction procedure and analyze its effect on the circuit and the polynomial. In Section 3.6, we prove Theorem 3.1.

3.2 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 [GKKS14, KSS14, KS15d, FLMS14]. For this measure, we show that the complexity of the NW 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 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, 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 $1/n^{\varepsilon}$. 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. Thus we define a slightly more refined random restriction procedure which keeps the NW 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 Subsection 3.5.1

3.3 Preliminaries and Notations

Homogeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ Circuits: A homogeneous $\Sigma\Pi\Sigma\Pi$ circuit 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.

Bounded support shifted partial derivatives: In this chapter, 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 3.3 (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*

$$\langle \partial^{=r} P \rangle_{(\ell,m)} \stackrel{def}{=} \mathbb{F}\text{-}span\{ \prod_{\substack{i \in S \\ S \subseteq [N] \\ |S|=m}} {x_i}^{j_i} \cdot g : \sum_{i \in S} j_i = \ell, j_i \geq 1, g \in \partial^{=r} P\} \tag{3.4}$$

$\Diamond$

The following property follows from the definition above.

Lemma 3.5. *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 chapter, 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 variant of the family of Nisan-Wigderson design polynomials. For the rest of this chapter, 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

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

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

1. The number of monomials in NW is exactly n^d.
2. Each of the monomials in NW 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,

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

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

Definition 3.6. *A monomial θ is said to be an extension of a monomial $\tilde{\theta}$, if θ divides $\tilde{\theta}$.* ◊

In this chapter, 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 [CM14a].

Definition 3.7 (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.* $\lozenge$

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

Observation 3.8. *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*

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

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

Fact 3.9. *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.

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

In this section, we will prove Theorem 3.2. 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 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 polynomial. Comparing the two will then imply our lower bound. The bound holds for NW for any $d = \delta n$, where δ is a constant such that $0 < \delta < 1$.

3.4.1 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 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 3.10. *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 ΣΠΣΠ circuit where each of the bottom level product gates is of support at most s, Lemma 3.10 immediately implies the following upper bound on the complexity of the circuit due to subadditivity from Lemma 3.5.

Corollary 3.11 (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 $m + rs \le \frac{N}{2}$ and $m + rs \le \frac{\ell}{2}$,*

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

3.4.2 Lower bound on the complexity of the NW polynomial

We will now prove a lower bound on the complexity of the NW polynomial. For this, we will first observe that distinct partial derivatives of the NW 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 3.12. *For any positive integer r such that $n - r > d$ and $r < d - 1$, the set $\{\partial_\alpha(\mathsf{NW}) : \alpha \in \mathcal{M}^{[r]}\}$ consists of $|\mathcal{M}^{[r]}| = n^r$ nonzero distinct polynomials.*

Proof. We need to show the following two statements.

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

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, $\partial_\alpha(\mathsf{NW}) = \sum_g \partial_\alpha(m_g)$ 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 $\partial_\alpha(\mathsf{NW})$ and $\partial_\beta(\mathsf{NW})$. 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, $\partial_\alpha(\mathsf{NW})$ and $\partial_\beta(\mathsf{NW})$ do not have a common monomial. Therefore, $\partial_\alpha(\mathsf{NW}) \neq \partial_\beta(\mathsf{NW})$. □

Remark 3.13. *Observe that there is nothing special about the set $\mathcal{M}^{[r]}$ and the Lemma 3.12 holds for $\{\mathcal{M}\}^{\mathcal{S}}$ 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 $\partial_\alpha(\mathsf{NW})$ and $\partial_\beta(\mathsf{NW})$ 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 $\{\partial_\alpha(\mathsf{NW}) : \alpha \in \mathcal{M}^{[r]}\}$ by monomials of support m and degree ℓ results in many linearly independent shifted partial derivatives. We will first prove the following lemma.

Lemma 3.14. *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 3.9. 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 Lemma 3.14 to the NW polynomial. For any monomial α and positive integers ℓ, m, we will denote by $S_{\ell,m}(\alpha)$ the set of all shifts of $\partial_\alpha \mathsf{NW}$ by monomials of degree ℓ and support m. More formally,

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

also, let

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

Lemma 3.15. *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 3.12, we have observed that the leading monomials of $\partial_\alpha(\mathsf{NW})$ and $\partial_\beta(\mathsf{NW})$ 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 $\gamma_\alpha \partial_\alpha(\mathsf{NW}) = \gamma_\beta \partial_\beta(\mathsf{NW})$, 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 3.14, 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 polynomial. For this, we will use the following proposition from [GKKS14], the proof of which is a simple application of Gaussian elimination.

Lemma 3.16 ([GKKS14])**.** *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 $\mathsf{Dim}(\langle \partial^{=r}\mathsf{NW}\rangle_{(\ell,m)})$, 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 3.17. *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,*

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

Proof. Recall that $\mathcal{M}^{[r]} = \{\prod_{i=1}^r \prod_{j\in[n]} x_{i,j}\}$. We have argued in Lemma 3.12 that for each $\alpha, \beta \in \mathcal{M}^{[r]}$, such that $\alpha \neq \beta$, $\partial_\alpha(\mathsf{NW}) \neq \partial_\beta(\mathsf{NW})$ 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 $\alpha \in \mathcal{M}^{[r]}$, $\mathsf{Lead\text{-}Mon}(\gamma\partial_\alpha(\mathsf{NW})) \neq \mathsf{Lead\text{-}Mon}(\tilde{\gamma}\partial_\alpha(\mathsf{NW}))$. 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 3.15,

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

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 2.13,

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

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,

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

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

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

□

Remark 3.18. *The proof above shows something slightly more general than a lower bound on just the complexity of the* NW *polynomial. The only property of the* NW *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*.* ◊

3.4.3 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 polynomial, where $d = \delta n$ for some constant δ between 0 and 1.

Theorem 3.19. *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 *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 3.11 and Lemma 3.17, the top fan-in of the circuit must be at least

$$\frac{0.5n^r \binom{N}{m}\binom{\ell-1}{m-1}}{\text{poly}(nrs)\binom{n+r}{r}\binom{N}{m+rs}\binom{\ell+n-r}{m+rs}} \tag{3.20}$$

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 3.11 and Lemma 3.17. 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 Equation 3.20. 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

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

- $\frac{\binom{N}{m}}{\binom{N}{m+rs}} = \frac{(N-m-rs)!(m+rs)!}{(N-m)!m!} = (\frac{m}{N-m})^{rs}$ up to 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. $\frac{\binom{\ell-1}{m-1}}{\binom{\ell+n-r}{m+rs}} = \frac{(\ell-1)!}{(\ell+n-r)} \times \frac{(m+rs)!}{(m-1)!} \times \frac{(\ell-m+n-r-rs)!}{(\ell-m)!} = \text{poly(n)} \times \frac{1}{\ell^{n-r}} \times m^{rs} \times \frac{(\ell-m)^{n-r}}{(\ell-m)^{rs}}$. This simplifies to $\text{poly(n)} \times \left(\frac{m}{\ell-m}\right)^{rs} \times \left(\frac{\ell-m}{\ell}\right)^{n-r}$.

- $\frac{n^r}{\binom{n+r}{r}} \geq \frac{n^r}{\left(\frac{2(n+r)}{r}\right)^r}$. 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 $2^{\Omega(r\log r)} = 2^{\Omega(n)}$. □

3.5 Random Restrictions

In this section, we will describe our random restriction algorithm and analyze the effect of random restrictions on ΣΠΣΠ circuits as well as the NW 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 3.21. *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.

3.5.1 Random Restriction Algorithm

Let $\varepsilon > 0$ 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 $\varepsilon k \times n$ 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 $\varepsilon k \times n$. 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 will survive after the random restriction.

Claim 3.22. $|A_n| \geq n^d / 2^{\varepsilon k n} = n^{d - \varepsilon n}$.

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. □

3.5.2 Effect of random restriction on NW

Let S_0 be the set of variables output by the random restriction procedure R_ε. Let $R_\varepsilon(\mathsf{NW})$ be the polynomial obtained from NW after setting the variables in S_0 to 0. In this section we will show that $R_\varepsilon(\mathsf{NW})$ 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 $R_\varepsilon(\mathsf{NW})$ 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 with respect to monomials in T_S for some choice of S.

Lemma 3.23 (Random restriction on NW)**.** *For every $\varepsilon > 0$, 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 $n^{r(1-\varepsilon n/d)}$ monomials in T_S are such that the partial derivative of $R_\varepsilon(\mathsf{NW})$ 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 e 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 3.22 we know that $|A_n| \geq n^d/2^{\varepsilon k n} = n^{d-\varepsilon n}$. Thus there exists $i \leq d/r$ such that $m_i \geq n^{r(1-\varepsilon n/d)}$. Thus there are $n^{r(1-\varepsilon n/d)}$ 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 $R_\varepsilon(\mathsf{NW})$ extending it, and hence the corresponding partial derivative is nonzero. From Remark 3.13 it follows that each of these partial derivative is distinct. □

3.5.3 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 $R_\varepsilon(C)$ 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 3.24 (Random restriction on $\Sigma\Pi\Sigma\Pi$ circuit)**.** *Let $\varepsilon > 0$ 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.

Lemma 3.25. *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 3.26 (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.* ◊

Lemma 3.27. *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. □

Lemma 3.28. *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 $\frac{1}{n^\varepsilon}$. 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 $|U|/|V| = \frac{1}{n^\varepsilon}$.

□

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

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 3.25, the probability of this happening is at most $\frac{1}{n^{\varepsilon \log t}}$.

□

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

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 3.28, we know that a variable in row i_j, $j \in [t]$, survives with probability at most $\frac{1}{n^\varepsilon}$. 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 $\frac{1}{n^\varepsilon}$. Hence, all of them survive with probability at most $\frac{1}{n^{\varepsilon t}}$. □

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

Lemma 3.31. *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 3.27, 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 3.29, Lemma 3.30 and Lemma 3.31 show that any monomial with large support survives only with a very small probability, which completes the proof of Lemma 3.24. We formally prove this below.

Proof of Lemma 3.24: From Lemma 3.29, 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 $\frac{1}{n^{\varepsilon(\log\frac{\beta}{100}+0.9\log\log n)}}$ (for n large enough). Hence, for any circuit of size at most $n^{\rho\log\log n}$, where $\rho < \varepsilon/2$, 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 3.30, a monomial with nonzero support in at least $\log\log n$ non-compact rows survives with probability at most $\frac{1}{n^{\varepsilon\log\log n}}$. Hence, for circuits of size $n^{\rho\log\log n}$, where $\rho < \varepsilon/2$, 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 $\rho < \varepsilon/2$. □

3.6 Proof of main theorem

In this section, we give a proof of our main theorem. We will heavily borrow from the proof of Theorem 3.19 in Section 3.4. The following lemma provides a lower bound on the complexity of the NW polynomial after restricting it via R_ε.

Lemma 3.32. *Let δ and ε be any constants such that $0 < \varepsilon, \delta < 1$. 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 $r \leq \frac{(n-d)\log\phi \pm O\left(\phi\frac{(n-d-r)^2}{N}\right)}{(1-\varepsilon n/d)\log n + \log\phi}$. Then, for every random restriction R_ε,*

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

Proof. The proof is analogous to the proof of Lemma 3.17 till the point we substitute the value of $\mathcal{M}^{[r]}$ in the calculations in the proof of Lemma 3.17. For $R_\varepsilon(\mathsf{NW})$, the value to be substituted is now $n^{r(1-\varepsilon n/d)}$ as shown in Lemma 3.23. So, we know that

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

as long the parameters satisfy

$$n^{r(1-\varepsilon n/d)} \times \frac{(N-(n-d-r))!}{N!} \times \frac{m!}{(m-(n-d-r))!} \leq 1 \tag{3.33}$$

Now, using the approximation from Lemma 2.13,

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

Now, taking logarithms on both sides in Equation Equation 3.33 and substituting these approximations, we get

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

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

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

and

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

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

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

□

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 $R_\varepsilon(\mathsf{NW})$ polynomial for a constant β. The proof of the lemma is essentially the same as the proof of Theorem 3.19.

Lemma 3.34. *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 $R_\varepsilon(\mathsf{NW})$ 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 3.11 and Lemma 3.17, the top fan-in of the circuit must be at least

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

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 3.11 and Lemma 3.32. 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 $\phi = \frac{N}{m}$, $r \leq \frac{(n-d)\log\phi \pm O\left(\phi\frac{(n-d-r)^2}{N}\right)}{(1-\varepsilon n/d)\log n + \log\phi}$

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 2.13 for approximations.

- $\frac{\binom{N}{m}}{\binom{N}{m+rs}} = \frac{(N-m-rs)!(m+rs)!}{(N-m)!m!} = (\frac{m}{N-m})^{rs}$ up to 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. $\frac{\binom{\ell-1}{m-1}}{\binom{\ell+n-r}{m+rs}} = \frac{(\ell-1)!}{(\ell+n-r)} \times \frac{(m+rs)!}{(m-1)!} \times \frac{(\ell-m+n-r-rs)!}{(\ell-m)!} = \text{poly(n)} \times \frac{1}{\ell^{n-r}} \times m^{rs} \times \frac{(\ell-m)^{n-r}}{(\ell-m)^{rs}}$. This simplifies to $\text{poly(n)} \times \left(\frac{m}{\ell-m}\right)^{rs} \times \left(\frac{\ell-m}{\ell}\right)^{n-r}$.
- $\frac{n^{(1-\varepsilon n/d)r}}{\binom{n+r}{r}} \geq \frac{n^{(1-\varepsilon n/d)r}}{\left(\frac{2(n+r)}{r}\right)^r}$. 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

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

Simplifying further, this is at least

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

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 $\varepsilon n/d = \varepsilon/\delta << 0.1$ 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 $(\varepsilon n/d) r \log n$. 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 3.35. *Let $d = \delta n$ for any constant δ such that $0 < \delta < 1$. Any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing the* NW *must have size at least $n^{\Omega(\log \log n)}$.*

Proof. For every value of δ, such that $0 < \delta < 1$, choose the parameters $\varepsilon = \tilde{\varepsilon}, \beta = \tilde{\beta}$ such that Lemma 3.34 is true for $\tilde{d} = \delta n$. Now, let us choose a constant $\rho = \tilde{\rho}$ such that Lemma 3.24 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 3.24 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 $R_{\tilde{\varepsilon}}(NW_{\tilde{d}})$. 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 3.34 above, the top fan-in of $R_{\tilde{\varepsilon}}(C)$ must be at least $2^{\Omega(n)}$. Hence, any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing $NW_{\tilde{d}}$ must have size at least $n^{\Omega(\log \log n)}$. □

Chapter 4

On the power of homogeneous depth-4 arithmetic circuits[1]

4.1 Introduction

In Chapter 3, we saw a proof of a lower bound of $n^{\Omega(\log \log n)}$ on the size of any homogeneous depth-4 arithmetic circuit computing an n variate polynomial in VNP . Simultaneously and independently of the results in Chapter 3, Kayal, Limaye, Saha and Srinivasan [KLSS14a] showed a lower bound of $n^{\Omega(\log n)}$ on the size of any homogeneous depth-4 circuit computing an n variate polynomial in VP. Subsequently, Kayal, Limaye, Saha and Srinivasan greatly improved these lower bounds to obtain exponential ($2^{\Omega(\sqrt{n} \log n)}$) lower bounds for a polynomial in VNP (over fields of characteristic zero). Notice that this result also extends the results of [GKKS14] and [KSS14] who proved similar exponential lower bounds for the more restricted class of homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits. The result by [KLSS14a] shows the same lower bound without the restriction of bottom fanin. Again, any asymptotic improvement of this lower bound in the exponent would separate VP from VNP.

This class of results represents an important step forward, since homogeneous depth-4 circuits seem a much more natural class of circuits than homogeneous depth-4 circuits with bounded bottom fanin. The results of the current chapter build upon and strengthen the results of Kayal et al [KLSS14a]. Before we describe our results we first highlight some important questions left open by [KLSS14a] and place them in the context of several of the other recent results in this area.

[1]The results in this chapter appear in [KS17].

- **Dependence on the field:** Several of the major results on depth reduction and lower bounds have heavily depended on the underlying field one is working over. In a beautiful result [GKKS13], it was shown that if one is working over the field of real numbers, one can get surprising depth reduction of general circuits to just *depth*-3 *circuits*[2]! Indeed it was shown that any arithmetic circuit over the reals (in particular one computing the determinant) can be reduced to a depth-3 circuit of size $n^{O(\sqrt{n})}$. Thus proving $n^{\omega(\sqrt{n})}$ lower bounds for depth-3 non-homogeneous circuits over the reals would imply super-polynomial lower bounds for general arithmetic circuits. We know that such a depth reduction is not possible over small finite fields. Lower bounds of the form $2^{\Omega(n)}$ were shown for depth-3 (non-homogeneous) circuits over small finite fields (even for the determinant) by Grigoriev and Karpinksi [GK98] and Grigoriev and Razborov [GR00] [3]. Thus at least for depth-3 circuits, we know that there is a vast difference between the computational power of circuits for different fields.

  The lower bounds of [KLSS14a] work only over fields of characteristic zero. This is because in order to bound the complexity of the polynomial being computed, the proof reduces the question to lower bounding the rank of a certain matrix. This computation ends up being highly nontrivial and is done by using bounds on eigenvalues. However a similar analysis does not go through for other fields. In particular it was an open question if working over characteristic zero was *necessary* in order to prove the lower bounds.

- **Explicitness of the hard polynomial:** The result of [KLSS14a] only proved a lower bound for a polynomial in VNP. It is conceivable/likely that much more should be true, that even polynomials in VP should not be computable by depth-4 homogeneous circuits. The best lower bound known for homogeneous depth-4 circuits computing a poly in VP is the lower bound of $n^{\Omega(\log n)}$ by [KLSS14a]. Recall that when one introduces the restriction on bounded bottom fanin, then

[2] albeit with loss of homogeneity.

[3] Recently, Chillara and Mukhopadhyay [CM14b] showed $2^{\Omega(n \log n)}$ lower bounds for depth-3 circuits over small finite fields for a polynomial in VP.

stronger exponential lower bounds are indeed known [FLMS14, KS15d]. This fact is also related to the next bullet point below.

- **Tightness of depth reduction:** The result of [FLMS14] (which showed an explicit polynomial of degree n in $n^{O(1)}$ variables in VP requiring an $n^{\Omega(\sqrt{n})}$ sized homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ to compute it), in particular showed the the depth reduction results of Koiran [Koi12] and Tavenas [Tav15] (showing that every polynomial of degree n in $n^{O(1)}$ variables in VP can be computed by an $n^{O(\sqrt{n})}$ sized homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuit) are tight. In [KSS14] it was shown that the depth reduction results can in fact be improved for the class of regular arithmetic formulas, thus suggesting that it might be improvable for general formulas or at least homogeneous formulas. This was shown to be false in [KS15d], where it was shown that the depth reduction results of Koiran and Tavenas are tight even for homogeneous formulas. In all these cases, when it was shown that depth reduction is tight, it was shown that if one wants to reduce to the class of homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits, then one cannot do better. The significance of studying depth reduction to homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits stemmed from the matching strong lower bounds for that class.

  Given the new lower bounds for the more natural class of depth-4 homogeneous circuits (with no restriction on bottom fanin), and especially the exponential lower bounds of [KLSS14a], the most obvious question that arises is the following: If one relaxes away the requirement of bounded bottom fanin, i.e. all one requires is to reduce to the class of general depth-4 homogeneous circuits, can one improve upon the upper bounds obtained by Koiran and Tavenas? If we could do this over the reals/complex numbers, then given the [KLSS14a] result, this would also suffice in separating VP from VNP!

- **Shifted partial derivatives and variants:** The results of [KS13, KLSS14a] all use variants of the method of shifted partial derivates to obtain the lower bounds. All 3 works use different variants and they are all able to give nontrivial results. This suggests that we do not really fully understand the potential of these

methods, and perhaps they can be used to give even much stronger lower bounds for richer classes of circuits. Thus it seems extremely worthwhile to develop and understand these methods - to understand how general a class of lower bounds they can prove as well as to understand if there are any limitations to these methods.

4.1.1 Our results

In this chapter, we show a lower bound of $2^{\Omega(\sqrt{n}\log n)}$ on the size of homogeneous depth-4 circuits computing a polynomial in VP. Moreover, this result holds over all fields. We use the notion of the dimension of *projected shifted partial derivatives* as a measure of complexity of a polynomial. This measure was first used in [KLSS14a]. Our results extend those of [KLSS14a] in two ways - they hold over all fields, and they also hold for a much simpler polynomial that is in VP.

We first give a new, more combinatorial proof of the $2^{\Omega(\sqrt{n}\log n)}$ lower bound for a polynomial in VNP, which holds over all fields. This result is much simpler to prove than our result for a polynomial in VP and thus we prove it first. This will also enable us to develop methods and tools for the more intricate analysis of the lower bounds for VP.

Theorem 4.1. *Let $\mathbb{F}$ be any field. There exists an explicit family of polynomials (over $\mathbb{F}$) of degree n and in $N = n^{O(1)}$ variables in VNP, such that any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing it has size at least $n^{\Omega(\sqrt{n})}$.*

The lower bound in Theorem 4.1 is shown for a family of polynomials (denoted by $NW_{n,D}$) whose construction is based on the idea of Nisan-Wigderson designs . These are the same polynomials for which [KLSS14a] show their lower bounds. We give a formal definition in Section 4.3. The main difference in our proof of the above result from the proof in [KLSS14a] is that our proof of the lower bound on the complexity of the polynomial is completely combinatorial, while the proof in [KLSS14a], used matrix analysis that works only over fields of characteristic zero. The combinatorial nature of our proof allows us to prove our results over all fields. The combinatorial nature of the

proof also gives us much more flexibility and this is what enables the proof of our lower bounds for a polynomial in VP. Though our lower bound for the polynomial in VP is at a high level similar to the VNP lower bound, the analysis is much more delicate and the choice of parameters ends up being quite subtle. We will elaborate more on this in the proof outline given inSection 4.2.

Theorem 4.2 (Main Theorem)**.** *Let $\mathbb{F}$ be any field. There exists an explicit family of polynomials (over $\mathbb{F}$) of degree n and in $N = n^{O(1)}$ variables in* VP*, such that any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing it has size at least $n^{\Omega(\sqrt{n})}$.*

We would like to remark that although we state our theorems for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits, the lower bounds continue to hold for $\Sigma\Pi\Sigma\Pi$ circuits which have low formal degree, but might not be homogeneous. The strict notion of homogeneity is not critical for any of our arguments.

As an immediate corollary of the result above, we conclude that the depth reduction results of Koiran [Koi12] and Tavenas [Tav15] are tight even when one wants to depth reduce to the class of general homogeneous depth-4 circuits.

Corollary 4.3 (Depth reduction is tight)**.** *There exists a polynomial in* VP *of degree n in $N = n^{O(1)}$ variables such that any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing it has size at least $n^{\Omega(\sqrt{n})}$. In other words, the upper bound in the depth reduction of Tavenas [Tav15] is tight, even when the bottom fan-in is unbounded.*

The polynomial in Theorem 4.2 is the *Iterated Matrix Multiplication* ($IMM_{\tilde{n},n}$) polynomial. From the fact that the determinant polynomial is complete for the class VQP [Val79], we obtain the first exponential lower bounds for the polynomial Det_n (which is the determinant of an $n \times n$ generic matrix) computed by a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit.

Corollary 4.4. *There exists a constant $\varepsilon > 0$ such that any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing the polynomial Det_n has size at least $2^{\Omega(n^{\varepsilon})}$.*

We have not optimized the value of ε in the statement above, but our proof gives a value of $\varepsilon > 1/22$.

4.1.2 Organisation of the paper

In Section 4.2, we provide a broad overview of the proofs of Theorem 4.1 and Theorem 4.2. In Section 4.3, we define some preliminary notions and set up some notations used in the rest of the paper. We prove an upper bound on the dimension of the projected shifted partial derivatives of a homogeneous depth-4 circuit of bounded bottom support in Section 4.4. We lay down our strategy for obtaining a lower bound on the complexity of the polynomials of interest in Section 4.5. Finally in Section 4.6 and Section 4.7, we prove Theorem 4.1 and in Section 4.8 and Section 4.9, we prove Theorem 4.2.

4.2 Proof Overview

Let C be a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing the polynomial P (either $NW_{n,D}$ or $IMM_{\tilde{n},n}$). The broad outline of the proof of lower bound on the size of C is as follows.

1. If C is *large* ($\geq n^{\varepsilon\sqrt{n}}$) to start with, we have nothing to prove. Else, the size of C is small ($< n^{\varepsilon\sqrt{n}}$).

2. We choose a random subset V of the variables from some carefully defined distribution $\mathcal{D}$, and then restrict P and C to be the resulting polynomial and circuit after setting the variables not in V to zero. We will let $C|_V$ and $P|_V$ be the resulting circuit and polynomial. Since C computed P, thus $C|_V$ still computes $P|_V$. This choice of distribution $\mathcal{D}$ has to be very carefully designed in order to enable the rest of the proof to go through. When $P = NW_{n,D}$, V will be a random subset of variables which is chosen by picking each variable independently with a certain probability. In the case that $P = IMM_{\tilde{n},n}$, our distribution is much more carefully designed. This step is similar to the random restriction step in [FLMS14], although the distributions are slightly different.

3. We show that with a very high probability over the choice of $V \leftarrow \mathcal{D}$, no product gate in the bottom level of $C|_V$ has large support. Thus $C|_V$ is a homogeneous $\Sigma\Pi\Sigma\Pi^{\{\sqrt{n}\}}$ circuit (this is the class of $\Sigma\Pi\Sigma\Pi$ circuits where every product gate at

the bottom layer has only $\sqrt{n}$ distinct variables feeding into it, and we formally define this class in Section 4.3).

4. For any homogeneous $\Sigma\Pi\Sigma\Pi^{\{\sqrt{n}\}}$ circuit, we obtain a good estimate on the upper bound on its complexity $\Phi_{\mathcal{M},m}(C|_V)$ (this is the complexity measure of projected shifted partial derivatives that we use, and we define it formally in Section 4.3) in terms of its size. This step is very similar to that in [KLSS14a], and is fairly straightforward.

5. We show that with a reasonably high probability over $V \leftarrow \mathcal{D}$, the complexity of $P|_V$ remains large. This step is the most technical and novel part of the proof. Unlike the proof of the earlier exponential bound by [KLSS14a], our proof is completely combinatorial. We lower bound the complexity measure $\Phi_{\mathcal{M},m}(P|_V)$ by counting the number of distinct *leading monomials* that can arise after differentiating, shifting and projecting. This calculation turns out to be quite challenging. We first define three related quantities T_1, T_2 and T_3 and show that $T_1 - T_2 - T_3$ is a lower bound on $\Phi_{\mathcal{M},m}(P|_V)$. We elaborate on what these quantities are in Section 4.5. These quantities are easier to compute when $P = NW_{n,D}$, and we are able to show that $\mathbb{E}_{V \leftarrow \mathcal{D}}[T_1 - T_2 - T_3]$ is large. Using variance bounds then lets us conclude that $\Phi_{\mathcal{M},m}(P|_V)$ is large with high probability. When $P = IMM_{\tilde{n},n}$ however, all we are able to show is that $T_2 + T_3$ is not too much larger than T_1 in expected value (it will still be exponentially larger). We then use some sampling arguments to handle this and deduce anyway that $\Phi_{\mathcal{M},m}(P|_V)$ is large. We elaborate more on this step in Subsection 4.5.1 and give formal proofs in Section 4.8 and Section 4.9. In this step of the proof, the choice of the distribution $\mathcal{D}$ turns out to be extremely crucial, and we need to construct it quite carefully. We describe the distribution in Section 4.8.

6. Then, we argue that both the events in the above two items happen simultaneously with non-zero probability. Now, comparing the complexities $P|_V$ and $C|_V$, we deduce that the size of $C|_V$ and hence C must be large.

At a high level, the proof uses several ingredients from [KS13] and [KLSS14a]. We

now highlight the differences between our proof and the proof in each of these.

Comparison to [KS13] The random restriction procedure and the complexity measure in [KS13] is different from the one we use in this work. However the high level strategy of lower bounding the complexity of the polynomial by counting the number of distinct leading monomials that can arise is the same. In this chapter these calculations use much more sophisticated arguments.

Comparison to [KLSS14a] Although the complexity measure and the random restrictions in this chapter are the same as the one used in [KLSS14a], the proofs are different in a key aspect. Kayal et al prove a lower bound on the complexity of the polynomial by using a lemma in real matrix analysis to transform the problem into that of bounding traces of some matrices. This transformation does not work over all fields. In this chapter, we lower bound the complexity of the polynomial using a purely combinatorial argument that counts the number of distinct *leading monomials* that can arise. Hence our proof works over all fields. Although it is hard to say that one of these proofs is simpler than the other (our calculations of the number of distinct leading monomials is fairly nontrivial), we remark that our proof is based on a set of more elementary combinatorial ideas, and the techniques seem to be more flexible (and this is what allowed us to prove the more explicit lower bounds for a polynomial in VP).

4.3 Preliminaries

Support of a polynomial: By the support of a polynomial P, denoted by $\mathrm{Supp}(P)$, we mean the set of monomials which have a non zero coefficient in P. When we consider this set, we will ignore the information in the coefficients of the monomials and just treat them to be 1. We will also use the notion of the support of a monomial α defined as the subset of variables which have degree at least 1 in α. We will follow the notation that when we invoke the function Supp for a monomial, we mean the support in the latter sense. When we invoke it for a polynomial, we mean it in the former sense.

For any monomial α and a set of polynomials $\mathcal{S}$, we define the set $\alpha \cdot \mathcal{S} = \{\alpha\beta : \beta \in \mathcal{S}\}$. For two monomials α and β, we say that α is disjoint from β if the supports of α

and β are disjoint.

Multilinear projections of a polynomial: For any monomial α, we define $\sigma(\alpha)$ to be α if α is multilinear and define it to be 0 otherwise. The map can be then extended by linearity to all polynomials and sets of polynomials.

Homogeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ Circuits: A homogeneous $\Sigma\Pi\Sigma\Pi$ circuit as in Equation 2.7, is said to be a $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuit if every product gate at the bottom level has support at most s (i.e. each monomial in each Q_{ij} has at most s distinct variables feeding into it). Observe that there is no restriction on the bottom fan-in except that implied by the restriction of homogeneity.

Restriction of homogeneous $\Sigma\Pi\Sigma\Pi$ circuit $C|_V$: For a homogeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuit C in variables $v_1, v_2, \ldots, v_N$, and a subset of variables $V \subset \{v_1, v_2, \ldots, v_N\}$, we define $C|_V$ to be the new homogeneous $\Sigma\Pi\Sigma\Pi$ circuit obtained after setting the variables outside V to zero. Equivalently we can think of this as the circuit obtained after removing all multiplication gates at the bottom layer which have a variable not in V that feeds into it.

The complexity measure:

The notion of *shifted partial derivatives* was first introduced in [Kay12] and was subsequently used as a complexity measure in proving several recent lower bound results [FLMS14, GKKS14, KSS14, KS13, KS15d]. In this chapter, we use a variant of the method which first introduced in [KLSS14a].

For a polynomial P and a monomial γ, we denote by $\partial_\gamma(P)$ the partial derivative of P with respect to γ. For every polynomial P and a set of monomials $\mathcal{M}$, we define $\partial_{\mathcal{M}}(P)$ to be the set of partial derivatives of P with respect to monomials in $\mathcal{M}$. We now define the space of $(\mathcal{M}, m)$-projected shifted partial derivatives of a polynomial P below.

Definition 4.5 ($(\mathcal{M}, m)$-projected shifted partial derivatives)**.** *For an N variate polynomial $P \in \mathbb{F}[x_1, x_2, \ldots, x_N]$, set of monomials $\mathcal{M}$ and a positive integer $m \geq 0$, the*

space of $(\mathcal{M}, m)$*-projected shifted partial derivatives of* P *is defined as*

$$\langle \partial_{\mathcal{M}}(P) \rangle_m \stackrel{def}{=} \mathbb{F}\text{-}span\{\sigma(\prod_{i \in S} x_i \cdot g) : g \in \partial_{\mathcal{M}}(P), S \subseteq [N], |S| = m\}$$

$\Diamond$

In this chapter, we carefully choose a set of monomials $\mathcal{M}$ and a parameter m and use the quantity $\Phi_{\mathcal{M},m}(P)$ defined as

$$\Phi_{\mathcal{M},m}(P) = \mathsf{Dim}(\langle \partial_{\mathcal{M}}(P) \rangle_m)$$

as a measure of complexity of the polynomial P.

We will now elaborate on this definition of the measure in words - we look at the space of $(\mathcal{M}, m)$-projected shifted partial derivatives as the space of polynomials obtained at the end of the following steps, starting with the polynomial P.

1. We fix a set of monomials $\mathcal{M}$ and a parameter m.

2. We take partial derivatives of P with every monomial in $\mathcal{M}$, to obtain the set $\partial_{\mathcal{M}}(P)$.

3. We obtain the set of shifted partial derivatives of P by taking the product of every polynomial in $\partial_{\mathcal{M}}(P)$ with every monomial of degree m. In this chapter, we will often be working with restrictions of polynomial P obtained by setting some of the input variables to zero. Even for such restrictions, we consider product of the derivatives by all multilinear monomials of degree m over the complete set of input variables $\{x_1, x_2, \ldots, x_N\}$.

4. Then, we consider each polynomial in the set defined in the item above and project it to the polynomial composed of only the multilinear monomials in its support. The span of this set over $\mathbb{F}$ is defined to be $\langle \partial_{\mathcal{M}}(P) \rangle_m$.

5. We define the complexity of the polynomial $\Phi_{\mathcal{M},m}(P)$ to be the dimension of $\langle \partial_{\mathcal{M}}(P) \rangle_m$ over $\mathbb{F}$.

It follows easily from the definitions that the complexity measure is subadditive. We formalize this in the lemma below.

Lemma 4.6 (Sub-additivity)**.** *Let P and Q be any two multivariate polynomials in $\mathbb{F}[x_1, x_2, \ldots, x_N]$ any set of monomials. Let $\mathcal{M}$ be any set of monomials and m be any positive integer. Then, for all scalars α and β*

$$\Phi_{\mathcal{M},m}(\alpha \cdot P + \beta \cdot Q) \leq \Phi_{\mathcal{M},m}(P) + \Phi_{\mathcal{M},m}(Q)$$

$P|_V$ and $\Phi_{\mathcal{M},m}(P|_V)$: For a polynomial P and a subset of its variables V, we define $P|_V$ to be the polynomial obtained after setting variables not in V to zero (i.e. removing all monomials containing a variable not in V in its support). When we consider $\Phi_{\mathcal{M},m}(P|_V)$, we will be computing the complexity of the new polynomial with respect to the original set of variables, not just the variables in V. I.e. we set the variables outside V to zero only in order to compute $P|_V$. Once we get this new polynomial, we do not think of the variables outside V to be set to zero when computing $\Phi_{\mathcal{M},m}(P|_V)$.

Nisan-Wigderson Polynomials: We now define a variant of the Nisan-Wigderson design polynomials, which is used for one of the lower bounds in this paper. Let $\mathbb{F}_n$ be a finite field of size n. Here, we are assuming for simplicity that n is a prime power. and let F_{n^2} be its quadratic extension. We identify the set $[n]$ with the field $\mathbb{F}_n$ and the set $[n^2]$ with the field F_{n^2}. For the set of $N = n^3$ variables $\{x_{i,j} : i \in [n], j \in [n^2]\}$ and $D < n$, we define the degree n homogeneous polynomial $NW_{n,D}$ as

$$NW_{n,D} = \sum_{\substack{f(z) \in \mathbb{F}_{n^2}[z] \\ deg(f) \leq D-1}} \prod_{i \in [n]} x_{i,f(i)}$$

From the definition, we can observe the following properties of $NW_{n,D}$.

1. The number of monomials in $NW_{n,D}$ is exactly n^{2D}.
2. Each of the monomials in $NW_{n,D}$ 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.

Iterated Matrix Multiplication: Let $M_1, M_2, M_3, \ldots, M_b$ be b generic square matrices, each of dimension $a \times a$. Then, we define the polynomial $IMM_{a,b}$ as the $(1,1)$ entry of the matrix $\prod_j M_j$. It is easy to see that this polynomial can be computed by a polynomial sized circuit, and so is in VP. In this chapter, we show that any homogeneous depth-4 circuit computing $IMM_{a,b}$ has exponential size.

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

Definition 4.7. *A monomial Θ is said to be an extension of a monomial $\tilde{\Theta}$, if $\tilde{\Theta}$ divides Θ.* ◊

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. With respect to this order, we will often look at the leading monomial of various polynomials.

We will use the following notion of distance between two monomials which was also used in [CM14a].

Definition 4.8 (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 chapter, we invoke this definition only for multilinear monomials of the same degree. In this special case, we have the following crucial observation.

Observation 4.9. *Let α and β be two multilinear monomials of the same degree which are at a distance Δ from each other. If Supp(α) and Supp(β) are the supports of α and β respectively, then*

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

For any two multilinear monomials α and β of equal degree, we say that α and β have agreement t if $|\text{Supp}(\alpha) \cap \text{Supp}(\beta)| = t$. When $t = 0$, we say that α and β are disjoint.

Probability lemmas: We will now state some lemmas using probability which will be useful to us in the course of the proof.

Lemma 4.10. *Let X be a random variable sampled from a distribution $\mathcal{R}$ supported on the set R. Let f and g be functions from R to the set of positive real numbers, such that the following are true:*

- *For each $x \in R$, $f(x) \leq g(x)$*
- $\mathbb{E}_{X \leftarrow \mathcal{R}}[f(X)] \geq 0.5 \cdot \mathbb{E}_{X \leftarrow \mathcal{R}}[g(X)]$
- $Pr_{X \leftarrow \mathcal{R}}[|g(X) - \mathbb{E}_{X \leftarrow \mathcal{R}}[g(X)]| \geq 0.1 \cdot (\mathbb{E}_{X \leftarrow \mathcal{R}}[g(X)])] \leq 0.01$

Then,

$$Pr_{X \leftarrow \mathcal{R}}[f(X) \geq 0.01 \cdot (\mathbb{E}_{X \leftarrow \mathcal{R}}[g(X)])] \geq 0.1$$

Proof. We will prove the lemma via contradiction.

So, for the sake of contradiction, let us assume that

$$Pr_{X \leftarrow \mathcal{R}}[f(X) \geq 0.01 \cdot (\mathbb{E}_{X \leftarrow \mathcal{R}}[g(X)])] < 0.1$$

For the rest of the proof, all the probabilities are over $X \leftarrow \mathcal{R}$. Define

- $R_1 = \{x : f(x) < 0.01 \cdot \mathbb{E}[g]\}$
- $R_2 = R \setminus R_1$
- $W = \{x \in R : 0.9 \cdot \mathbb{E}[g] \leq g(x) \leq 1.1 \cdot \mathbb{E}[g]\}$

We know that $Pr[X \in W] \geq 0.99$. If possible, let the assertion of the lemma be false. This implies that $Pr[X \in R_1] \geq 0.9$ and $Pr[X \in R_2] \leq 0.1$. Let $Z \subseteq W \cap R_1$ be a subset of R such that $Pr[X \in Z] = 0.89$. Now

$$\mathbb{E}[g] = \sum_{x \in R} Pr[X = x]g(x) = \sum_{x \in Z} Pr[X = x]g(x) + \sum_{x \in R \setminus Z} Pr[X = x]g(x)$$

Substituting the values now, we get

$$\mathbb{E}[g] \geq Pr[X \in Z] \cdot 0.9 \cdot \mathbb{E}[g] + \sum_{x \in R \setminus Z} Pr[X = x]g(x)$$

Simplifying further, we get

$$\sum_{x \in R \setminus Z} Pr[X = x]g(x) \leq \mathbb{E}[g] \cdot (1 - 0.9 \cdot Pr[X \in Z]) \leq 0.2 \cdot \mathbb{E}[g]$$

We will now compute an upper bound on the expected value of f and arrive at a contradiction.

$$\mathbb{E}[f] = \sum_{x \in R} Pr[X = x]f(x) = \sum_{x \in Z} Pr[X = x]f(x) + \sum_{x \in R \setminus Z} \Pr[X = x]f(x)$$

Observe that

- $\sum_{x \in Z} Pr[X = x]f(x) \leq 0.01 \cdot \mathbb{E}[g] \cdot Pr[X \in Z] \leq 0.01 \times 0.89 \times \mathbb{E}[g] = 0.0089 \cdot \mathbb{E}[g]$
- $\sum_{x \in R \setminus Z} Pr[X = x]f(x) \leq \sum_{x \in R \setminus Z} Pr[X = x]g(x) \leq 0.2 \cdot \mathbb{E}[g]$

  So, we obtain

$$\mathbb{E}[f] \leq 0.3 \cdot \mathbb{E}[g] < 0.5 \cdot \mathbb{E}[g]$$

which is a contradiction. □

We will also need the following lemma, which could be thought of as a strengthened inclusion-exclusion proved using sampling.

Lemma 4.11 (Strong Inclusion-Exclusion)**.** *Let $W_1, W_2, W_3, \ldots, W_l$ be subsets of a finite set W. For a parameter $\lambda \geq 1$, let the following be true.*

$$\sum_{i,j \in [l], i \neq j} |W_i \cap W_j| \leq \lambda \sum_{i \in [l]} |W_i|$$

Then, $\left| \bigcup_{i \in [l]} W_i \right| \geq \frac{1}{4\lambda} \sum_{i \in [l]} |W_i|$.

Proof. Let $\lambda' > \lambda$ be any constant. For each $i \in [l]$, we construct the set $\tilde{W}_i$ by picking every element of W_i independently with probability $\frac{1}{\lambda'}$. By linearity of expectations, $\mathbb{E}(|\tilde{W}_i|) = \frac{1}{\lambda'}|W_i|$. Similarly, for any $i \neq j$, $\mathbb{E}(|\tilde{W}_i \cap \tilde{W}_j|) = \frac{1}{\lambda'^2}|W_i \cap W_j|$. By the principle of inclusion-exclusion, $|\cup_{i \in [l]} \tilde{W}_i| \geq \sum_{i \in [l]} |\tilde{W}_i| - \sum_{i,j \in [l], i \neq j} |\tilde{W}_i \cap \tilde{W}_j|$. By the linearity of expectations, $\mathbb{E}(|\cup_{i \in [l]} \tilde{W}_i|) \geq \sum_{i \in [l]} \mathbb{E}(|\tilde{W}_i|) - \sum_{i,j \in [l], i \neq j} \mathbb{E}(|\tilde{W}_i \cap \tilde{W}_j|)$, which is at least $(1/\lambda' - \lambda/\lambda'^2) \sum_{i \in [l]} |W_i|$. Hence, there is some choice of random bits, such that the size of $\cup_{i \in [l]} \tilde{W}_i$ is at least $(1/\lambda' - \lambda/\lambda'^2) \sum_{i \in [l]} |W_i|$. Now, taking $\lambda' = 2\lambda$ completes the proof. □

4.4 Upper bound on the complexity of homogeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuits

In this section, we state and prove the upper bound on the complexity of a $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuit. A very similar bound was proved by Kayal et al in [KLSS14a]. We include a proof for completeness.

Lemma 4.12. *Let C be a depth-4 homogeneous circuit computing a polynomial of degree u in N variables such that the support of the bottom product gates in C is at most s. Let $\mathcal{M}$ be a set of monomials of degree equal to r and let m be a positive integer. Then,*

$$\Phi_{\mathcal{M},m}(C) \leq (rs+1) \cdot \mathit{Size}(C) \binom{\lceil \frac{2u}{s} \rceil + r}{r} \binom{N}{m+rs}$$

for any choice of m, r, s, N satisfying $m + rs \leq N/2$.

Proof. Let us consider a product gate $Q = \prod_{i=1}^{l} P_i$ in C. Without loss of generality, we can assume that there is at most one i such that degree of P_i is less than $\frac{s}{2}$. Otherwise, we could multiply two such low degree P_i and increase the degree of the polynomials. Observe that if the support of the bottom product gates in C was at most s to start with, this operation preserves that property, since we are only multiplying two polynomials if there degree is at most $\frac{s}{2}$.Therefore, $l \leq \lceil \frac{2u}{s} \rceil$.

Now, let α be a monomial of degree r. The derivative of Q with respect to α is a sum, where each summand is of the form $\partial_\alpha(\prod_{i \in S} P_i) \cdot \prod_{j \in [l] \setminus S} P_j$ where S is a subset of $[l]$ of size at most r.

We will now focus on one such summand. When this derivative is shifted by a multilinear monomial γ of degree m, we get a polynomial of the form $\gamma \cdot \partial_\alpha(\prod_{i \in S} P_i) \cdot \prod_{j \in [l] \setminus S} P_j$. Let us focus our attention on monomials in $\gamma \cdot \partial_\alpha(\prod_{i \in S} P_i)$. Every monomial here has support at least m and most $m+rs$ since γ has support m, each P_i has support at most s and $|S| \leq r$. This implies that the polynomial $\gamma \cdot \partial_\alpha(\prod_{i \in S} P_i) \cdot \prod_{j \in [l] \setminus S} P_j$ is in the linear span of the polynomials $\{\beta \cdot \prod_{j \in [l] \setminus S} P_j : m \leq \mathrm{Supp}(\beta) \leq m + rs\}$. Moreover, even after taking the multilinear projections, it is true that the polynomial $\sigma(\gamma \cdot \partial_\alpha(\prod_{i \in S} P_i) \cdot \prod_{j \in [l] \setminus S} P_j)$ is in the linear span of the polynomials $\{\sigma(\beta \cdot \prod_{j \in [l] \setminus S} P_j) :$

$m \leq \text{Supp}(\beta) \leq m + rs\}$. Note that the set of polynomials $\{\sigma(\beta \cdot \prod_{j \in [l] \setminus S} P_j) : m \leq \text{Supp}(\beta) \leq m + rs\}$ does not depend upon α. In particular, for all α of degree r, it is true that $\sigma(\gamma \cdot \partial_\alpha(\prod_{i \in S} P_i) \cdot \prod_{j \in [l] \setminus S} P_j)$ is in the linear span of the polynomials $\{\sigma(\beta \cdot \prod_{j \in [l] \setminus S} P_j) : m \leq \text{Supp}(\beta) \leq m + rs\}$. Observe that any polynomial of the form $\beta \cdot \prod_{j \in [l] \setminus S} P_j$ will be set to zero under multilinear projections if β is not multilinear. So, $\sigma(\gamma \cdot \partial_\alpha(\prod_{i \in S} P_i) \cdot \prod_{j \in [l] \setminus S} P_j)$ is in fact in the linear span of the polynomials $\{\sigma(\beta \cdot \prod_{j \in [l] \setminus S} P_j) : m \leq \text{degree}(\beta) = \text{Supp}(\beta) \leq m + rs\}$. The dimension of the space $\{\sigma(\beta \cdot \prod_{j \in [l] \setminus S} P_j) : m \leq \text{degree}(\beta) = \text{Supp}(\beta) \leq m + rs\}$ is at most the number of multilinear monomials β of degree between m and $m + rs$. This is at most $\sum_{i=0}^{rs} \binom{N}{m+i}$, which is at most $(rs+1) \cdot \binom{N}{m+rs}$ since $m + rs \leq \frac{N}{2}$ and so the terms in the summation increase with an increase in i.

From the above discussion, we can conclude that for a fixed subset S of $[l]$ of size at most r, the multilinear projections of the shifts of $\partial_\alpha(\prod_{i \in S} P_i) \cdot \prod_{j \in [l] \setminus S} P_j$ lie in a space of dimension at most $(rs+1) \cdot \binom{N}{m+rs}$. From this it follows that the set of projected shifted partial derivatives of order r of Q lie in a linear space of polynomials of dimension at most $(rs+1) \cdot \binom{N}{m+rs} \cdot \binom{\lceil \frac{2u}{s} \rceil + r}{r}$ since there are at most $\binom{\lceil \frac{2u}{s} \rceil + r}{r}$ subsets of $[l]$ of size at most r.

The bound on the complexity of the circuit now just follows from sub-additivity of the complexity measure.

□

4.5 Strategy for proving a lower bound on the complexity of $NW_{n,D}$ and $IMM_{\tilde{n},n}$

To show a lower bound on the complexity of the polynomial P(which will be $IMM_{\tilde{n},n}$ or $NW_{n,D}$ in this chapter), we choose an appropriate set of monomials $\mathcal{M}$ and a parameter m and then obtain a lower bound on the value of $\Phi_{\mathcal{M},m}(P)$. When $\mathcal{M}$ and m are clear from the context, we use $\Phi_{\mathcal{M},m}(P)$ and $\Phi(P)$ interchangeably. We will now try to gain a more concrete understanding of the space of polynomials, whose dimension we want to lower bound. We will need some notations first.

We denote by $M(\alpha)$ the set of monomials $\text{Supp}(\partial_\alpha(P))$. We will use the two interchangeably. For any monomial $\alpha \in \mathcal{M}$ and any monomial $\beta \in \text{Supp}(\partial_\alpha(P))$, define the set

$$S_m^P(\alpha, \beta) = \{\gamma : \deg(\gamma) = \text{Supp}(\gamma) = m \text{ and } \text{Supp}(\gamma) \cap \text{Supp}(\beta) = \phi\}$$

to be the set of all multilinear monomials of degree m which are disjoint from β. We define the set $\tilde{S}_m^P(\alpha, \beta)$ to be the subset of multilinear monomials γ in $S_m^P(\alpha, \beta)$ such that $\beta \cdot \gamma$ is the leading monomial of $\sigma(\gamma \cdot \partial_\alpha(P))$. Define

$$A_m^P(\alpha, \beta) = \{\gamma \cdot \beta : \gamma \in \tilde{S}_m^P(\alpha, \beta)\}$$

When the polynomial P is clear from the context, we drop the P from $A_m^P(\alpha, \beta)$, $S_m^P(\alpha, \beta)$ and $\tilde{S}_m^P(\alpha, \beta)$ and instead denote them by $A_m(\alpha, \beta)$, $S_m(\alpha, \beta)$ and $\tilde{S}_m(\alpha, \beta)$ respectively.

The following lemma relates the size of the union of the sets $A_m(\alpha, \beta)$ to $\Phi_{\mathcal{M},m}(P)$

Lemma 4.13. *Let P be a polynomial in N variables and let $\mathcal{M}$ be any set of monomials on these variables. Let $m \leq N$ be a positive integer and let $\Phi_{\mathcal{M},m}(P)$ and $A_m(\alpha, \beta)$ be as defined. Then,*

$$\Phi_{\mathcal{M},m}(P) \geq \left| \bigcup_{\substack{\alpha \in \mathcal{M} \\ \beta \in Supp(\partial_\alpha(P))}} A_m(\alpha, \beta) \right|$$

Proof. To prove the lemma, it suffices to show that for $\alpha \in \mathcal{M}$ and $\beta \in \text{Supp}(\partial_\alpha(P))$, $A_m(\alpha, \beta)$ are a subset of leading monomials of polynomials in

$$\mathbb{F} - \text{span}\left\{\sigma(\gamma \cdot \partial_{\mathcal{M}}(P)) : \text{Supp}(\gamma) = \deg(\gamma) = m\right\}.$$

This fact just follows from the definition of $A_m(\alpha, \beta)$. The lemma then follows from the fact that for any linear space of polynomials, its dimension is at least the number of distinct leading monomials in the space. □

By the principle of inclusion-exclusion, we get the following corollary.

Corollary 4.14. *Let P be a polynomial in N variables and let $\mathcal{M}$ be any set of monomials on these variables. Let $m \leq N$ be a positive integer and let $\Phi_{\mathcal{M},m}(P)$ and $A_m(\alpha, \beta)$ be as defined. Then,*

$$\Phi_{\mathcal{M},m}(P) \geq \sum_{\substack{\alpha \in \mathcal{M} \\ \beta \in Supp(\partial_\alpha(P))}} |A_m(\alpha, \beta)| - \sum_{\substack{\alpha_1, \alpha_2 \in \mathcal{M} \\ \beta_1 \in Supp(\partial_{\alpha_1}(P)) \\ \beta_2 \in Supp(\partial_{\alpha_2}(P)) \\ (\alpha_1, \beta_1) \neq (\alpha_2, \beta_2)}} |A_m(\alpha_1, \beta_1) \cap A_m(\alpha_2, \beta_2)|$$

Therefore, to get a lower bound on $\Phi_{\mathcal{M},m}(P)$, we show that $\sum_{\alpha \in \mathcal{M}, \beta \in \partial_\alpha(P)} |A_m(\alpha, \beta)|$ is large and the second term in the expression above is small. The following lemma relates $\sum_{\beta \in \partial_\alpha(P)} |A_m(\alpha, \beta)|$ to the size of the sets $S_m(\alpha, \beta)$, which, in principle are somewhat simpler objects to describe.

Lemma 4.15. *Let P be a polynomial in N variables and let $\alpha \in \mathcal{M}$ be a monomial on these variables. Let $S_m(\alpha, \beta)$ and $A_m(\alpha, \beta)$ be sets as defined. Then,*

$$\sum_{\beta \in Supp(\partial_\alpha(P))} |A_m(\alpha, \beta)| \geq \left| \bigcup_{\beta \in Supp(\partial_\alpha(P))} S_m(\alpha, \beta) \right|$$

Proof. Consider the sets $Z = \{(\beta, \gamma) : \beta \in \text{Supp}(\partial_\alpha(P)), \gamma \in A_m(\alpha, \beta)\}$ and $W = \bigcup_{\beta \in \text{Supp}(\partial_\alpha(P))} S_m(\alpha, \beta)$. To prove the lemma, we show the existence of a one one map from W to Z. Consider any $\gamma \in W$. By definition, this means that there exists a $\beta \in \text{Supp}(\partial_\alpha(P))$, such that $\gamma \in S_m(\alpha, \beta)$. This implies that $\gamma \cdot \beta \in \text{Supp}(\sigma(\gamma \cdot \partial_\alpha(P)))$. In particular, $\sigma(\gamma \cdot \partial_\alpha(P))$ is not the identically zero polynomial. So, there exists a $\beta' \in \text{Supp}(\partial_\alpha(P))$ such that $\gamma \cdot \beta'$ is the leading monomial of $\sigma(\gamma \cdot \partial_\alpha(P))$. From the definitions, this implies that $\gamma \cdot \beta' \in A_m(\alpha, \beta')$. So, we map γ to $(\beta', \gamma \cdot \beta')$. Clearly, this map is one one, since the pre-image of (ρ, ψ) is given by ψ/ρ. Hence, the cardinality of Z is at least the cardinality of W. □

4.5.1 Obtaining the lower bound on $\Phi_{\mathcal{M},m}(P)$

For a polynomial P, a set of monomials $\mathcal{M}$ and a positive integer m, we now outline the general sequence of arguments which we use to lower bound $\Phi_{\mathcal{M},m}(P)$. The exact sequence of arguments used in the proofs vary slightly for $NW_{n,D}$ and $IMM_{\tilde{n},n}$. To express this outline more concretely, we will need some notations. For a polynomial P

and monomials $\alpha, \alpha' \in \mathcal{M}$, we define

$$T_1(\alpha, P) = \sum_{\beta \in \text{Supp}(\partial_\alpha(P))} |S_m(\alpha, \beta)|$$

$$T_2(\alpha, P) = \sum_{\substack{\beta_1, \beta_2 \in \text{Supp}(\partial_\alpha(P)) \\ \beta_1 \neq \beta_2}} |S_m(\alpha, \beta_1) \cap S_m(\alpha, \beta_2)|$$

and

$$T_3(\alpha, \alpha', P) = \sum_{\substack{\beta_1 \in \text{Supp}(\partial_\alpha(P)) \\ \beta_2 \in \text{Supp}(\partial_{\alpha'}(P)) \\ (\alpha, \beta_1) \neq (\alpha', \beta_2)}} |A_m(\alpha, \beta_1) \cap A_m(\alpha', \beta_2)|$$

We also define

$$T_1(P) = \sum_{\alpha \in \mathcal{M}} T_1(\alpha, P)$$

$$T_2(P) = \sum_{\alpha \in \mathcal{M}} T_2(\alpha, P)$$

and

$$T_3(P) = \sum_{\alpha, \alpha' \in \mathcal{M}} T_3(\alpha, \alpha', P)$$

At places where P is clear from the context, we drop the P in $T_1(\alpha, P), T_2(\alpha, P)$ and $T_3(\alpha, \alpha', P)$ and denote them by $T_1(\alpha), T_2(\alpha)$ and $T_3(\alpha, \alpha')$ respectively.

From the Corollary 4.14 and Lemma 4.15, it follows that for any polynomial P, set of monomials $\mathcal{M}$ and a parameter m,

$$\Phi_{\mathcal{M},m}(P) \geq T_1(P) - T_2(P) - T_3(P)$$

Outline for Nisan-Wigderson polynomials In the proof of the lower bound for the $NW_{n,D}$ polynomial, we observe that over the random restrictions of $NW_{n,D}$, the expected value of $T_1 - T_2 - T_3$ is almost as large as the expected value of T_1. We will then use Lemma 4.10 to argue that with a sufficiently high probability, the complexity of a random restriction of $NW_{n,D}$ is high.

Outline for Iterated Matrix Multiplication For iterated matrix multiplication, it turns out that the expected value of T_2 and T_3 are in fact larger than the expected value of T_1. So, we first use tail inequalities to argue that for a random restriction P of $IMM_{\tilde{n},n}$, with a high probability all of T_1, T_2, T_3 take values close to their expected

values. We pick such a restriction P. Since the value of $T_2(P) + T_3(P)$ is larger than $T_1(P)$, $T_1(P) - T_2(P) - T_3(P)$ does not give us a meaningful lower bound on $\Phi_{\mathcal{M},m}(P)$.

To get around this problem, we take the help of Lemma 4.11, which can be seen as an strengthened form of the principle of Inclusion-Exclusion. We first show that for such a restriction P , there is a large subset $\mathcal{G} \subseteq \mathcal{M}$ of monomials such that

1. For each α in $\mathcal{G}$, $T_1(\alpha)$ is large.
2. For each α in $\mathcal{G}$, $T_2(\alpha)$ is not too large compared to $T_1(\alpha)$.
3. $\sum_{\alpha_1,\alpha_2 \in \mathcal{G}} T_3(\alpha_1, \alpha_2)$ is not too large when compared to $\sum_{\alpha \in \mathcal{G}, \beta \in \mathrm{Supp}(\partial_\alpha(P))} |A_m(\alpha, \beta)|$.

We now argue that by multiple invocations of Lemma 4.11, this suffices to show that the complexity of P is large.

- For each $\alpha \in \mathcal{G}$, since $T_1(\alpha)$ is large, it follows that $\sum_{\beta \in \mathrm{Supp}(\partial_\alpha(P))} |S_m(\alpha, \beta)|$ is large.
- For each $\alpha \in \mathcal{G}$, since $T_2(\alpha)$ is not much larger than $T_1(\alpha)$, Lemma 4.11 and Lemma 4.15 imply that for each $\alpha \in \mathcal{G}$, $\sum_{\beta \in \mathrm{Supp}(\partial_\alpha(P))} |A_m(\alpha, \beta)|$ is large.
- We also know that $\sum_{\alpha_1,\alpha_2 \in \mathcal{G}} T_3(\alpha_1, \alpha_2) = \sum_{\substack{\alpha_1,\alpha_2 \in \mathcal{G} \\ \beta_1 \in \mathrm{Supp}(\partial_{\alpha_1}(P)) \\ \beta_2 \in \mathrm{Supp}(\partial_{\alpha_2}(P)) \\ (\alpha_1,\beta_1) \neq (\alpha_2,\beta_2)}} |A_m(\alpha_1, \beta_1) \cap A_m(\alpha_2, \beta_2)|$ is not much larger than $\sum_{\alpha \in \mathcal{G}, \beta \in \mathrm{Supp}(\partial_\alpha(P))} |A_m(\alpha, \beta)|$.
- Lemma 4.11 will then imply that $\left| \bigcup_{\substack{\alpha \in \mathcal{G} \\ \beta \in \mathrm{Supp}(\partial_\alpha(P))}} A_m(\alpha, \beta) \right|$ is large. Hence, by Lemma 4.13, $\Phi_{\mathcal{G},m}(P)$ is large.

4.6 Lower bound for $NW_{n,D}$

In this section, we prove lower bound on the size of homogeneous $\Sigma\Pi\Sigma\Pi$ circuits which compute the $NW_{n,D}$ polynomial.

4.6.1 Random restrictions and proof outline

From the definition, it follows that the total number of variables N in $NW_{n,D}$ is $N = n^3$. Let the set of all these variables be $\mathcal{V}$. We will now define our random restriction

procedure by defining a distribution $\mathcal{D}$ over subsets $V \subset \mathcal{V}$. The random restriction procedure will sample $V \leftarrow \mathcal{D}$ and then keep only those variables "alive" that come from V and set the rest to zero. The restriction of the set of variables induces a restriction on any polynomial of these variables. We will use the notation $NW_{n,D}|_V$ for the restriction of $NW_{n,D}$ obtained by setting every variable outside V to 0. Therefore, any distribution $\mathcal{D}$ also induces a distribution on the set of restrictions of $NW_{n,D}$. Similarly, the distribution $\mathcal{D}$ also induces a distribution over the restrictions of any circuit computing a polynomial over $\mathcal{V}$. We will use the notation $C|_V$ for the restriction of a circuit C obtained by setting every input gate in C which is labelled by a variable outside V to 0.

The distribution: Each variable in $\mathcal{V}$ is independently kept alive with a probability $p = n^{-\varepsilon}$, where ε is an absolute constant such that $0 < \varepsilon \leq 0.01$. This gives a distribution over the subsets of $\mathcal{V}$. We call it $\mathcal{D}$.

Steps in the proof: The proof consists of three main steps.

- We consider a depth-4 homogeneous circuit C computing the polynomial $NW_{n,D}$. If C was *large* to start with, we have nothing to prove. Else, C was *small*. We then analyze the behavior of C under random restrictions as defined above.

- We show that with high probability, none of the product gates in the bottom level of C which has support at least $s = \sqrt{n}$ survives the random restriction procedure if the original circuit had size $2^{O(\sqrt{n}\log n)}$. So, we are left with a low support circuit computing a restriction of $NW_{n,D}$.

- We then argue that with good probability, a random restriction of $NW_{n,D}$ has high projected shifted partials complexity.

- Finally, we show that both the events above together happen with some non zero probability. Then, comparing the complexity of the restriction of $NW_{n,D}$ and the restricted circuit, gives us the lower bound.

4.6.2 Choice of parameters

We enumerate the values of the parameters used in this proof below.

1. n. (This is the degree of the polynomial $NW_{n,D}$)
2. $N = n^3$. (This is the total number of variables)
3. $r = \frac{1.1\sqrt{n}}{5}$. (This is the order of the derivatives involved)
4. $s = \sqrt{n}$. (This indicates the support of a product gate in the circuit after random restrictions)
5. $m = \frac{N}{2}(1 - \frac{\ln n}{5\sqrt{n}})$. (This is the degree of the multilinear shifts)
6. ε is any absolute constant such that $0 < \varepsilon < 0.01$.
7. $p = n^{-\varepsilon}$. (This is the probability with which each variable is kept alive independently)
8. $k = n - r$. (This is the size of the support of the monomials in any r^{th} order derivative of $NW_{n,D}$)
9. $d = \Theta\left(\frac{n}{\log n}\right)$ is a parameter chosen such that $1/4 \cdot n^{-4}\frac{\binom{N-k}{m}}{\binom{N-2k}{m-k}} \leq n^{2d} \leq 1/4 \cdot n^{-2}\frac{\binom{N-k}{m}}{\binom{N-2k}{m-k}}$.
10. $D = \frac{\varepsilon n}{2} + d$. (This is the parameter D in $NW_{n,D}$)
11. $\mathcal{D}$. (This is the distribution on the subsets of $\mathcal{V}$ obtained by keeping each variable in $\mathcal{V}$ alive independently with a probability $p = n^{-\varepsilon}$)

In the rest of this chapter, we always invoke the definition of the Nisan-Wigderson polynomials for $D = \frac{\varepsilon n}{2} + d$. So, for the rest of the proof, we use the notation NW for $NW_{n,D}$.

4.6.3 Effect of random restrictions on the circuit

The following lemma gives us an upper bound on the complexity of *small* circuits under the random restrictions.

Lemma 4.16. *Let $s = \sqrt{n}, r = \frac{1.1\sqrt{n}}{5}$ and let m be a parameter such that $m+rs \leq N/2$ and let $\varepsilon > 0$ be a constant. Let $\mathcal{M}$ be any set of monomials of degree equal to r. Let C be a homogeneous depth-4 circuit of size at most $2^{\frac{\varepsilon}{2}\sqrt{n}\log n}$ computing the polynomial NW. Then, with probability at least $1 - o(1)$ over $V \leftarrow \mathcal{D}$*

$$\Phi_{\mathcal{M},m}(C|_V) \leq O(n) \cdot \mathit{Size}(C) \binom{\lceil \frac{2n}{s} \rceil + r}{r} \binom{N}{m+rs}$$

Proof. When the variables are kept alive with probability $n^{-\varepsilon}$ independently, then the probability that a bottom product gate with support at least $\sqrt{n}$ survives equals $n^{-\varepsilon\sqrt{n}}$. Therefore, the probability that some gate with support at least $s = \sqrt{n}$ survives in $C|_V$ is at most $\text{Size}(C)/n^{\varepsilon\sqrt{n}}$. Substituting the value of size of C, we see that this is at most $n^{-\frac{\varepsilon}{2}\sqrt{n}}$ which is $o(1)$.

Now, by Lemma 4.12, the complexity of the restricted circuit is at most $O(n) \cdot \text{Size}(C) \cdot \binom{\lceil \frac{2n}{s} \rceil + r}{r} \cdot \binom{N}{m+rs}$, with probability at least $1 - o(1)$. □

Observe that we have just argued that if the circuit was of size at most $2^{\frac{\varepsilon}{2}\sqrt{n}\log n}$, then with probability at least $1 - o(1)$, at the end of the random restriction process, none of the product gates with support larger than $s = \sqrt{n}$ at the bottom level is alive. Otherwise, the size of the circuit was larger than $2^{\frac{\varepsilon}{2}\sqrt{n}\log n}$ to start with, in which case, we have nothing to prove.

4.6.4 Effect of random restrictions on $NW_{n,D}$

In this section, we show that with a reasonably high probability, a random restriction of NW has a large complexity. We outline the plan and set some notations below.

Plan of the proof: We will show that for $V \leftarrow \mathcal{D}$ expected value of the expression $T_1|_V - T_2|_V - T_3|_V$ is large and then use this to obtain a lower bound on the complexity of a random restriction of NW. We will do this by proving a lower bound on the expected value of $T_1|_V$ and upper bounds on the expected values of $T_2|_V$ and $T_3|_V$. At this point, we would like to argue that the complexity remains close to the expectation with a reasonably high probability. This observation is proved using Lemma 4.10 and the bound on the variance of the number of monomials alive at the end of random restrictions obtained in [KLSS14a].

Recall that $D = \frac{n.\varepsilon}{2} + d$ for some constant ε and a parameter $d = \Theta(\frac{n}{\log n})$.

Let $\mathcal{M}^{[r]} = \{\prod_{i\in[r]} x_{i,j} : j \in [n^2]\}$ be a set of monomials. Observe that for $r < D$, every monomial in $\mathcal{M}^{[r]}$ has an extension in Supp(NW). This implies that for every $\alpha \in \mathcal{M}^{[r]}$, $\partial_\alpha(NW)$ is non zero. In fact, it is a sum of exactly $n^{2(D-r)}$ monomials. For our partial derivatives, we consider the set of partial derivatives of NW with respect to monomials from $\mathcal{M}^{[r]}$. For brevity, we call this set $\mathcal{M}$ for the rest of the proof.

We will now prove that with a high probability over $V \leftarrow \mathcal{D}$, $\Phi_{\mathcal{M},m}(NW|_V)$ is large. Recall that from the discussion in Section 4.5, it will suffice to show that $\Phi_{\mathcal{M},m}(NW|_V) = T_1(NW|_V) - T_2(NW|_V) - T_3(NW|_V)$ is large with a good probability. To this end, we first show that $\Phi_{\mathcal{M},m}(NW)$ is large in expectation and then argue that with a good probability the complexity measure is not too much less the mean.

Observe that according to our definitions here, the set of monomials $\mathcal{M}$ is fixed and does not depend upon the random restrictions. Also, the contribution of any monomial $\alpha \in \mathcal{M}$ is a random variable. For example, for any $\alpha \in \mathcal{M}$ and $\beta \in M(\alpha)$, if α and β both survive the random restriction procedure, then the contribution of β to $|A_m(\alpha,\beta)|$ is $|S_m(\alpha,\beta)| = \binom{N-k}{m}$ whereas if either of them is set to zero during the random restrictions, then the contribution is 0. Similarly for T_2 and T_3. Taking this into account, we state the definitions of T_1, T_2, T_3 which we use in our expectations calculations below. We need a piece of notation first. For monomials $\alpha_1, \alpha_2, \ldots, \alpha_j$, we define $1_{\alpha_1,\alpha_2,\ldots,\alpha_j}$ to be the event that every monomial in $\{\alpha_1, \alpha_2, \ldots, \alpha_j\}$ survives the random restriction procedure.

- $T_1(NW|_V) = \sum_{\substack{\alpha\in\mathcal{M}^{[r]} \\ \beta\in M(\alpha)}} 1_{\alpha,\beta} \cdot |S_m(\alpha,\beta)|$
- $T_2(NW|_V) = \sum_{\substack{\alpha\in\mathcal{M}^{[r]} \\ \beta,\gamma\in M(\alpha) \\ \beta\neq\gamma}} 1_{\alpha,\beta,\gamma} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|$
- $T_3(NW|_V) = \sum_{\substack{\alpha_1,\alpha_2\in\mathcal{M}^{[r]} \\ \beta_1\in M(\alpha_1) \\ \beta_2\in M(\alpha_2) \\ (\alpha_1,\beta_1)\neq(\alpha_2,\beta_2)}} 1_{\alpha_1,\alpha_2,\beta_1,\beta_2} \cdot |A_m(\alpha_1,\beta_1) \cap A_m(\alpha_2,\beta_2)|$

For the ease of notations, for the rest of the proof of lower bound for NW, we denote $T_1(NW|_V)$ by $T_1|_V$. Similarly, we use $T_2|_V$ for $T_2(NW|_V)$ and $T_3|_V$ for $T_3(NW|_V)$. We

know that for any restriction $NW|_V$,

$$\Phi_{\mathcal{M},m}(NW|_V) \geq T_1|_V - T_2|_V - T_3|_V \tag{4.17}$$

Therefore, by the linearity of expectation is, the expected complexity of a random restriction of NW,

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[\Phi_{\mathcal{M},m}(NW|_V)] \geq \mathbb{E}_{V\leftarrow\mathcal{D}}[T_1|_V] - \mathbb{E}_{V\leftarrow\mathcal{D}}[T_2|_V] - \mathbb{E}_{V\leftarrow\mathcal{D}}[T_3|_V] \tag{4.18}$$

We will now bound the expected values of $T_1|_V$, $T_2|_V$, $T_3|_V$ under random restrictions. More precisely, we prove the following.

Lemma 4.19.

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[T_1|_V] = \binom{N-k}{m} \cdot n^{2d}$$

Lemma 4.20.

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[T_2|_V] \leq n^{4d-2r+\varepsilon r+1} \cdot \binom{N-2k}{m}$$

Lemma 4.21.

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[T_3|_V] \leq n^{4d+2} \cdot \binom{N-2k}{m-k}$$

We will now use the bounds given by the lemmas above to complete the proof of the lower bound. We will prove the above lemmas in Section 4.7.

4.6.5 Lower bound on the complexity of $NW_{n,D}$

Lemma 4.22. *For any choice of parameters $m, r, d, \varepsilon, n, N, k$ such that*

- $n^{2d-2r+\varepsilon r+1} \leq 1/4 \cdot \frac{\binom{N-k}{m}}{\binom{N-2k}{m}}$
- $n^{2d+2} \leq 1/4 \cdot \frac{\binom{N-k}{m}}{\binom{N-2k}{m-k}}$

the following is true

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[\Phi_{\mathcal{M},m}(NW|_V)] \geq 0.5 \cdot \mathbb{E}_{V\leftarrow\mathcal{D}}[T_1|_V]$$

Proof. From the choice of parameters and Lemma 4.19, Lemma 4.20 and Lemma 4.21, it easily follows that $\mathbb{E}_{V\leftarrow\mathcal{D}}[T_1|_V] \geq 4 \cdot \mathbb{E}_{V\leftarrow\mathcal{D}}[T_2|_V]$ and $\mathbb{E}_{V\leftarrow\mathcal{D}}[T_1|_V] \geq 4 \cdot \mathbb{E}_{V\leftarrow\mathcal{D}}[T_3|_V]$. Thus

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[\Phi_{\mathcal{M},m}(NW|_V)] \geq 0.5 \cdot \mathbb{E}_{V\leftarrow\mathcal{D}}[T_1].$$

□

Thus for the above choice of parameters, we get a lower bound on the expected value of $\Phi_{\mathcal{M},m}(NW|_V)$. We would like to conclude that with a decent (≥ 0.1) probability, the complexity is large. Observe that we cannot directly use Markov's inequality. However we are still able to prove such a statement (see Lemma 4.27). We make the following crucial observation.

Lemma 4.23. *For any $V \subseteq \mathcal{V}$,*

$$\Phi_{\mathcal{M},m}(NW|_V) \leq |Supp(NW|_V)|\binom{N-k}{m}$$

.

Proof. To prove the lemma, we prove an upper bound on the size of the set

$$\bigcup_{\alpha \in \mathcal{M}^{[r]}} \text{Supp}(\partial_\alpha(NW|_V))$$

in the following claim.

Claim 4.24. *For any $V \subseteq \mathcal{V}$, the following is true.*

$$\left| \bigcup_{\alpha \in \mathcal{M}^{[r]}} Supp(\partial_\alpha(NW|_V)) \right| \leq |Supp(NW|_V)|$$

Proof. To prove this claim, we argue that there is a one-one map from the set $\bigcup_{\alpha \in \mathcal{M}^{[r]}} \text{Supp}(\partial_\alpha(NW|_V))$ to the set $\text{Supp}(NW|_V)$. From the definition of $\mathcal{M}^{[r]}$, it follows that all the monomials in $\mathcal{M}^{[r]}$ are of degree r and contain exactly one variable from the set $\{x_{i,j} : j \in [n^2]\}$ for each $i \in [r]$. Also, from the definition of NW, it follows that for every monomial β in $\text{Supp}(NW|_V)$, there is exactly one monomial $\alpha \in \mathcal{M}^{[r]}$ such that β is an extension of α. Or, in other words, for each $\beta \in \text{Supp}(NW|_V)$, there is exactly one $\alpha \in \mathcal{M}^{[r]}$ such that $\partial_\alpha(\beta) \in \text{Supp}(\partial_\alpha(NW|_V))$. Therefore, the function which maps $\partial_\alpha(\beta)$ to β is a one-one map. □

Now, observe that for any monomial γ in the support of any polynomial in the set

$$\{\sigma(\prod_{i \in S} x_i \cdot g) : g \in \partial_{\mathcal{M}^{[r]}}(NW|_V), S \subseteq [N], |S| = m\}$$

there exists an $\alpha \in \mathcal{M}^{[r]}$, a monomial $\beta \in \text{Supp}(NW|_V)$ and a multilinear monomial ρ of degree m such that the supports of $\partial_\alpha(\beta)$ and ρ are disjoint and $\gamma = \partial_\alpha(\beta) \cdot \rho$. For any such β, the number of ρ, which are multilinear of degree m and disjoint from $\partial_\alpha(\beta)$ is equal to $\binom{N-k}{m}$, since $\partial_\alpha(\beta)$ is a multilinear monomial of degree equal to k. Therefore, the number of distinct monomials in the union of supports of all polynomials in $\{\sigma(\prod_{i\in S} x_i \cdot g) : g \in \partial_{\mathcal{M}^{[r]}}(NW|_V), S \subseteq [N], |S| = m\}$ is at most the product of $|\bigcup_{\alpha\in\mathcal{M}^{[r]}} \text{Supp}(\partial_\alpha(NW|_V))|$ and $\binom{N-k}{m}$. The lemma follows from the claim above. $\square$

We will now use Lemma 4.10 to argue that with a decent probablity, a random restriction of NW has a complexity very close to its expected value. For a restriction $P = NW|_V$ of NW, define $g(P) = |\text{Supp}(P)| \cdot \binom{N-k}{m}$ and define $f(P) = \Phi_{\mathcal{M}^{[r]},m}(P)$. Lemma 4.23 implies that for every restriction $P = NW|_V$ of NW, $f(P) \leq g(P)$. Lemma 4.22 implies that $\mathbb{E}_{V\leftarrow\mathcal{D}}[f] \geq 1/2 \cdot \mathbb{E}_{V\leftarrow\mathcal{D}}[g]$. The following lemma of Kayal et al [KLSS14a] tells us that g takes values very close to its expected value with a high probability.

Lemma 4.25 ([KLSS14a])**.** $Pr_{V\leftarrow\mathcal{D}}[|g(NW|_V) - \mathbb{E}_{V'\leftarrow\mathcal{D}}[g]| \geq 0.1 \cdot \mathbb{E}_{V'\leftarrow\mathcal{D}}[g]] \leq 0.01$.

The functions f and g now satisfy the hypothesis of Lemma 4.10. Therefore, we get the following lemma.

Lemma 4.26. $Pr_{V\leftarrow\mathcal{D}}[f(NW|_V) \geq 0.01 \cdot \mathbb{E}_{V'\leftarrow\mathcal{D}}[g]] \geq 0.1$.

Therefore, the following lemma is true.

Lemma 4.27. *For any choice of parameters $m, r, d, \varepsilon, n, N, k$ such that*

- $n^{2d-2r+\varepsilon r+1} \leq 1/4 \cdot \frac{\binom{N-k}{m}}{\binom{N-2k}{m}}$
- $n^{2d+2} \leq 1/4 \cdot \frac{\binom{N-k}{m}}{\binom{N-2k}{m-k}}$

the following is true

$$Pr_{V\leftarrow\mathcal{D}}[\Phi_{\mathcal{M},m}(NW|_V) \geq 0.005 \cdot n^{2d}\binom{N-k}{m}] \geq 0.1$$

4.6.6 Wrapping up the proof

We now complete the proof of the lower bound for the case of NW polynomial which implies Theorem 4.1.

Theorem 4.28. *Let C be any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing $NW_{n,D}$. Then, the size of C is at least $n^{\Omega(\sqrt{n})}$.*

Proof. Recall that, from our choice of parameters, we have $s = \sqrt{n}$, $r = \frac{1.1\sqrt{n}}{5}$, $N = n^3$, $m = \frac{N}{2}(1 - \frac{\ln n}{5\sqrt{n}}) = \frac{N}{2}(1 - \frac{\ln n}{5s})$, d such that $n^{2d} = 1/4 \cdot n^{-2}\frac{\binom{N-k}{m}}{\binom{N-2k}{m-k}}$, $k = n - r$, and $\varepsilon < 0.01$. Observe that $m + rs < \frac{N}{2}$. Let C be a circuit computing the polynomial NW.

If the size of the circuit is at least $n^{\frac{\varepsilon}{2}\sqrt{n}}$, then we are done. Else, the size of C is at most $n^{\frac{\varepsilon}{2}\sqrt{n}}$. Lemma 4.16 implies that with probability at least $1 - o(1)$ the complexity of the circuit is at most $O(n) \cdot \text{Size}(C)\binom{\lceil \frac{2n}{s} \rceil + r}{r}\binom{N}{m+rs}$.

We will first show that for the choice of paramters made above, the hypotheses of Lemma 4.22 hold.

Claim 4.29. *For $m, r, d, \varepsilon, n, N, k$ as chosen above,*

- $n^{2d-2r+\varepsilon r+1} \leq 1/4 \cdot \frac{\binom{N-k}{m}}{\binom{N-2k}{m}}$
- $n^{2d+2} \leq 1/4 \cdot \frac{\binom{N-k}{m}}{\binom{N-2k}{m-k}}$

Proof. By the choice of d, the second constraint is met.

We now need to verify that for the choice of parameters the first constraint is met, i.e.

$$n^{2d-2r+\varepsilon r} \leq 1/4 \cdot n^{-1}\frac{\binom{N-k}{m}}{\binom{N-2k}{m}}.$$

In other words, we would like to show that

$$n^{2d-2r+\varepsilon r} \cdot 4n \cdot \frac{\binom{N-2k}{m}}{\binom{N-k}{m}} \leq 1.$$

Now,

$$\begin{aligned}
n^{2d-2r+\varepsilon r}\cdot 4n\cdot\frac{\binom{N-2k}{m}}{\binom{N-k}{m}} &= n^{-2r+\varepsilon r}\cdot\frac{1}{n}\cdot\frac{\binom{N-2k}{m}}{\binom{N-2k}{m-k}} && \text{substituting value of } n^{2d}\\
&= n^{-2r+\varepsilon r}\cdot\frac{1}{n}\cdot\frac{(N-m-k)!}{(N-m-2k)!}\times\frac{(m-k)!}{m!}\\
&\approx n^{-2r+\varepsilon r}\cdot\frac{1}{n}\cdot\left(\frac{N-m}{m}\right)^k && \text{By Lemma 2.13}\\
&= n^{-2r+\varepsilon r}\cdot\frac{1}{n}\cdot\left(\frac{1+\frac{\ln n}{5s}}{1-\frac{\ln n}{5s}}\right)^k && \text{substituting choice of } m\\
&\leq n^{-2r+\varepsilon r}\cdot\frac{1}{n}\cdot e^{2.01k\frac{\ln n}{5s}} && \text{for large enough } n\\
&= n^{-2r+\varepsilon r}\cdot\frac{1}{n}\cdot n^{2.01k/5s}
\end{aligned}$$

Here, $\approx$ indicated equality up to a polynomial factor in n.

Substituting $r=\frac{1.1\sqrt{n}}{5}, s=\sqrt{n}, k=n-r$ and $\varepsilon<0.01$, it can be verified that the expression above is at most 1. □

Thus by the claim above and Lemma 4.27, we conclude that with

$$\Pr_{V\leftarrow\mathcal{D}}\left[\Phi_{\mathcal{M},m}(NW|_V)=\Omega\left(n^{2d}\binom{N-k}{m}\right)\right]\geq 0.1.$$

So, with probability at least $0.1-o(1)$, the complexity of $C|_V$ is low while at the same time the complexity of the $NW|_V$ remains high. Comparing the bounds, we have

$$\text{Size(C)}=\Omega\left(\frac{n^{2d-1}\binom{N-k}{m}}{\binom{\lceil\frac{2n}{s}\rceil+r}{r}\binom{N}{m+rs}}\right)$$

Putting in $n^{2d}=1/4\cdot n^{-2}\frac{\binom{N-k}{m}}{\binom{N-2k}{m-k}}$, we have

$$\text{Size(C)}=\Omega\left(n^{-3}\cdot\frac{\binom{N-k}{m}\binom{N-k}{m}}{\binom{\lceil\frac{2n}{s}\rceil+r}{r}\binom{N}{m+rs}\binom{N-2k}{m-k}}\right)$$

We will first estimate the ratio of binomial coefficients one by one.

- $\frac{\binom{N-k}{m}}{\binom{N}{m+rs}}=\frac{(N-k)!}{N!}\times\frac{(m+rs)!}{m!}\times\frac{(N-m-rs)!}{(N-m-k)!}\approx\left(\frac{m}{N-m}\right)^{rs}\times\left(\frac{N-m}{N}\right)^k$
- $\frac{\binom{N-k}{m}}{\binom{N-2k}{m-k}}=\frac{(N-k)!}{(N-2k)!}\times\frac{(m-k)!}{m!}\approx\frac{N^k}{m^k}$

- $\binom{\lceil \frac{2n}{s} \rceil + r}{r}$ is $2^{O(r)}$ for our choice of r and s

Plugging these bounds back, we have

$$\text{Size}(C) \geq n^{-3} \cdot \left(\frac{N-m}{m} \right)^{k-rs} \times 2^{-O(r)}$$

Now, we plug in the value of m, which gives us

$$\text{Size}(C) \geq \left(\frac{1 + \frac{\ln n}{5s}}{1 - \frac{\ln n}{5s}} \right)^{k-rs} \times 2^{-O(r)}$$

This gives us

$$\text{Size}(C) \geq \left(1 + \frac{\ln n}{5s} \right)^{k-rs} \times 2^{-O(r)}$$

which implies

$$\text{Size}(C) \geq n^{\frac{k-rs}{5s}} \times 2^{-O(r)}$$

Substituting the values of k, r, s, we get

$$\text{Size}(C) \geq n^{\Omega(\sqrt{n})}$$

□

4.7 Calculations for $NW_{n,D}$

In this sections, we provide the proofs of Lemma 4.19, Lemma 4.20 and Lemma 4.21.

4.7.1 Expected value of $T_1(NW_{n,D}|_V)$

This computation is quite straight forward.

$$\begin{aligned} \mathbb{E}_{V \leftarrow \mathcal{D}}[T_1|_V] &= \sum_{\substack{\alpha \in \mathcal{M}^{[r]} \\ \beta \in M(\alpha)}} \mathbb{E}[1_{\alpha,\beta}] \cdot |S_m(\alpha, \beta)| \\ &= \binom{N-k}{m} \cdot \sum_{\substack{\alpha \in \mathcal{M}^{[r]} \\ \beta \in M(\alpha)}} \mathbb{E}[1_{\alpha,\beta}] \end{aligned}$$

Now observe that $1_{\alpha,\beta} = 1$ when all the variables in the support of the monomial $\alpha\beta$ stay alive. This happens with probability exactly p^n since $\alpha \cdot \beta$ is a multilinear monomial

of degree equal to n. The number of pairs α, β such that $\alpha \in \mathcal{M}^{[r]}$ and $\beta \in M(\alpha)$ is exactly equal to n^{2D}, since $|\mathcal{M}^{[r]}| = n^{2r}$ and for each such α, the number of $\beta \in M(\alpha)$ equals $n^{2(D-r)}$. Plugging this back, we obtain

$$\begin{aligned}\mathbb{E}_{V \leftarrow \mathcal{D}}[T_1|_V] &= \binom{N-k}{m} \cdot n^{2D} p^n \\ &= \binom{N-k}{m} \cdot n^{2d}\end{aligned}$$

4.7.2 Expected value of $T_2(NW_{n,D}|_V)$

By linearity of expectation,

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V] = \sum_{\substack{\alpha \in \mathcal{M}^{[r]} \\ \beta,\gamma \in M(\alpha) \\ \beta \neq \gamma}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha,\beta,\gamma} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|]$$

For any fixed α, β, we partition the set of all $\gamma \in M(\alpha)$ based upon the size of the intersection of the supports of β and γ

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V] = \sum_{0 \leq w \leq D-r} \sum_{\substack{\alpha \in \mathcal{M}^{[r]} \\ \beta \in M(\alpha) \\ \gamma \in M(\alpha) \\ \gamma \neq \beta \\ |\mathrm{Supp}(\gamma) \cap \mathrm{Supp}(\beta)| = w}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha,\beta,\gamma} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|]$$

Observe that we only need to sum up to $w = D - r$ since for any $\beta \neq \gamma \in M(\alpha)$, the maximum size of the intersection of $\mathrm{Supp}(\beta)$ and $\mathrm{Supp}(\gamma)$ can be $D - r$. This is due to the observation that for $\beta \neq \gamma \in M(\alpha)$, there exist distinct univariate polynomials f_β and f_γ of degree at most $D - 1$ in $\mathbb{F}_{n^2}[Z]$ such that $\alpha \cdot \gamma = \prod_{i \in [n]} x_{i,f_\gamma(i)}$ and $\alpha \cdot \beta = \prod_{i \in [n]} x_{i,f_\beta(i)}$. Rearranging the order of summation, we obtain

$$\begin{aligned}\mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V] &= \sum_{\substack{\alpha \in \mathcal{M}^{[r]} \\ \beta \in M(\alpha)}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha,\beta}] \\ &\quad \sum_{0 \leq w \leq D-r} \sum_{\substack{\gamma \in M(\alpha) \\ \gamma \neq \beta \\ |\mathrm{Supp}(\gamma) \cap \mathrm{Supp}(\beta)| = w}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\gamma|\beta} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|]\end{aligned}$$

where $1_{\gamma|\beta}$ is the event $1_{\gamma'}$ where $\gamma' = \prod_{X \in \mathrm{Supp}(\gamma) \setminus \mathrm{Supp}(\beta)} X$. Since the support of α is disjoint from the support of β and γ, so the dependence is only between γ and β. In

the claim below, we derive an upper bound on the expression

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\gamma|\beta} \cdot |S_m(\alpha, \gamma) \cap S_m(\alpha, \beta)|]$$

for fixed values of $\alpha \in \mathcal{M}^{[r]}, \beta \in M(\alpha)$ and $0 \leq w \leq D - r$.

Claim 4.30. *Let α, β be monomials such that $\alpha \in \mathcal{M}^{[r]}$ and $\beta \in M(\alpha)$ and w be an integer such that $0 \leq w \leq D - r$. Then*

$$\sum_{\substack{\gamma \in M(\alpha) \\ \gamma \neq \beta \\ |Supp(\gamma) \cap Supp(\beta)| = w}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\gamma|\beta} \cdot |S_m(\alpha, \gamma) \cap S_m(\alpha, \beta)|] \leq \binom{k}{w} \cdot n^{2(D-r-w)} \cdot p^{k-w} \cdot \binom{N - 2k + w}{m}$$

Proof. From the definition of NW, for any $\alpha \in \mathcal{M}^{[r]}$ and $\beta \in M(\alpha)$, $\alpha\beta$ is a monomial in $\text{Supp}(NW)$. Moreover, there is a unique univariate polynomial $f_\beta(Z) \in \mathbb{F}_{n^2}[Z]$ of degree at most $D - 1$ such that $\alpha \cdot \beta = \prod_{i \in [n]} x_{i, f_\beta(i)}$. The summation above is over all $f_\gamma \in \mathbb{F}_{n^2}[Z]$ of degree at most $D - 1$ satisfying

- $\prod_{i \in [r]} x_{i, f_\gamma(i)} = \alpha$
- $|\{i \in [n] \setminus [r] : f_\gamma(i) = f_\beta(i)\}| = w$

The first condition above can also be written as $f_\beta(j) = f_\gamma(j)$ for every $j \in [r]$. Thus, f_β agrees with f_γ over all the elements in set $[r]$ and over w elements of the set $[n] \setminus [r]$. Since any univariate polynomial of degree at most $D - 1$ can be uniquely determined by its evaluations on any D points, there is a one-one map from the set of f_γ satisfying the constraints above to tuples (U_1, U_2) where

- $U_1 \subseteq [n] \setminus [r]$ is the set of w elements in $[n] \setminus [r]$ where f_β and f_γ agree
- U_2 is a set of input, value pairs for some $D - r - w$ points in $[n] \setminus ([r] \cup U_1)$

Therefore, the number of such f_γ is at most $\binom{k}{w} \cdot n^{2(D-r-w)}$. We will now get an upper bound on the value of $\mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\gamma|\beta} \cdot |S_m(\alpha, \gamma) \cap S_m(\alpha, \beta)|]$ for each such γ. Observe that $1_{\gamma|\beta}$ is 1 when all the variables in the set $\text{Supp}(\gamma) \setminus \text{Supp}(\beta)$ are alive. This happens with probability equal to $p^{|\text{Supp}(\gamma) \setminus \text{Supp}(\beta)|} = p^{k-w}$. The quantity $|S_m(\alpha, \gamma) \cap S_m(\alpha, \beta)|$ is the number of multilinear monomials of degree m which are disjoint from both β

and γ (where $|\mathrm{Supp}(\gamma) \setminus \mathrm{Supp}(\beta)| = w$), and hence $|S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)| = \binom{N-2k+w}{m}$ (Recall that we shift with all multilinear monomials of degree m regardless of V). So,

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\gamma|\beta} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|] = p^{k-w} \cdot \binom{N-2k+w}{m}$$

Multiplying this by the bound on the number of terms in the summation completes the proof of the claim. □

We will now upper bound the sum

$$\sum_{0 \le w \le D-r} \sum_{\substack{\gamma \in M(\alpha) \\ \gamma \neq \beta \\ |\mathrm{Supp}(\gamma) \cap \mathrm{Supp}(\beta)| = w}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha,\beta,\gamma} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|]$$

Claim 4.31. *Let α, β be monomials such that $\alpha \in \mathcal{M}^{[r]}$ and $\beta \in M(\alpha)$. Then*

$$\sum_{0 \le w \le D-r} \sum_{\substack{\gamma \in M(\alpha) \\ \gamma \neq \beta \\ |Supp(\gamma) \cap Supp(\beta)| = w}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha,\beta,\gamma} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|] \le n^{2d-2r+\varepsilon r+1} \cdot \binom{N-2k}{m}$$

Proof. Claim 4.30 implies that

$$\sum_{0 \le w \le D-r} \sum_{\substack{\gamma \in M(\alpha) \\ \gamma \neq \beta \\ |\mathrm{Supp}(\gamma) \cap \mathrm{Supp}(\beta)| = w}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha,\beta,\gamma} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|]$$

is at most

$$\sum_{0 \le w \le D-r} \binom{k}{w} \cdot n^{2(D-r-w)} \cdot p^{k-w} \cdot \binom{N-2k+w}{m}$$

Let us set $g(w) = \binom{k}{w} \cdot n^{2(D-r-w)} \cdot p^{k-w} \cdot \binom{N-2k+w}{m}$ and $g'(w) = g(w)/\binom{N-2k}{m}$. By our choice of parameters, $w^2 = O(n^2)$, $k^2 = O(n^2)$ and $N = \Omega(n^2)$. So by Lemma 2.13

$$\frac{\binom{N-2k+w}{m}}{\binom{N-2k}{m}} \approx \left(\frac{N-2k}{N-m-2k}\right)^w$$

We also know from our choice of parameters that $\frac{N-2k}{N-m-2k} = \Theta(1)$. So, $g'(w) = \binom{k}{w} \cdot n^{2(D-r-w)} \cdot p^{k-w} \cdot \Theta(1)^w$. For $p = n^{-\varepsilon}$ and $k = \Theta(n)$, $g'(w) \le k^w \cdot n^{2D-2r-2w} \cdot p^{k-w} \cdot \Theta(1)^w$. In particular, $g'(w)$ is upper bounded by a decreasing function of w and takes the maximum value $n^{2D-2r}p^k$ at $w = 0$. So

$$\sum_{0 \le w \le D-r} \sum_{\substack{\gamma \in M(\alpha) \\ \gamma \neq \beta \\ |\mathrm{Supp}(\gamma) \cap \mathrm{Supp}(\beta)| = w}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha,\beta,\gamma} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|] \le D \cdot n^{2D-2r} \cdot p^k \cdot \binom{N-2k}{m}$$

Now, substituting $D = \frac{\varepsilon n}{2} + d$, $p = n^{-\varepsilon}$ and $k = n - r$, we get

$$\sum_{0 \le w \le D-r} \sum_{\substack{\gamma \in M(\alpha) \\ \gamma \neq \beta \\ |\mathrm{Supp}(\gamma) \cap \mathrm{Supp}(\beta)| = w}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha,\beta,\gamma} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|] \le n^{2d-2r+\varepsilon r+1} \cdot \binom{N-2k}{m}$$

□

Putting this value back into the equality

$$\begin{aligned} \mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V] \quad = \quad & \sum_{\substack{\alpha \in \mathcal{M}^{[r]} \\ \beta \in M(\alpha)}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha,\beta}] \\ & \sum_{0 \le w \le D-r} \sum_{\substack{\gamma \in M(\alpha) \\ \gamma \neq \beta \\ |\mathrm{Supp}(\gamma) \cap \mathrm{Supp}(\beta)| = w}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\gamma|\beta} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|] \end{aligned}$$

we obtain

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V] \le \sum_{\substack{\alpha \in \mathcal{M}^{[r]} \\ \beta \in M(\alpha)}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha,\beta}] \cdot n^{2d-2r+\varepsilon r+1} \cdot \binom{N-2k}{m}$$

Now observe that $1_{\alpha,\beta} = 1$ when all the variables in the support of the monomial $\alpha\beta$ stay alive. This happens with probability exactly p^n since $\alpha \cdot \beta$ is a multilinear monomial of degree equal to n. The number of pairs α, β such that $\alpha \in \mathcal{M}^{[r]}$ and $\beta \in M(\alpha)$ is exactly equal to n^{2D}, since $|\mathcal{M}^{[r]}| = n^{2r}$ and for each such α, the number of $\beta \in M(\alpha)$ equals $n^{2(D-r)}$. So,

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V] \le p^n \cdot n^{2D} \cdot n^{2d-2r+\varepsilon r+1} \cdot \binom{N-2k}{m}$$

Plugging back the values of p and D, we get Lemma 4.20.

4.7.3 Expected values of $T_3(NW_{n,D}|_V)$

We will again proceed as in the above case, but we have to be a little more careful.

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_3|_V] = \sum_{\substack{\alpha_1,\alpha_2 \in \mathcal{M}^{[r]} \\ \beta_1 \in M(\alpha_1) \\ \beta_2 \in M(\alpha_2) \\ (\alpha_1,\beta_1) \neq (\alpha_2,\beta 2)}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha_1,\alpha_2,\beta_1,\beta_2} \cdot |A_m(\alpha_1,\beta_1) \cap A_m(\alpha_2,\beta_2)|]$$

We will again split the sum based upon the number of agreements between α_1, α_2 and the number of agreements between β_1, β_2. We can rewrite $\mathbb{E}_{V \leftarrow \mathcal{D}}[T_3|_V]$ as

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[T_3|_V] = \sum_{\substack{0\le w_1\le r, 0\le w_2\le k\\ w_1+w_2\le D}} \sum_{\substack{\alpha_1,\alpha_2\in\mathcal{M}^{[r]}\\ \beta_1\in M(\alpha_1)\\ \beta_2\in M(\alpha_2)|\mathrm{Supp}(\alpha_1)\cap\mathrm{Supp}(\alpha_2)|=w_1\\ |\mathrm{Supp}(\beta_1)\cap\mathrm{Supp}(\beta 2)|=w_2}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_1,\alpha_2,\beta_1,\beta_2}\cdot|A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|]$$

Observe that we can drop the constraint $(\alpha_1,\beta_1)\neq(\alpha_2,\beta_2)$ since the sum of number of agreements between α_1 and α_2 and between β_1 and β_2 is at most D which is strictly smaller than n. Rearranging the order of summation, we get

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[T_3|_V] = \sum_{\substack{\alpha_1\in\mathcal{M}^{[r]}\\ \beta_1\in M(\alpha)}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_1,\beta_1}] \sum_{\substack{0\le w_1\le r, 0\le w_2\le k\\ w_1+w_2\le D}} \sum_{\substack{\alpha_2\in\mathcal{M}^{[r]}\\ \beta_2\in M(\alpha_2)\\ |\mathrm{Supp}(\alpha_1)\cap\mathrm{Supp}(\alpha_2)|=w_1\\ |\mathrm{Supp}(\beta_1)\cap\mathrm{Supp}(\beta 2)|=w_2}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_2|\alpha_1}\cdot 1_{\beta_2|\beta_1}\cdot|A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|]$$

where $1_{\alpha_2|\alpha_1}$ is the event $1_{\alpha'}$ where $\alpha'=\prod_{X\in\mathrm{Supp}(\alpha_2)\setminus\mathrm{Supp}(\alpha_1)}X$ and similarly for $1_{\beta_2|\beta_1}$. In the claim below, we upper bound the expression

$$\sum_{\substack{\alpha_2\in\mathcal{M}^{[r]}\\ \beta_2\in M(\alpha_2)\\ |\mathrm{Supp}(\alpha_1)\cap\mathrm{Supp}(\alpha_2)|=w_1\\ |\mathrm{Supp}(\beta_1)\cap\mathrm{Supp}(\beta 2)|=w_2}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_2|\alpha_1}\cdot 1_{\beta_2|\beta_1}\cdot|A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|]$$

for any fixed $\alpha_1\in\mathcal{M}^{[r]},\beta_1\in M(\alpha_1),w_1,w_2$.

Claim 4.32. *Let α_1,β_1 be monomials such that $\alpha_1\in\mathcal{M}^{[r]}$ and $\beta_1\in M(\alpha_1)$. Let $0\le w_1\le r$ and $0\le w_2\le k$ be positive integers such that $w_1+w_2\le D$. Then*

$$\sum_{\substack{\alpha_2\in\mathcal{M}^{[r]}\\ \beta_2\in M(\alpha_2)\\ |Supp(\alpha_1)\cap Supp(\alpha_2)|=w_1\\ |Supp(\beta_1)\cap Supp(\beta 2)|=w_2}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_2|\alpha_1}\cdot 1_{\beta_2|\beta_1}\cdot|A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|]$$
$$\le \binom{r}{w_1}\cdot\binom{k}{w_2}\cdot n^{2(D-w_1-w_2)}\cdot p^{k+r-w_1-w_2}\cdot\binom{N-2k+w_2}{m-k+w_2}$$

Proof. Recall that every monomial in NW corresponds to a univariate polynomial $f\in\mathbb{F}_{n^2}[Z]$ of degree at most $D-1$. So, every pair $\alpha_1\in\mathcal{M}^{[r]}$ and $\beta_1\in M(\alpha_1)$ satisfies

$\alpha_1\beta_1 = \prod_{i\in[n]} x_{i,f_1(i)}$ for $f_1 \in \mathbb{F}_{n^2}[Z]$ of degree at most $D-1$. For a fixed $\alpha_1 \in \mathcal{M}^{[r]}$ and $\beta_1 \in M(\alpha)$ and w_1, w_2, the summation above runs over precisely the set of polynomials $f_2 \in \mathbb{F}_{n^2}[Z]$ of degree at most $D-1$ that satisfy the following two properties:

- $|\{i \in [r] : f_1(i) = f_2(i)\}| = w_1$
- $|\{i \in [n] \setminus [r] : f_1(i) = f_2(i)\}| = w_2$

Since every polynomial of degree $D-1$ is uniquely determined by its evaluation at some D points, the number polynomial f_2 satisfying the above properties is at most $\binom{r}{w_1} \cdot \binom{k}{w_2} \cdot n^{2(D-w_1-w_2)}$. This follows from the observation there is an one-one map from the set of polynomials f_2 satisfying the above properties and the set of tuples (U_1, U_2, U_3), where

- $U_1 \subseteq [r]$ is the set of w_1 elements of $[r]$ where f_1 and f_2 agree
- $U_2 \subseteq [n] \setminus [r]$ is the set of w_2 elements of $[n] \setminus [r]$ where f_1 and f_2 agree
- U_3 specifies the evaluation of f_2 on some $D - w_1 - w_2$ elements of $[n] \setminus (U_1 \cup U_2)$.

Thus, the number of summands in the sum equals $\binom{r}{w_1} \cdot \binom{k}{w_2} \cdot n^{2(D-w_1-w_2)}$.

Now observe that for every such fixed $\alpha_1, \alpha_2, \beta_1, \beta_2$, $1_{\alpha_2|\alpha_1}$ is 1 when all the variables in $\mathrm{Supp}(\alpha_2) \setminus \mathrm{Supp}(\alpha_1)$ survive the random restriction procedure and it is zero otherwise. So, $1_{\alpha_2|\alpha_1}$ is 1 with probability $p^{|\mathrm{Supp}(\alpha_2)\setminus\mathrm{Supp}(\alpha_1)|} = p^{r-w_1}$. Similarly, $1_{\beta_2|\beta_1}$ is 1 with probability p^{k-w_2}. Moreover, $1_{\alpha_2|\alpha_1}$ and $1_{\beta_2|\beta_1}$ are independent events. Also, observe that $|A_m(\alpha_1, \beta_1) \cap A_m(\alpha_2, \beta_2)|$ is upper bounded by the number of multilinear monomials γ of degree $m+k$ which are divisible by both β_1 and β_2. This is at most $\binom{N-2k+w_2}{m-(k-w_2)}$. Hence,

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_2|\alpha_1} \cdot 1_{\beta_2|\beta_1} \cdot |A_m(\alpha_1, \beta_1) \cap A_m(\alpha_2, \beta_2)|] \leq p^{r-w_1} \cdot p^{k-w_2} \cdot \binom{N-2k+w_2}{m-(k-w_2)}$$

The bound in the lemma follows by multiplying the above bound with the upper bound on the number of summands in the summation. □

Using the bound in Claim 4.32, we now upper bound the expression

$$\sum_{\substack{0\le w_1\le r, 0\le w_2\le k\\ w_1+w_2\le D}} \sum_{\substack{\alpha_2\in\mathcal{M}^{[r]}\\ \beta_2\in M(\alpha_2)\\ |\mathrm{Supp}(\alpha_1)\cap\mathrm{Supp}(\alpha_2)|=w_1\\ |\mathrm{Supp}(\beta_1)\cap\mathrm{Supp}(\beta 2)|=w_2}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_2|\alpha_1}\cdot 1_{\beta_2|\beta_1}\cdot |A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|]$$

Claim 4.33. *Let α_1, β_1 be monomials such that $\alpha_1 \in \mathcal{M}^{[r]}$ and $\beta_1 \in M(\alpha_1)$. Then*

$$\sum_{\substack{0\le w_1\le r, 0\le w_2\le k\\ w_1+w_2\le D}} \sum_{\substack{\alpha_2\in\mathcal{M}^{[r]}\\ \beta_2\in M(\alpha_2)\\ |Supp(\alpha_1)\cap Supp(\alpha_2)|=w_1\\ |Supp(\beta_1)\cap Supp(\beta 2)|=w_2}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_2|\alpha_1}\cdot 1_{\beta_2|\beta_1}\cdot |A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|] \le n^{2d+2}\cdot\binom{N-2k}{m-k}$$

Proof. From Claim 4.32, it follows that

$$\sum_{\substack{0\le w_1\le r, 0\le w_2\le k\\ w_1+w_2\le D}} \sum_{\substack{\alpha_2\in\mathcal{M}^{[r]}\\ \beta_2\in M(\alpha_2)\\ |\mathrm{Supp}(\alpha_1)\cap\mathrm{Supp}(\alpha_2)|=w_1\\ |\mathrm{Supp}(\beta_1)\cap\mathrm{Supp}(\beta 2)|=w_2}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_2|\alpha_1}\cdot 1_{\beta_2|\beta_1}\cdot |A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|]$$

is at most

$$\sum_{\substack{0\le w_1\le r, 0\le w_2\le k\\ w_1+w_2\le D}} \binom{r}{w_1}\cdot\binom{k}{w_2}\cdot n^{2(D-w_1-w_2)}\cdot p^{k+r-w_1-w_2}\cdot\binom{N-2k+w_2}{m-k+w_2}$$

By separating out the parts dependent upon w_1 and w_2, the expression above is equal to

$$p^{k+r}\cdot n^{2(D)}\cdot\sum_{0\le w_1\le r}\binom{r}{w_1}\cdot n^{-2w_1}p^{-w_1}\cdot\sum_{0\le w_2\le D-w_1}\binom{k}{w_2}\cdot n^{-2w_2}\cdot p^{-w_2}\cdot\binom{N-2k+w_2}{m-k+w_2}$$

Let $g(w_2)=\binom{k}{w_2}\cdot n^{-2w_2}\cdot p^{-w_2}\cdot\binom{N-2k+w_2}{m-k+w_2}$. Let us consider the expression $g'(w_2)=g(w_2)/\binom{N-2k}{m-k}$. By our choice of parameters, $w_1^2=O(n^2)$, $k^2=O(n^2)$ and $N=\Omega(n^2)$. So by Lemma 2.13

$$\frac{\binom{N-2k+w_2}{m-k+w_2}}{\binom{N-2k}{m-k}}\approx\left(\frac{N-2k}{m-k}\right)^{w_2}$$

We also know from our choice of parameters that $\frac{N-2k}{m-k}=\Theta(1)$. So, $g'(w_2)=\binom{k}{w_2}\cdot n^{-2w_2}\cdot p^{-w_2}\cdot\Theta(1)^{w_2}$. For $p=n^{-\varepsilon}$ and $k=\Theta(n)$, $g'(w_2)\le k^{w_2}\cdot n^{\varepsilon w_2-2w_2}\cdot\Theta(1)^{w_2}$.

In particular, $g'(w_2)$ is upper bounded by a decreasing function of w_2 and takes the maximum value 1 at $w_2 = 0$. Hence,

$$\sum_{0 \le w_2 \le D - w_1} g(w_2) \le D \cdot \binom{N-2k}{m-k}$$

By a similar reasoning,

$$\sum_{0 \le w_1 \le r} \binom{r}{w_1} \cdot n^{-2w_1} p^{-w_1} \le r \cdot 1$$

So

$$\sum_{\substack{0 \le w_1 \le r, 0 \le w_2 \le k \\ w_1 + w_2 \le D}} \binom{r}{w_1} \cdot \binom{k}{w_2} \cdot n^{2(D-w_1-w_2)} \cdot p^{k+r-w_1-w_2} \cdot \binom{N-2k+w_2}{m-k+w_2}$$

is upper bounded by

$$p^{k+r} \cdot n^{2D} \cdot D \cdot \binom{N-2k}{m-k} \cdot r$$

For $k = n - r$, $D = \frac{\varepsilon n}{2} + d$ and $p = n^{-\varepsilon}$, this is at most

$$n^{2d+2} \cdot \binom{N-2k}{m-k}$$

□

Now, plugging this bound back into Equation 4.32, we get

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_3|_V] \quad \le \sum_{\substack{\alpha_1 \in \mathcal{M}^{[r]} \\ \beta_1 \in M(\alpha)}} \mathbb{E}_{V \leftarrow \mathcal{D}}[1_{\alpha_1, \beta_1}] \cdot n^{2d+2} \cdot \tbinom{N-2k}{m-k}$$

Now, $1_{\alpha_1,\beta_1} = 1$ when all the variables in the supports of α and β are alive. This happens with probability exactly p^n since $\alpha\beta$ is a multilinear monomial of degree n. Also, there are n^{2r} possible α and for each of these, there are exactly $n^{2(D-r)}$ many β in $M(\alpha)$. So,

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_3|_V] \le p^n \cdot n^{2r} \cdot n^{2(D-r)} \cdot n^{2d+2} \cdot \binom{N-2k}{m-k}$$

Putting in $D = \frac{\varepsilon n}{2} + d$ and $p = n^{-\varepsilon}$, we get

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_3|_V] \le n^{4d+2} \cdot \binom{N-2k}{m-k}$$

So, we obtain Lemma 4.21.

4.8 Lower bound for $IMM_{\tilde{n},n}$

In this section, we prove the lower bound on the size of homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing an entry in the product of generic matrices. The proof is similar in spirit to the proof of lower bound for the Nisan-Wigderson polynomials. In fact, the choice of parameters in this proof is strongly motivated by the choice of parameters in the earlier proof.

We will first introduce some notation needed for the proof.

4.8.1 Notation

Let $IMM_{\tilde{n},n}$ be the the polynomial computed by the $(1,1)$ coordinate of the product of n different $\tilde{n} \times \tilde{n}$ matrices, where the entries of the matrices are distinct variables. Thus there are $\tilde{n}^2 \times n$ variables in total.

Let $\tilde{n}, n, r', k'$ be positive integers such that $(k'+2)r' = n$. Let $IMM_{\tilde{n},n}{}^*(\tilde{n}, n, r', k')$ be an n-tuple of $\tilde{n} \times \tilde{n}$ matrices of the following form: The n tuples will be composed of r' blocks, each block having $k'+2$ matrices. In each block, the first matrix will be a special matrix, the next k' will be regular matrices, and the last one will be the all 1s matrix that we call J. Note that regular and special matrices continue to have $\tilde{n}^2$ distinct variables in them. In the ith block, we call the special matrix $Y^{(i)}$, the regular matrices are $X^{(i,1)}, X^{(i,2)}, \ldots, X^{(i,k')}$, and the last all 1s matrix is $J^{(i)}$. In the n-tuple, we arrange the matrices of the first block first, in the order described above, then the matrices of the second block, and so on. Thus the ith block, which we call $B^{(i)}$ is a $(k'+2)$-tuple of the form

$$\left(Y^{(i)}, X^{(i,1)}, X^{(i,2)}, \ldots, X^{(i,k')}, J^{(i)}\right),$$

and the n-tuple $IMM_{\tilde{n},n}{}^*(\tilde{n}, \tilde{k}, r', k')$ is a concatenation of the different blocks $B^{(i)}$, for $i \in [r']$.

Thus $IMM_{\tilde{n},n}{}^*(\tilde{n}, n, r', k')$ is of the following form:

$$\left(Y^{(1)}, X^{(1,1)}, X^{(1,2)}, \ldots, X^{(1,k')}, J^{(1)}, \ldots\ldots\ldots, Y^{(r')}, X^{(r',1)}, X^{(r',2)}, \ldots, X^{(r',k')}, J^{(r')}\right).$$

We will select the parameters $(\tilde{n}, n, r', k')$ right in the beginning and then use these fixed parameters for the rest of the paper. Thus for ease of notation we will often suppress the parameters and let $IMM_{\tilde{n},n}{}^* = IMM_{\tilde{n},n}{}^*(\tilde{n}, n, r', k')$.

For any matrix M, we let $m_{i,j}$ be the variable in the (i,j)th entry of M. We will use capital letters to denote the name of the matrix and the small letter to denote the variables in the matrix. For instance, the (i,j)th entry of the matrix $X^{(u,v)}$ is $x_{i,j}^{(u,v)}$.

Let $IMM_{\tilde{n},n}{}^{\times}$ be the matrix which is the product of all n matrices in $IMM_{\tilde{n},n}{}^*(\tilde{n}, n, r', k')$ in the order given above.

For $i, j \in [\tilde{n}]$, let P_{ij} be the polynomial computed at the (i,j) entry of $IMM_{\tilde{n},n}{}^{\times}$.

For our proof, we will initially fix a value of $\tilde{n}$ and n and work with it. So for the rest of the paper, we will supress the subscript $\tilde{n}, n$ from our notations.

Let $\overline{IMM}$ be $\mathsf{supp}(P_{11})$.

Let $\overline{IMM}_X$ be the set of monomials obtained from $\overline{IMM}$ after setting all the variables in the special matrices to 1. (When we talk about the set of monomials obtained, we disregard the information in the coefficients of the monomials obtained, and just treat them all to be monic.)

Let $\overline{IMM}_X^{(i)}$ be the set of monomials obtained from $\overline{IMM}$ after setting all the variables in all the matrices except the regular matrices of the ith block to 1. (Again, we disregard the coefficients of the monomials and treat them as monic monomials.)

Notice that

$$\overline{IMM}_X = \prod_{i \in [r']} \overline{IMM}_X^{(i)},$$

where every element of the product set is identified with the monomial formed by the product of the monomials from the individual sets.

Let $\overline{IMM}_Y$ be the set of monomials (all monomials are treated as monic in the set) obtained from $\overline{IMM}$ after setting all the variables in the regular matrices to 1. Notice that $|\overline{IMM}_Y| = (\tilde{n}^2)^{r'}$, since we get a monomial for every r'-tuple of variables where the ith element is a variable in $Y^{(i)}$.

For $\alpha \in \overline{IMM}_Y$, let $\overline{IMM}(\alpha)$ be the set of monomials β in $\overline{IMM}_X$ such $\alpha \cdot \beta$ is an element of $\overline{IMM}$.

For $\alpha \in \overline{IMM}_Y$, let $\overline{IMM}(\alpha)^{(i)}$ be the set of monomials in $\overline{IMM}(\alpha)$ obtained after all the variables that are not in the ith block have been set to 1.

4.8.2 Choice of parameters

We will pick the following choice of parameters:

1. n. (This denotes the total number of matrices in ${IMM_{\tilde{n},n}}^*$)
2. $r = \sqrt{n}$. (This will be the order of partial derivatives in the complexity measure)
3. $\tilde{n} = n^5$. (This is the dimension of the matrices)
4. $s = \frac{\sqrt{n}}{64}$. (This indicates the target support of a product gate in the circuit after random restrictions)
5. $\Lambda = 32$. (This is a parameter used in the proof)
6. $r' = \Lambda r$. (This is the number of blocks)
7. $k = n - 2r'$. (This is the number of regular matrices.)
8. $k' = k/r'$. (This is the number of regular matrices per block)
9. $N = (n - r') \cdot \tilde{n}^2$. (This is the total number of variables in ${IMM_{\tilde{n},n}}^*$)
10. Γ is a parameter (it will be a number very close to 2) which is chosen so that the following equalities hold. Set $m = \frac{N}{2}\left(1 - \frac{\ln n}{\Gamma\sqrt{n}}\right)$. Then choose Γ so that

$$n^r \cdot \left(\frac{N}{N-m}\right)^k \approx \left(\frac{N}{m}\right)^k.$$

Here, $\approx$ is used to indicate an equality up to a $n^{O(1)}$ factor.

Thus

$$n^r \approx \left(\frac{N-m}{m}\right)^k.$$

Using the choices of $r = \sqrt{n}, k = n - 2r'$ and $m = \frac{N}{2}\left(1 - \frac{\ln n}{\Gamma\sqrt{n}}\right)$, we get that

$$n = \left(\frac{\left(1 + \frac{\ln n}{\Gamma\sqrt{n}}\right)}{\left(1 - \frac{\ln n}{\Gamma\sqrt{n}}\right)}\right)^{\sqrt{n}-(2/\Lambda)+o(1)} = n^{\frac{2+o(1)}{\Gamma}}.$$

So, $\Gamma = 2 + o(1)$.

11. $m = \frac{N}{2}\left(1 - \frac{\ln n}{\Gamma\sqrt{n}}\right)$. (This is the degree of the multilinear shifts)

12. $D = N/(N-m)$. Thus $D^k = \left(\frac{N}{(N-m)}\right)^k$. (This is an indicator of the number of monomials in the support of the resulting polynomial after applying a restriction from our distribution and taking partial derivative with respect to a suitable monomial. Note that D is a number slightly smaller than 2 for our choice of m)

13. η is a parameter chosen so that

$$n^{\eta \cdot r'} \cdot 2^{k-(2\log n+1)r'} = D^k$$

Thus

$$\left(\frac{n^{\eta-2}}{2}\right)^{r'} \cdot 2^k = D^k = 2^k \cdot \left(\frac{1}{1+\frac{\ln n}{\Gamma\sqrt{n}}}\right)^k.$$

Thus

$$\frac{n^{\eta-2}}{2} = \left(\frac{1}{1+\frac{\ln n}{\Gamma\sqrt{n}}}\right)^{k'} = \left(\frac{1}{1+\frac{\ln n}{\Gamma\sqrt{n}}}\right)^{(1+o(1))\sqrt{n}/\Lambda} = n^{-\frac{1+o(1)}{\Gamma\Lambda}}.$$

Thus $\eta = 2 - \frac{1+o(1)}{\Gamma\Lambda}$.

4.8.3 Random restrictions

The total number of variables N in $IMM_{\tilde{n},n}{}^*$ is $N = \tilde{n}^2 \times (n - r')$. There are $(\tilde{n}^2 \times r')$ y-variables and $(\tilde{n}^2 \times k'r')$ x-variables. Let this total set of variables be $\mathcal{V}$. We will randomly set certain of these variables to zero, to get a distribution over *restrictions* of $IMM_{\tilde{n},n}$. We will now define a distribution $\mathcal{D}$ over subsets $V \subset \mathcal{V}$. The random restriction procedure will sample $V \leftarrow \mathcal{D}$ and then keep only those variables "alive" that come from V and set the rest to zero.

For each matrix in $IMM_{\tilde{n},n}{}^*$ we specify a random procedure for deciding which variables to set to zero, and then we will apply this procedure independently for each matrix.

Random restriction for special matrices

- For each special matrix $Y^{(i)}$, choose $\tilde{n}^{3/4}$ entries uniformly at random from the first row and keep those nonzero. Set all other variables to zero. The choice fo

this parameter is governed by some parameter constraints, which will be clear at a later stage.

Random restriction for regular matrices

Let $2 > \eta > 1$ be the parameter that was set in item 13 above.

- For each regular matrix of the form $X^{(i,1)}$ (i.e. the first regular matrix in any block), in each row, pick n^η distinct variables (uniformly at random), and keep them nonzero. Set the remaining variables to zero. Do this independently for each row.
- For each regular matrix of the form $X^{(i,j)}$, where $j > k' - 2\log n$ (i.e. the last $2\log n$ regular matrices in any block), in each row, pick 1 distinct variable (uniformly at random), and keep it nonzero. Set the remaining variables to zero. Do this independently for each row.
- For each regular matrix of the form $X^{(i,j)}$, where $2 \leq j \leq k' - 2\log n$, in each row, pick 2 distinct variable (uniformly at random), and keep them nonzero. Set the remaining variables to zero. Do this independently for each row.

In this manner, independently for each matrix in ${IMM_{\tilde{n},n}}^*$ we only keep a random subset of variables alive, and thus we get a distribution $\mathcal{D}$ over subsets $V \subset \mathcal{V}$ where V is the total set of alive variables. Notice that every $V \leftarrow \mathcal{D}$ is such that

$$|V| = r' \cdot (\tilde{n}^{3/4} + \tilde{n} \cdot n^\eta + (k' - 2\log n - 1) \cdot \tilde{n} \cdot 2 + 2\log n \cdot \tilde{n}).$$

Notation for restricted matrices

For each random subset of variables $V \leftarrow \mathcal{D}$ obtained in this way, let $IMM|_V^*$ be the the n-tuple of matrices ${IMM_{\tilde{n},n}}^*$ where only the variables in V are kept alive and the rest have been set to zero. Let $IMM|_V$ be the $(1,1)$ entry of the product of the matrices in $IMM|_V^*$. Let $(X^{(i,j)})|_V$ be the jth regular matrix of the ith block in $IMM|_V^*$. Let $(Y^{(i)})|_V$ be the ith special matrix in $IMM|_V^*$.

Let $\overline{IMM}|_V$,$(\overline{IMM}|_V)_X$, $(\overline{IMM}|_V)_X^{(i)}$, $(\overline{IMM}|_V)_Y$, $\overline{IMM}|_V(\alpha)$ and $\overline{IMM}|_V(\alpha)^{(i)}$ be obtained from $\overline{IMM}$,$\overline{IMM}_X$, $\overline{IMM}_X^{(i)}$, $\overline{IMM}_Y$, $\overline{IMM}(\alpha)$ and $\overline{IMM}(\alpha)^{(i)}$ respectively by keeping only those variables 'alive' that are present in V, and setting the remaining to zero.

Viewing $IMM|_V^*$ as a graph

Note than one can view any $\tilde{n} \times \tilde{n}$ matrix as the incidence matrix of a bipartite graph with $\tilde{n}$ left vertices and $\tilde{n}$ right vertices. For each entry in the (i, j) location that is nonzero, we add an edge from the ith left vertex to the jth right vertex with the variable written in the (i, j)th entry now written on the edge. (In the case of the J matrices (of all 1s), we just label the edges with 1.

Thus one can view any $IMM|_V^*$ as an n-tuple of bipartite graphs, where for any two adjacent matrices M, M' in the n-tuple, we identify the right vertices of M with the left vertices of M'. Thus we get a layered bipartite graph, with n layers, and each monomial in $\overline{IMM}|_V$ corresponds to a path from the leftmost layer to the rightmost layer. We define the ith layer in $IMM|_V^*$ to be precisely the bipartite graph corresponding the ith matrix in $IMM|_V^*$. The *degree* of a layer is defined to be the left-degree of the corresponding bipartite graph. Notice that at least for all the regular matrices, the corresponding bipartite graphs (after restricting to V) are regular with respect to the left-degrees. For the regular matrix $X^{(i,j)}|_V$, we let $\mathsf{Deg}(X^{(i,j)}|_V)$ denote the left degree of the corresponding bipartite graph, and by the random restriction process, note that this is a number only depending on the value of j. For ease of notation, we may some times refer to this quantity as $\mathsf{Deg}(j)$. For every left vertex of this graph (of degree $\mathsf{Deg}(j)$), we give each of the outgoing edge a distinct label from 1 to $\mathsf{Deg}(j)$. This choice of labels is assigned independently and uniformly at random for each left vertex. Thus for instance, for every left vertex, if we follow the edge labelled 1 that leaves it, we get a uniformly random element of $[\tilde{n}]$ as the right vertex.

Any element of $(\overline{IMM}|_V)_X^{(i)}$ is a monomial of degree k', and it corresponds to a path of length k' in the k'-layered bipartite graph corresponding to the regular matrices of the ith block. Each such monomial can thus be fully specified by first specifying the

start vertex, i.e. an element of $[\tilde{n}]$, and the labels of the edges along the path, i.e. a k'-tuple where the jth entry is free to vary in $[\mathsf{Deg}(X^{(i,j)}|_V)]$. This correspondence will be very useful in the arguments that will be coming up.

4.8.4 Choosing a set of monomials

From our definition of the complexity measure Φ, it depends upon two parameters. The degree of multilinear shift m has already been set by our choice of parameters. For every $V \leftarrow \mathcal{D}$, we will first choose an appropriate set of monomials of degree r' denoted by $\mathcal{T}(IMM|_V)$. The final set of monomials with respect to which we will take derivatives will be a large subset of $\mathcal{T}(IMM|_V)$. As we will see, the complexity of the circuit just depends on the parameter r' and is totally independent of the precise set of monomials with respect to which partial derivatives are taken. Hence, choosing the set of monomials dependent upon V does not lead to a problem.

For any $V \leftarrow \mathcal{D}$, let $\mathcal{T}(IMM|_V)$ be a subset of $(\overline{IMM}|_V)_Y$ chosen such that the following properties hold:

- $|\mathcal{T}(IMM|_V)| = n^r$
- For any two distinct monomials $\alpha, \beta \in \mathcal{T}(IMM|_V)$,

$$|\mathrm{Supp}(\alpha) \setminus \mathrm{Supp}(\beta)| = |\mathrm{Supp}(\beta) \setminus \mathrm{Supp}(\alpha)| \geq r' - r$$

The following lemma shows that such a set exists with a probability 1 over $V \leftarrow \mathcal{D}$.

Lemma 4.34. *For any $V \subseteq \mathcal{V}$ such that V lies in the support of the distribution $\mathcal{D}$, there exists $\mathcal{T}(IMM|_V) \subseteq (\overline{IMM}|_V)_Y$ such that the following two properties hold.*

- $|\mathcal{T}(IMM|_V)| = n^r$
- *For any two distinct monomials $\alpha, \alpha' \in \mathcal{T}(IMM|_V)$,*

$$|Supp(\alpha) \setminus Supp(\alpha')| = |Supp(\alpha') \setminus Supp(\alpha)| \geq r' - r$$

Proof. From the definition of the random restriction procedure, it follows that for each of Y matrices, $\tilde{n}^{3/4}$ variables in the first row are kept alive. We will identify the set of these variables with elements in the field $\mathbb{F}_q$ with $q = \tilde{n}^{3/4}$ for each of the Y matrices. We note that if $\tilde{n}^{3/4}$ is not a prime power then we can just take q to be something slightly larger and the analysis still works. For simplicity we assume for now that it is a prime power. Then, the cartesian product of the subset of alive (i.e. nonzero) variables in each of the Y matrices can be identified with $\mathbb{F}_q^{r'}$. For $r < r'$, we consider the set of all codewords of the Reed-Solomon codes corresponding to polynomials of degree at most $r-1$, and evaluated at r' distinct field elements. This gives is a subset of $\mathbb{F}_q^{r'}$ of size $q^r = \tilde{n}^{3r/4} = n^{15r/4}$ such that the distance between any two elements (which are r'-tuples) is at least $r' - r$. We take, $\mathcal{T}(IMM|_V)$ to be any subset of these codewords of size exactly n^r. □

Eventually in our proof, we will only look at derivatives of $IMM|_V$ with respect to a *good* subset $\mathcal{G}$ of monomials in $\mathcal{T}(IMM|_V)$. We will argue that with a high probability this set will have some good properties, which will help us lower bound the complexity of $IMM|_V$.

4.8.5 Proof overview

The proof of the lower bound for $IMM_{\tilde{n},n}$ is a little more subtle than the proof of lower bounds for $NW_{n,D}$.

- If the circuit was large to start with, we have nothing to prove. Else, we will argue that under the random restrictions given by the distribution $\mathcal{D}$, with high probability none of product gates in the bottom layer C has high support (all the high support gates set to zero).

- Assuming that the circuit has bounded support, we will obtain a good upper bound on its complexity. This is similar to the corresponding step in $NW_{n,D}$.

- We will then show that with a good probability, the complexity of a random restriction of $IMM_{\tilde{n},n}$ remains high. This is the most technical part of the proof.

We elaborate more on this step next.

- We will argue that the probability that both of the above items happen together is high. Then, comparing the complexity of the circuit and the polynomial $IMM|_V$ completes the proof.

Lower bound on the complexity of a random restriction of $IMM_{\tilde{n},n}$: In spirit, this proof is like that for $NW_{n,D}$. Analogous to the definitions of the expressions T_1, T_2, T_3 for $NW_{n,D}$, for every restriction $V \leftarrow \mathcal{D}$, and with respect to a set of monomials $\mathcal{T}(IMM|_V)$ as given by the Lemma 4.34, we define

- $T_1(IMM|_V) = \sum_{\substack{\alpha \in \mathcal{T}(IMM|_V) \\ \beta \in \text{Supp}(\partial_\alpha(IMM_{\tilde{n},n}))}} 1_{\alpha,\beta} \cdot |S_m(\alpha,\beta)|$
- $T_2(IMM|_V) = \sum_{\substack{\alpha \in \mathcal{T}(IMM|_V) \\ \beta,\gamma \in \text{Supp}(\partial_\alpha(IMM_{\tilde{n},n})) \\ \beta \neq \gamma}} 1_{\alpha,\beta,\gamma} \cdot |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|$
- $T_3(IMM|_V) = \sum_{\substack{\alpha_1,\alpha_2 \in \mathcal{T}(IMM|_V) \\ \beta_1 \in \text{Supp}(\partial_{\alpha_1}(IMM_{\tilde{n},n})) \\ \beta_2 \in \text{Supp}(\partial_{\alpha_2}(IMM_{\tilde{n},n})) \\ (\alpha_1,\beta_1) \neq (\alpha_2,\beta_2)}} 1_{\alpha_1,\alpha_2,\beta_1,\beta_2} \cdot |A_m(\alpha_1,\beta_1) \cap A_m(\alpha_2,\beta_2)|$

We will use $T_1|_V$ for $T_1(IMM|_V)$, $T_2|_V$ for $T_2(IMM|_V)$ and $T_3|_V$ for $T_3(IMM|_V)$. Observe that the definitions above are equivalent to the following definitions.

- $T_1|_V = \left[\sum_{\substack{\alpha \in \mathcal{T}(IMM|_V) \\ \beta \in \overline{IMM}|_V(\alpha)}} |S_m(\alpha,\beta)|\right] = \left[\sum_{\substack{\alpha \in \mathcal{T}(IMM|_V) \\ \beta \in \overline{IMM}|_V(\alpha)}} \binom{N-k}{m}\right]$,
  where the last equality holds because $S(\alpha,\beta)$ is the set of all multilinear monomials of degree m which are disjoint from β.

- $$T_2|_V = \sum_{\alpha \in \mathcal{T}(IMM|_V)} \left(\sum_{\substack{\beta,\gamma \in \overline{IMM}|_V(\alpha) \\ \beta \neq \gamma}} |S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)| \right)$$

  $$= \sum_{\alpha \in \mathcal{T}(IMM|_V)} \left(\sum_{\substack{\beta,\gamma \in \overline{IMM}|_V(\alpha) \\ \beta \neq \gamma}} \binom{N-k-\Delta(\beta,\gamma)}{m} \right)$$

  Where the last equality holds because $|S_m(\alpha,\gamma) \cap S_m(\alpha,\beta)|$ counts the number of multilinear monomials of degree m which are disjoint from both β and γ.

- $$T_3|_V = \sum_{\substack{\alpha_1,\alpha_2 \in \mathcal{T}(IMM|_V) \\ \beta_1 \in \overline{IMM}|_V(\alpha_1) \\ \beta_2 \in \overline{IMM}|_V(\alpha_2) \\ (\alpha_1,\beta_1) \neq (\alpha_2,\beta_2)}} |A_m(\alpha_1,\beta_1) \cap A_m(\alpha_2,\beta_2)|$$

  $$T_3|_V \leq \sum_{\substack{\alpha_1,\alpha_2 \in \mathcal{T}(IMM|_V) \\ \beta_1 \in \overline{IMM}|_V(\alpha_1) \\ \beta_2 \in \overline{IMM}|_V(\alpha_2) \\ (\alpha_1,\beta_1) \neq (\alpha_2,\beta_2)}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)}$$

  Where the last inequality holds since $|A_m(\alpha_1,\beta_1) \cap A_m(\alpha_2,\beta_2)|$ is upper bounded by the number of multilinear monomials of degree $m+k$ which are divisible by both β_1 and β_2.

For every pair of monomials $\alpha, \alpha' \in \mathcal{T}(IMM|_V)$, we define

- $T_1|_V(\alpha) = \sum_{\beta \in \overline{IMM}|_V(\alpha)} |S_m(\alpha,\beta)|$
- $T_2|_V(\alpha) = \sum_{\substack{\beta,\gamma \in \overline{IMM}|_V(\alpha) \\ \beta \neq \gamma}} \binom{N-k-\Delta(\beta,\gamma)}{m}$
- If $\alpha = \alpha'$, then $T_3|_V(\alpha,\alpha') = \sum_{\substack{\beta,\gamma \in \overline{IMM}|_V(\alpha) \\ \beta \neq \gamma}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)}$
- If $\alpha \neq \alpha'$, $T_3|_V(\alpha,\alpha') = \sum_{\substack{\beta \in \overline{IMM}|_V(\alpha) \\ \gamma \in \overline{IMM}|_V(\alpha')}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)}$

We will now describe the strategy to prove to a lower bound on the complexity of $IMM|_V$. We compute the expected values of expression $T_1|_V$, $T_2|_V$ and $T_3|_V$ for V sampled according to $\mathcal{D}$. Then, we argue that with a high probability, $T_2|_V$ and $T_3|_V$ have values not much larger than their expectations and $T_1|_V$ has value close to its expectation. For such *good* restrictions, we show the existence of a set $\mathcal{G}_V \subseteq \mathcal{T}(IMM|_V)$ with the following properties.

1. For each α in $\mathcal{G}_V$, $T_1|_V(\alpha)$ is large.
2. For each α in $\mathcal{G}_V$, $T_2|_V(\alpha)$ is not too large compared to $T_1(\alpha)$.
3. $\sum_{\alpha_1,\alpha_2 \in \mathcal{G}_V} T_3|_V(\alpha_1,\alpha_2)$ is not too large when compared to

$$\sum_{\alpha \in \mathcal{G}_V, \beta \in \mathrm{Supp}(\partial_\alpha(IMM|_V))} |A_m(\alpha,\beta)|.$$

Then, we show that these conditions suffice to show that $\Phi_{\mathcal{G}_V,m}(IMM|_V)$ is large. This argument has the following major steps.

- For each $\alpha \in \mathcal{G}_V$, since $T_1|_V(\alpha)$ is large, it follows that $\sum_{\beta \in \overline{IMM}|_V(\alpha)} |S_m(\alpha,\beta)|$ is large.
- For each $\alpha \in \mathcal{G}_V$, since $T_2|_V(\alpha)$ is not much larger than $T_1|_V(\alpha)$, Lemma 4.11 and Lemma 4.15 imply that for each $\alpha \in \mathcal{G}_V$, $\sum_{\beta \in \overline{IMM}|_V(\alpha)} |A_m(\alpha,\beta)|$ is large.
- We also know that $\sum_{\alpha_1,\alpha_2 \in \mathcal{G}_V} T_3|_V(\alpha_1,\alpha_2) = \sum_{\substack{\alpha_1,\alpha_2 \in \mathcal{G}_V \\ \beta_1 \in \overline{IMM}|_V(\alpha_1) \\ \beta_2 \in \overline{IMM}|_V(\alpha_2) \\ (\alpha_1,\beta_1) \neq (\alpha_2,\beta_2)}} |A_m(\alpha_1,\beta_1) \cap A_m(\alpha_2,\beta_2)|$ is not much larger than $\sum_{\alpha \in \mathcal{G}_V, \beta \in \overline{IMM}|_V(\alpha)} |A_m(\alpha,\beta)|$.
- Lemma 4.11 will then imply that $\left| \bigcup_{\substack{\alpha \in \mathcal{G}_V \\ \beta \in \overline{IMM}|_V(\alpha)}} A_m(\alpha,\beta) \right|$ is large. Hence, by Lemma 4.13, $\Phi_{\mathcal{G}_V,m}(IMM|_V)$ is large.

4.8.6 Effect of random restrictions on the circuit

We will now analyze the effect of the random restrictions on a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing the polynomial $IMM_{\tilde{n},n}$ and show that with a high probability, no large support product gate survives.

Lemma 4.35. *Let C be a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit of size at most $n^{\frac{\sqrt{n}}{128}}$ computing the polynomial $IMM_{\tilde{n},n}$. Then, with a probability at least $1 - o(1)$ over $V \leftarrow \mathcal{D}$, $C|_V$ is a $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuit, for $s = \frac{\sqrt{n}}{64}$.*

Proof. We will analyze the probability that a fixed product gate at the bottom layer of C (that computes a monomial) of support size s (we will later set $s = \frac{\sqrt{n}}{64}$) survives[4] the random restriction procedure. Observe that the events that two variables in different matrices in $IMM_{\tilde{n},n}{}^*$ survive are independent, but the probability that two variables within the same matrix survive are correlated. We will first upper bound the probability that a monomial has support t within any layer (i.e. t distinct variables that all come

[4]We say that a product gate survives the random restriction if none of the variables feeding in to it are set to zero.

from the same layer) survives the random restriction procedure, based on the type of the layer. We will think of t to be $O(\sqrt{n})$.

- **Special matrices:** In a special layer, a random subset of $\tilde{n}^{3/4}$ variables in the first row is kept alive. The probability that a monomial of support t within this layer survives is, therefore at most $\frac{\binom{\tilde{n}-t}{\tilde{n}^{3/4}-t}}{\binom{\tilde{n}}{\tilde{n}^{3/4}}}$. Since t is $O(\sqrt{n})$ and $\tilde{n} = n^5$, so $\tilde{n}$ and $\tilde{n}^{3/4}$ are both $\Omega(t^2)$. Hence, $\frac{\binom{\tilde{n}-t}{\tilde{n}^{3/4}-t}}{\binom{\tilde{n}}{\tilde{n}^{3/4}}} \approx \frac{\tilde{n}^{-t}}{\tilde{n}^{-3t/4}}$, by Lemma 2.13. So, the probability of survival is at most $\frac{1}{\tilde{n}^{t/4}} < \frac{1}{n^t}$.

- **Regular matrices of the form $X^{(i,1)}$:** Here, in each row exactly n^η random variables are kept alive. For $\eta \geq 1$, the probability that a fixed monomial with support at least $t' = O(\sqrt{n})$ within any row survives is at most $\frac{\binom{\tilde{n}-t'}{n^\eta-t'}}{\binom{\tilde{n}}{n^\eta}} \approx \frac{\tilde{n}^{-t'}}{n^{-\eta \cdot t'}}$. Also, the events across different rows are independent. So, the probability that a monomial with support at least t in the variables in this matrix survives is at most $\frac{\tilde{n}^{-t}}{n^{-\eta \cdot t}} \leq n^{(\eta-5)\cdot t} < n^{-t}$.

- **Regular matrices $X^{i,j}$ for $j > k' - 2\log n$:** In these matrices, exactly one variable in each row is kept alive uniformly at random. So, the probability that a monomial of support at least t within one of these matrices survives the random restriction procedure is at most $\tilde{n}^{-t}$.

- **Regular matrices $X^{i,j}$ for $2 \leq j \leq k' - 2\log n$:** In these matrices, from each row, two distinct variables chosen uniformly at random are kept alive by the random restriction procedure. So, the probability that a fixed variable within a fixed row survives is at most $2 \cdot \tilde{n}^{-1}$. Therefore, the probability that a monomial of support at least t in such a matrix survives is at most $2^t \cdot \tilde{n}^{-t}$. For $\tilde{n} = n^5$, this is at most n^{-t}.

From the above bounds, it follows that for $t = O(\sqrt{n})$, the probability that a monomial that has support at least t within any single layer survives is at most $\frac{1}{n^t}$. Also, the events are independent across different layers. So the probability that any monomial with support at least t across all layers survives is at most $\frac{1}{n^t}$. Therefore, by the union bound, the probability that at least one gate with support larger than s survives is at

most $\frac{\text{Size}(C)}{n^s}$. For C such that $Size(C) \leq n^{\frac{\sqrt{n}}{128}}$ and $s = \frac{\sqrt{n}}{64}$, the probability that any product gate with support at least s survives the random restriction procedure is at most $n^{-\frac{\sqrt{n}}{128}}$. So, the lemma follows. □

4.8.7 Effect of random restrictions on $IMM_{\tilde{n},n}$

In this subsection, we will show that with a high probability over the random restrictions, the complexity of $IMM_{\tilde{n},n}$ remains high, assuming that the bounds given by the following lemmas.

Lemma 4.36. *For all $V \leftarrow \mathcal{D}$, and for every $\alpha \in \mathcal{T}(IMM|_V)$,*

$$T_1|_V(\alpha) = D^k \cdot \binom{N-k}{m}.$$

Lemma 4.37.

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}}[T_2|_V] \leq n^r \cdot D^k \cdot \binom{N-k}{m} \cdot n^{o(r)}$$

Lemma 4.38.

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}}[T_3|_V] \leq n^r \cdot D^k \cdot O(n^{(4/\Lambda)r}) \cdot \binom{N-k}{m}$$

We will also need the following lemma, which implies Lemma 4.37 via linearity of expectations.

Recall that for $\alpha \in \mathcal{T}(IMM|_V)$, we define $T_2|_V(\alpha) = \sum_{\beta,\gamma \in \overline{IMM}|_V(\alpha)} \binom{N-k-\Delta(\beta,\gamma)}{m}$. When $\alpha \notin \mathcal{T}(IMM|_V)$, we define $T_2|_V(\alpha) = 0$.

Lemma 4.39. $\forall \alpha \in (\overline{IMM}|_V)_Y$,

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}}[T_2|_V(\alpha)] \leq D^k \cdot \binom{N-k}{m} \cdot n^{o(r)}$$

We will prove these lemmas in Section 4.9

We will now show using Markov's inequality that $T_2|_V$ and $T_3|_V$ take values close to their expected values with a high probability.

Lemma 4.40.

$$Pr_{V \leftarrow \mathcal{D}}\left[T_2|_V < 20 \cdot \mathop{\mathbb{E}}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}] \wedge T_3|_V < 20 \cdot \mathop{\mathbb{E}}_{V' \leftarrow \mathcal{D}}[T_3|_{V'}]\right] \geq 0.9$$

Proof. The proof follows from the Markov's inequality and the union bound. □

Lemma 4.41. *With probability at least* 0.9 *over* $V \leftarrow \mathcal{D}$*, there exists a set* $\mathcal{G}_V \subseteq \mathcal{T}(IMM|_V)$ *such that the following are true:*

$$|\mathcal{G}_V| \geq \frac{4}{5} \cdot |\mathcal{T}(IMM|_V)|$$

And

$$\forall \alpha \in \mathcal{G}_V, T_2|_V(\alpha) \leq 100 \cdot \mathop{\mathbb{E}}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]/(n^r)$$

Proof. Let $V \subseteq \mathcal{V}$ be such that the bounds in Lemma 4.40 hold. Let $\mathcal{G}_V$ be the set of $\alpha \in \mathcal{T}(IMM|_V)$ such that $T_2|_V(\alpha) \leq 100 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]/(n^r)$. We will now argue that $|\mathcal{G}_V| \geq \frac{4}{5} \cdot |\mathcal{T}(IMM|_V)|$. Let us assume this is not true, then $\sum_{\alpha \in \mathcal{T}(IMM|_V)} T_2|_V(\alpha) \geq \sum_{\alpha \in \mathcal{T}(IMM|_V) \setminus \mathcal{G}_V} T_2|_V(\alpha) > \frac{1}{5} \cdot 100 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]/(n^r) \cdot |\mathcal{T}(IMM|_V)| = 20 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]$ which contradicts the fact that $\sum_{\alpha \in \mathcal{T}(IMM|_V)} T_2|_V(\alpha) = T_2|_V < 20 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]$.

□

Lemma 4.42. *With probability at least* 0.9 *over* $V \leftarrow \mathcal{D}$*, there exists a set of monomials* $\mathcal{G}_V$*, each of degree equal to* r' *such that*

$$\Phi_{\mathcal{G}_V, m}(IMM|_V) \geq \frac{n^r}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot D^k \cdot \binom{N-k}{m}$$

Proof. Lemma 4.41 guarantees that with a probability at least 0.9 over $V \leftarrow \mathcal{D}$, there exists a subset $\mathcal{G}_V \subseteq \mathcal{T}(IMM|_V)$, satisfying

$$|\mathcal{G}_V| \geq \frac{4}{5} \cdot |\mathcal{T}(IMM|_V)|$$

And,

$$\forall \alpha \in \mathcal{G}_V, T_2|_V(\alpha) \leq 100 \cdot \mathop{\mathbb{E}}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]/(n^r).$$

Moreover, $T_2|_V < 20 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]$ and $T_3|_V < 20 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_3|_{V'}]$. From the definition of sets $S_m(\alpha, \beta)$, and the above mentioned bounds, it follows that for all $\alpha \in \mathcal{G}_V$

$$T_1|_V(\alpha) = \sum_{\beta \in \overline{IMM}|_V(\alpha)} |S_m(\alpha, \beta)| = D^k \cdot \binom{N-k}{m}$$

and

$$T_2|_V(\alpha) = \sum_{\substack{\beta_1,\beta_2 \in \overline{IMM}|_V(\alpha) \\ \beta_1 \neq \beta_2}} |S_m(\alpha,\beta_1) \cap S_m(\alpha,\beta_2)| \leq 100 \cdot n^{o(r)} \cdot D^k \cdot \binom{N-k}{m}$$

Hence, by Lemma 4.11, we get that for all $\alpha \in \mathcal{G}_V$,

$$\left| \bigcup_{\beta \in \overline{IMM}|_V(\alpha)} S_m(\alpha,\beta) \right| \geq \frac{1}{O(n^{o(r)})} \cdot D^k \cdot \binom{N-k}{m}$$

By Lemma 4.15, it follows that for all $\alpha \in \mathcal{G}_V$

$$\begin{aligned} \sum_{\beta \in \overline{IMM}|_V(\alpha)} |A_m(\alpha,\beta)| &\geq \left| \bigcup_{\beta \in \overline{IMM}|_V(\alpha)} S_m(\alpha,\beta) \right| \\ &\geq \frac{1}{O(n^{o(r)})} \cdot D^k \cdot \binom{N-k}{m} \end{aligned}$$

Consequently,

$$\begin{aligned} \sum_{\alpha \in \mathcal{G}_V} \sum_{\beta \in \overline{IMM}|_V(\alpha)} |A_m(\alpha,\beta)| &\geq \frac{1}{O(n^{o(r)})} \cdot D^k \cdot \binom{N-k}{m} \cdot |\mathcal{G}_V| \\ &\geq \frac{n^r}{O(n^{o(r)})} \cdot D^k \cdot \binom{N-k}{m} \end{aligned}$$

Also,

$$\begin{aligned} \sum_{\alpha_1,\alpha_2 \in \mathcal{G}_V} T_3|_V(\alpha_1,\alpha_2) &\leq \sum_{\alpha_1,\alpha_2 \in \mathcal{T}(IMM|_V)} T_3|_V(\alpha_1,\alpha_2) \\ &= T_3|_V \\ &< 20 \mathop{\mathbb{E}}_{V' \leftarrow \mathcal{D}} [T_3|_{V'}] \\ &\leq O(n^{(4/\Lambda)r}) \cdot n^r D^k \cdot \binom{N-k}{m} \end{aligned}$$

and hence

$$\begin{aligned} \sum_{\substack{\alpha_1,\alpha_2 \in \mathcal{G}_V \\ \beta_1 \in \overline{IMM}|_V(\alpha_1) \\ \beta_2 \in \overline{IMM}|_V(\alpha_2) \\ (\alpha_1,\beta_1) \neq (\alpha_2,\beta_2)}} |A_m(\alpha_1,\beta_1) \cap A_m(\alpha_2,\beta_2)| & \\ = \sum_{\alpha_1,\alpha_2 \in \mathcal{G}_V} T_3|_V(\alpha_1,\alpha_2) & \\ \leq O(n^{(4/\Lambda)r}) \cdot n^r D^k \cdot \binom{N-k}{m} & \end{aligned}$$

So, we have

$$\begin{aligned}\sum_{\substack{\alpha_1,\alpha_2\in\mathcal{G}_V \\ \beta_1\in\overline{IMM}|_V(\alpha_1) \\ \beta_2\in\overline{IMM}|_V(\alpha_2) \\ (\alpha_1,\beta_1)\neq(\alpha_2,\beta_2)}} & |A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)| \\ \leq\quad & O(n^{(4/\Lambda)r})\cdot n^{o(r)}\cdot\sum_{\alpha\in\mathcal{G}_V}\sum_{\beta\in\overline{IMM}|_V(\alpha)}|A_m(\alpha,\beta)|\end{aligned}$$

Therefore, by Lemma 4.11, we have

$$\begin{aligned}\left|\bigcup_{\substack{\alpha\in\mathcal{G}_V \\ \beta\in\overline{IMM}|_V(\alpha)}} A_m(\alpha,\beta)\right| &\geq \frac{1}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}}\cdot\sum_{\substack{\alpha\in\mathcal{G}_V \\ \beta\in\overline{IMM}|_V(\alpha)}}|A_m(\alpha,\beta)| \\ &\geq \frac{n^r}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}}\cdot D^k\cdot\binom{N-k}{m}\end{aligned}$$

Now by Lemma 4.13,

$$\Phi_{\mathcal{G}_V,m}(IMM|_V)\geq\frac{n^r}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}}\cdot D^k\cdot\binom{N-k}{m}$$

□

4.8.8 Wrapping up the proof

We will now complete the proof of the main theorem.

Theorem 4.43. *Any homogeneous* ΣΠΣΠ *circuit computing the polynomial* $IMM_{\tilde{n},n}$ *has size at least* $2^{\Omega(\sqrt{n}\log n)}$.

Proof. Let C be a homogeneous ΣΠΣΠ circuit computing the polynomial $IMM_{\tilde{n},n}$. If $\text{Size}(C)\geq n^{\frac{\sqrt{n}}{128}}$, then we have nothing to prove and we are done, else Lemma 4.35 implies that with a probability $1-o(1)$, the circuit $C|_V$ does not have any product gate in the bottom layer of support larger than $s=\frac{\sqrt{n}}{64}$. Also, $\text{Size}(C|_V)\leq\text{Size}(C)$. Therefore, for any set $\mathcal{G}_V$ of monomials of degree r' and any positive integer m,

$$\Phi_{\mathcal{G}_V,m}(C|_V)\leq O(n)\cdot\text{Size}(C|_V)\cdot\binom{\lceil\frac{2n}{s}\rceil+r'}{r'}\cdot\binom{N}{m+r's}$$

From Lemma 4.42, we also know that with a probability at least 0.9, for random restriction $V\leftarrow\mathcal{D}$, there exists a set $\mathcal{G}_V$ of monomials of degree r' such that

$$\Phi_{\mathcal{G}_V,m}(IMM|_V)\geq\frac{n^r}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}}\cdot D^k\cdot\binom{N-k}{m}$$

Therefore, with a probability at least $0.9 - o(1)$, both these bounds hold. Since the circuit $C|_V$ computes the polynomial $IMM|_V$. Hence, $\Phi_{\mathcal{G}_V,m}(C|_V) \geq \Phi_{\mathcal{G}_V,m}(IMM|_V)$ for all V. Plugging back the values from above, and the observation that $\text{Size}(C|_V) \leq \text{Size}(C)$, we get

$$\text{Size}(C) \geq \frac{\frac{n^r}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot D^k \cdot \binom{N-k}{m}}{O(n) \cdot \binom{\lceil \frac{2n}{s} \rceil + r'}{r'} \cdot \binom{N}{m+r's}}$$

From our choice of parameters

- $r' = \Lambda r$
- $n^r \cdot D^k = \left(\frac{N}{m}\right)^k$
- $k = n - 2r'$
- $m = \frac{N}{2}\left(1 - \frac{\ln n}{\Gamma\sqrt{n}}\right)$
- $s = \frac{\sqrt{n}}{64}$
- $\Lambda = 32$

For these choice of parameters, observe that

- $\binom{\lceil \frac{2n}{s} \rceil + r'}{r'} = 2^{O(\sqrt{n})}$
- $\frac{\binom{N-k}{m}}{\binom{N}{m+r's}} = \frac{N-k!}{N!} \cdot \frac{(m+r's)!}{m!} \cdot \frac{(N-m-r's)!}{(N-m-k)!} \approx \frac{m^{r's}}{N^k} \cdot \frac{(N-m)^k}{(N-m)^{r's}}$

Plugging the value of the parameters and the bounds above back into equation Theorem 4.8.8, we get

$$\begin{aligned}
\text{Size}(C) &\geq \frac{\frac{n^r}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}} \cdot D^k \cdot \binom{N-k}{m}}{O(n)\cdot \binom{\lceil \frac{2n}{s} \rceil + r'}{r'} \cdot \binom{N}{m+r's}} \\
&\geq \frac{1}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}} \cdot \left(\frac{N}{m}\right)^k \cdot 2^{-O(\sqrt{n})} \cdot \frac{m^{r's}}{N^k} \cdot \frac{(N-m)^k}{(N-m)^{r's}} \\
&= \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}} \cdot \left(\frac{N-m}{m}\right)^{k-r's} \\
&= \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}} \cdot \left(\frac{1+\frac{\ln n}{\Gamma\sqrt{n}}}{1-\frac{\ln n}{\Gamma\sqrt{n}}}\right)^{k-r's} \\
&\geq \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}} \cdot \left(1+\frac{\ln n}{\Gamma\sqrt{n}}\right)^{k-r's} \\
&\geq \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}} \cdot e^{0.5\cdot(n-2r'-r's)\frac{\ln n}{\Gamma\sqrt{n}}} \\
&\geq \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}} \cdot n^{0.5(\frac{\sqrt{n}}{\Gamma} - \frac{r'(2+s)}{\Gamma\sqrt{n}})} \\
&\geq \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}} \cdot n^{\frac{\sqrt{n}}{2\Gamma} - \frac{\Lambda r(2+s)}{2\Gamma\sqrt{n}}} \\
&\geq \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)\sqrt{n}})\cdot n^{o(r)}} \cdot n^{\frac{\sqrt{n}-\Lambda s}{2\Gamma}} \qquad \text{by substituting } r=\sqrt{n}
\end{aligned}$$

Here we used our choice of parameters, namely $m = \frac{N}{2}\left(1 - \frac{\ln n}{\Gamma\sqrt{n}}\right)$, $k = n - 2r'$ and the fact that for $x = o(1), e^{x/2} \leq 1 + x$.

Now, by substituting $\Lambda = 32$, $\Gamma = 2 + o(1)$ and $s = \frac{\sqrt{n}}{64}$, we obtain

$$\text{Size(C)} \geq 2^{-O(\sqrt{n})} \cdot n^{\Omega(\sqrt{n})}.$$

□

4.9 Calculations for $IMM_{\tilde{n},n}$

In this section, we provide the calculations which establish the bounds in Lemma 4.36, Lemma 4.37, Lemma 4.38. In the next section, we will first prove technical results that will be the building blocks of the lemmas.

4.9.1 Preliminary lemmas

Here, the union is over all V in the support of the distribution $\mathcal{D}$. The reader should think of α as one of the partial derivatives of the polynomial.

Proposition 4.44. *Let α be any monomial in $\overline{IMM}_Y$. For every fixed $\beta \in \overline{IMM}_X$,*

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}} \left[\sum_{\gamma \in \overline{IMM}|_V(\alpha)} D^{-\Delta(\beta,\gamma)} \right] \leq n^{o(r)}.$$

The proof follows from Lemma 4.45 that we state and prove below. We give the formal proof at the end of the subsection.

For any monomial $\beta \in \overline{IMM}_X$, we define $\beta^{(i)} \in \overline{IMM}_X^{(i)}$ to be the resulting monomial after setting all the nonzero variables that are not in the ith block to 1.

Lemma 4.45. *For all $\beta^{(i)} \in \overline{IMM}_X^{(i)}$,*

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}} \left[\sum_{\gamma^{(i)} \in (\overline{IMM}|_V)(\alpha)^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right] \leq O(1).$$

Proof. The proof follows immediately from Lemmas Lemma 4.46 and Lemma 4.47 below by taking a sum of the two bounds. □

For all $\beta^{(i)} \in \overline{IMM}_X^{(i)}$, we define the following two sets.

- $\mathcal{A}_V^{(i)}(\beta^{(i)})$ is the set of all $\gamma^{(i)} \in (\overline{IMM}|_V)(\alpha)^{(i)}$ such that there is some $j \in [k'-1]$ such that $\gamma^{(i,j)} \neq \beta^{(i,j)}$ and $\gamma^{(i,j+1)} = \beta^{(i,j+1)}$
- $\mathcal{B}_V^{(i)}(\beta^{(i)})$ is the set of all $\gamma^{(i)} \in (\overline{IMM}|_V)(\alpha)^{(i)}$ such that if for $j, j' \in [k']$ $\gamma^{(i,j)} = \beta^{(i,j)}$ and $\gamma^{(i,j')} \neq \beta^{(i,j')}$, then $j' > j$.

Observe that $\mathcal{A}_V^{(i)}(\beta^{(i)}) \cup \mathcal{B}_V^{(i)}(\beta^{(i)}) = (\overline{IMM}|_V)(\alpha)^{(i)}$.

Thus we have partitioned the set of $\gamma^{(i)} \in (\overline{IMM}|_V)(\alpha)^{(i)}$ into two sets $\mathcal{A}_V^{(i)}(\beta^{(i)})$ and $\mathcal{B}_V^{(i)}(\beta^{(i)})$, and we estimate the expression in Lemma 4.45 separately as $\gamma^{(i)}$ varies in these sets. This calculation is carried out in Lemmas Lemma 4.46 and Lemma 4.47 below.

Lemma 4.46. *For all* $\beta^{(i)} \in \overline{IMM}_X^{(i)}$,

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}} \left[\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)}, \gamma^{(i)})} \right] \leq O(1).$$

Proof. We partition $\mathcal{B}_V^{(i)}(\beta^{(i)})$ into $k'+1$ sets, based on the number of locations j for which $\gamma^{(i,j)} = \beta^{(i,j)}$. For $0 \leq j \leq k'$, let $\mathcal{B}_V^{(i,j)}(\beta^{(i)})$ be the set of all $\gamma^{(i)} \in (\overline{IMM}|_V)(\alpha)^{(i)}$ such that $\gamma^{(i)}$ and $\beta^{(i)}$ agree on exactly the first j variables.

We now bound the size of $\mathcal{B}_V^{(i,j)}(\beta^{(i)})$. Notice that once we fix $\beta^{(i)}$, the first j variables of any $\gamma^{(i)}$ in $\mathcal{B}_V^{(i,j)}(\beta^{(i)})$ are determined. For each of the remaining variables $\gamma^{(i,j')}$ such that $j' > j$, the total number different choices they can take is at most $\mathsf{Deg}(X^{(i,j')})$.

Thus

$$|\mathcal{B}_V^{(i,j)}(\beta^{(i)})| \leq \prod_{j'=j+1}^{k'} \mathsf{Deg}(X^{(i,j')}).$$

Now, observe that $\prod_{j'=1}^{k'} \mathsf{Deg}(X^{(i,j')}) = D^{k'}$. This follows from the exact choice of degrees and value of D as set in the choice of parameters in Subsection 4.8.2. Thus we get that

$$\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i,j)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)}, \gamma^{(i)})} \leq \prod_{j'=j+1}^{k'} \mathsf{Deg}(X^{(i,j')}) \cdot D^{-(k'-j)}$$

Thus,

$$\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i,j)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)}, \gamma^{(i)})} \leq D^j \prod_{j'=1}^{j} \mathsf{Deg}(X^{(i,j')})^{-1}$$

Now for $j = 0$, the expression above equals 1. For $j > k' - 2\log n$, since $\mathsf{Deg}(X^{(i,j)}) = 1$, thus

$$\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i,j)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)}, \gamma^{(i)})} \leq D^{-(k'-j)}.$$

For $j \leq k' - 2\log n$, using the fact that $D < 2$, $\mathsf{Deg}(X^{(i,1)}) = n^{\eta}$ and $\mathsf{Deg}(X^{(i,j')}) = 2$ for $2 \leq j' \leq k' - 2\log n$, we get that

$$\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i,j)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)}, \gamma^{(i)})} \leq D^j \prod_{j'=1}^{j} \mathsf{Deg}(X^{(i,j')})^{-1}$$

$$= \frac{D}{\mathsf{Deg}(X^{(i,1)})} \cdot \prod_{j'=2}^{j} \frac{D}{\mathsf{Deg}(X^{(i,j')})}$$
$$\leq \frac{2}{n^\eta}$$

Putting together these values for all values of j, and using the fact that $k' < n/2 \leq \frac{n^\eta}{2}$, we get that

$$\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} = \sum_{j=0}^{k'} \sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i,j)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}$$
$$\leq 1 + (k' - 2\log n) \cdot \frac{2}{n^\eta} + \sum_{j=k'-2\log n+1}^{k'} D^{-(k'-j)}$$
$$\leq 2 + \sum_{j=0}^{2\log n} D^{-j}$$
$$\leq 2 + \frac{1}{1 - D^{-1}}$$
$$\leq 5$$

□

Lemma 4.47. *For all $\beta^{(i)} \in \overline{IMM}_X^{(i)}$,*

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}} \left[\sum_{\gamma^{(i)} \in \mathcal{A}_V^{(i)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right] \leq O(1/n).$$

Proof. For $\gamma^{(i)} \in \mathcal{A}_V^{(i)}(\beta^{(i)})$, we call a coordinate j such that $2 \leq j \leq k'$ a *switch* if either $\gamma^{(i,j-1)} \neq \beta^{(i,j-1)}$ and $\gamma^{(i,j)} = \beta^{(i,j)}$ or if $\gamma^{(i,j-1)} = \beta^{(i,j-1)}$ and $\gamma^{(i,j)} \neq \beta^{(i,j)}$. In the first case we call it an *agree switch* and in the latter case we call it a *disagree switch*. It is clear from this definition that the sequence of switches for any $\gamma^{(i)}$ in $\mathcal{A}_V^{(i)}(\beta^{(i)})$ must alternate between agree switch and disagree switch. We also know that each member of $\mathcal{A}_V^{(i)}(\beta^{(i)})$ has at least one agree switch (by definition).

We partition the set $\mathcal{A}_V^{(i)}(\beta^{(i)})$ according the the number of switch coordinates of its members. Let $\mathcal{A}_{V,t}^{(i)}(\beta^{(i)})$ be the set of all $\gamma^{(i)} \in \mathcal{A}_V^{(i)}(\beta^{(i)})$ containing exactly t switches.

Thus, to specify an element of $\mathcal{A}_{V,t}^{(i)}(\beta^{(i)})$ one needs to specify the locations $S_t \subseteq [k']$ ($|S| = t$) of its switch coordinates, and whether the first switch is an agree switch or

a disagree switch, which can be specified by a bit $b \in \{0,1\}$. Once this information is known, this fully determines the set of coordinates j for which $\gamma^{(i,j)} \neq \beta^{(i,j)}$. Let $\mathsf{Dis}_{S_t,b}$ be this set of coordinates - we call these the disagreeing coordinates. For each one of these coordinates j in $\mathsf{Dis}_{S_t,b}$, one needs to specify the value of $\gamma^{(i,j)}$.

Given the values of all coordinates before the jth coordinate, the value of $\gamma^{(i,j)}$ can be one of only $\mathsf{Deg}(X^{(i,j)})$ many choices, as it is determined by the *label* of the outgoing edge in the graph of $X^{(i,j)}$. Thus, once $\mathsf{Dis}_{S_t,b}$ is determined , if $\mathsf{Dis}_{S_t,b} = \{t_1, t_2, \ldots, t_s\} \subseteq [k']$ is the set of disagreeing coordinates, let $L(\mathsf{Dis}_{S_t,b}) = \{(a_{t_1}, a_{t_2}, \ldots, a_{t_s}) : a_{t_j} \in [\mathsf{Deg}(X^{(i,t_j)})]\}$ be set of labels of edges the disagreeing coordinates could correspond to. Thus every $\gamma^{(i)}$ corresponding to the set $\mathsf{Dis}_{S_t,b}$ of disagreeing coordinates would also correspond to some element of $L(\mathsf{Dis}_{S_t,b})$.

Thus the maximum number of possible choices for $\gamma^{(i)} \in \mathcal{A}^{(i)}_{V,t}(\beta^{(i)})$ is at most the number of ways of choosing the set $\mathsf{Dis}_{S_t,b}$, which is $\binom{k'}{t} \cdot 2$, multiplied by $\prod_{j \in T} \mathsf{Deg}(X^{(i,j)})$.

However, not every element of $L(\mathsf{Dis}_{S_t,b})$ would correspond to a choice of $\gamma^{(i)} \in \mathcal{A}^{(i)}_{V,t}(\beta^{(i)})$. The reason being that when a disagreeing coordinate appears right before an *agree switch*, the only way there can be an "agree" after a "disagree" is that the endpoint of a disagreeing edge coincides with the start point of an agree edge in the corresponding layered graph. However, for every edge label of the disagreeing edge, the end point was chosen to be a uniformly random element of $\tilde{n}$ in the distribution $\mathcal{D}$. Thus this event happens only with probability exactly $1/\tilde{n}$ for $V \leftarrow \mathcal{D}$, and this is independent for each agree switch. Thus for every fixing of $\mathsf{Dis}_{S_t,b}$ coordinates corresponding to the disagreeing coordinates, and every sequence $s_t \in L(\mathsf{Dis}_{S_t,b})$, the probability that the sequence corresponds to a $\gamma^{(i)} \in \mathcal{A}^{(i)}_{V,t}(\beta^{(i)})$ is at most the probability that for each agree switch, the endpoint of a disagreeing edge coincides with the start point of an agree edge. For each agree switch this happens independently with probability $1/\tilde{n}$. Recall that the number of agree switches is at least $\max\{1, (t-1)/2\}$.

Let $\mathcal{A}^{(i)}_{V,t,T}(\beta^{(i)})$ be the set of all $\gamma^{(i)} \in \mathcal{A}^{(i)}_{V,t}(\beta^{(i)})$ containing exactly t switches and such that T is the set of disagreeing coordinates.

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}} \left[|\mathcal{A}^{(i)}_{V,t,T}(\beta^{(i)})| \right] \leq \prod_{j \in T} \mathsf{Deg}(X^{(i,j)}) \cdot \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}}.$$

Before the final computation, we need the following simple lemma:

Lemma 4.48. $\forall i \in [r']$, $\forall T \subseteq [k']$, $\left(\prod_{j\in T} \mathsf{Deg}(X^{(i,j)})\right) \cdot D^{-|T|} \leq n^2$.

Proof. Observe that since $1 < D < 2$, thus for all j such that $1 \leq j \leq k' - 2\log n$, we have that $\mathsf{Deg}(X^{(i,j)}) > D$, and for all j such that $k'-2\log n < j \leq k'$, $\mathsf{Deg}(X^{(i,j)}) < D$. Thus the expression $\left(\prod_{j\in T} \mathsf{Deg}(X^{(i,j)})\right) \cdot D^{-|T|}$ is maximized for $T = [k' - 2\log n]$, and for this choice of T, $\prod_{j\in T} \mathsf{Deg}(X^{(i,j)}) = D^{k'}$ and $D^{|T|} = \frac{D^{k'}}{D^{2\log n}}$. Thus $\prod_{j\in T} \mathsf{Deg}(X^{(i,j)}) \cdot D^{-|T|} \leq D^{2\log n} \leq n^2$. □

Thus

$$\begin{aligned}
\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}[\sum_{\gamma^{(i)}\in\mathcal{A}^{(i)}_{V,t,T}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}] &\leq \textstyle\prod_{j\in T} \mathsf{Deg}(X^{(i,j)}) \cdot \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}} \cdot D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \\
&= \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}} \cdot \left(\textstyle\prod_{j\in T} \mathsf{Deg}(X^{(i,j)})\right) \cdot D^{-|T|} \\
&\leq \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}} \cdot n^2. \qquad \text{(by Lemma 4.48)}
\end{aligned}$$

Now, given t, there are at most $2 \cdot \binom{k'}{t}$ ways of choosing the set T. Thus $\mathcal{A}^{(i)}_{V,t}(\beta^{(i)})$ can be written as a union of at most $2 \cdot \binom{k'}{t}$ different sets of the form $\mathcal{A}^{(i)}_{V,t,T}(\beta)$. Thus

$$\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}\left[\sum_{\gamma^{(i)}\in\mathcal{A}^{(i)}_{V,t}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}\right] \leq \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}} \cdot n^2 \cdot 2 \cdot \binom{k'}{t}.$$

Summing over the various choices of t, we get that

$$\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}\left[\sum_{\gamma^{(i)}\in\mathcal{A}^{(i)}_{V}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}\right] \leq \sum_{t=1}^{k'} \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}} \cdot n^2 \cdot 2 \cdot \binom{k'}{t}.$$

Since $\tilde{n} = n^5$ and $k' = O(\sqrt{n})$, it is easily verified that

$$\mathbb{E}[\sum_{\gamma^{(i)}\in\mathcal{A}^{(i)}_{V}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}] \leq O(1/n).$$

□

We now give a proof of Proposition 4.44.

Proof of Proposition 4.44. For all $\beta \in \overline{IMM}_X$, observe that

$$\sum_{\gamma \in (\overline{IMM}|_V)(\alpha)} D^{-\Delta(\beta,\gamma)} = \prod_{i \in [r']} \sum_{\gamma^{(i)} \in (\overline{IMM}|_V)(\alpha)^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}.$$

Moreover, since the choice of $V \leftarrow \mathcal{D}$ chooses variables in distinct matrices independently, thus

$$\begin{aligned} \mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}} \left[\sum_{\gamma \in (\overline{IMM}|_V)(\alpha)} D^{-\Delta(\beta,\gamma)} \right] &= \prod_{i \in [r']} \mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}} \left[\sum_{\gamma^{(i)} \in (\overline{IMM}|_V)(\alpha)^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right] \\ &\leq (O(1))^{r'} \\ &\leq n^{o(r)} \end{aligned}$$

Here the second to last inequality follows from Lemma 4.45, and the last inequality follows from the fact that $r' = O(r)$.

□

4.9.2 Expected value of $T_1(IMM|_V)$

We now prove Lemma 4.36.

Proof of Lemma 4.36. For all $\alpha \in \mathcal{T}(IMM|_V)$,

$$\begin{aligned} T_1|_V(\alpha) &= \textstyle\sum_{\beta \in \overline{IMM}|_V(\alpha)} |S_m(\alpha,\beta)| \\ &= \textstyle\sum_{\beta \in \overline{IMM}|_V(\alpha)} \binom{N-k}{m} \\ &= D^k \cdot \textstyle\binom{N-k}{m} \end{aligned}$$

□

4.9.3 Expected value of $T_2(IMM|_V)$

Let $V \leftarrow \mathcal{D}$. Recall that

$$T_2|_V = \sum_{\alpha \in \mathcal{T}(IMM|_V)} \left(\sum_{\substack{\beta,\gamma \in \overline{IMM}|_V(\alpha) \\ \beta \neq \gamma}} \binom{N-k-\Delta(\beta,\gamma)}{m} \right).$$

For $\alpha \in (\overline{IMM}|_V)_Y$ and $\beta \in \overline{IMM}|_V(\alpha)$, let

$$T_2|_V(\alpha,\beta) = \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} \binom{N-k-\Delta(\beta,\gamma)}{m}.$$

For $\alpha \notin (\overline{IMM}|_V)_Y$ or $\beta \notin \overline{IMM}|_V(\alpha)$, let $T_2|_V(\alpha,\beta) = 0$. For every fixed $\alpha \in (\overline{IMM}|_V)_Y$ and $\beta \in \overline{IMM}|_V(\alpha)$, $T_2|_V(\alpha,\beta)$ counts for every $\gamma \in \overline{IMM}|_V(\alpha)$ such that $\gamma \neq \beta$, the number of multilinear shifts of degree m that are disjoint from both β and γ. It then takes the sum of this quantity over all $\gamma \in \overline{IMM}|_V(\alpha)$. We now prove Lemma 4.37 and Lemma 4.39. In order to do so, we first bound $\mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V(\alpha,\beta)]$, and then sum over α and β as appropriate to obtain Lemma 4.37 and Lemma 4.39.

Lemma 4.49. *For $\alpha \in \overline{IMM}_Y$ and $\beta \in \overline{IMM}(\alpha)$,*

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}}[T_2|_V(\alpha,\beta)] \leq \binom{N-k}{m} \cdot n^{o(r)}.$$

Proof.

$$\begin{aligned}
T_2|_V(\alpha,\beta) &= \textstyle\sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} \binom{N-k-\Delta(\beta,\gamma)}{m} \\
&= \textstyle\sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} \binom{N-k-\Delta(\beta,\gamma)}{m} \cdot \frac{\binom{N-k}{m}}{\binom{N-k}{m}} \\
&= \textstyle\binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} \frac{\binom{N-k-\Delta(\beta,\gamma)}{m}}{\binom{N-k}{m}} \\
&\approx \textstyle\binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} \left(\frac{N-m}{N}\right)^{\Delta(\beta,\gamma)} \qquad \text{by Lemma 2.13} \\
&\leq \textstyle\binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} D^{-\Delta(\beta,\gamma)}
\end{aligned}$$

Thus,

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}}[T_2|_V(\alpha,\beta)] \leq \binom{N-k}{m} \cdot \mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}}\left[\sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} D^{-\Delta(\beta,\gamma)}\right] \leq \binom{N-k}{m} \cdot n^{o(r)},$$

where the second inequality follows from Proposition 4.44. □

Proof of Lemma 4.39. $\forall \alpha \in (\overline{IMM}|_V)_Y$,

$$\begin{aligned}
\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}}[T_2|_V(\alpha)] &\leq \textstyle\sum_{\beta \in \overline{IMM}|_V(\alpha)} \mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V(\alpha,\beta)] \\
&= D^k \cdot \textstyle\binom{N-k}{m} \cdot n^{o(r)}.
\end{aligned}$$

□

Proof of Lemma 4.37.

$$\begin{aligned}\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}[T_2|_V] &= \textstyle\sum_{\alpha\in\mathcal{T}(IMM|_V)} \mathbb{E}_{V\leftarrow\mathcal{D}}[T_2|_V(\alpha)] \\ &\leq \textstyle\sum_{\alpha\in\mathcal{T}(IMM|_V)} D^k \cdot \binom{N-k}{m} \cdot n^{o(r)} \\ &= n^r \cdot D^k \cdot \binom{N-k}{m} \cdot n^{o(r)}\end{aligned}$$

□

4.9.4 Expected value of $T_3(IMM|_V)$

We now prove Lemma 4.38.

Proof of Lemma 4.38. Let $V \leftarrow \mathcal{D}$. Let

$$T_3^{=}|_V = \sum_{\alpha\in\mathcal{T}(IMM|_V)} \left(\sum_{\substack{\beta,\gamma\in\overline{IMM}|_V(\alpha)\\ \beta\neq\gamma}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \right)$$

Let

$$T_3^{\neq}|_V = \sum_{\substack{\alpha,\alpha'\in\mathcal{T}(IMM|_V)\\ \alpha\neq\alpha'}} \left(\sum_{\substack{\beta\in\overline{IMM}|_V(\alpha)\\ \gamma\in\overline{IMM}|_V(\alpha')}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \right)$$

Observe that

$$T_3|_V = T_3^{=}|_V + T_3^{\neq}|_V$$

For $\alpha \in \overline{IMM}_Y$, let

$$T_3^{=}|_V(\alpha) = \sum_{\substack{\beta,\gamma\in\overline{IMM}|_V(\alpha)\\ \beta\neq\gamma}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \tag{4.50}$$

For $\alpha, \alpha' \in \overline{IMM}_Y$ such that $\alpha \neq \alpha'$, let

$$T_3^{\neq}|_V(\alpha,\alpha') = \sum_{\substack{\beta\in\overline{IMM}|_V(\alpha)\\ \gamma\in\overline{IMM}|_V(\alpha')}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \tag{4.51}$$

For every α and α', $T_3^{\neq}|_V(\alpha,\alpha')$ counts for every β extending α and γ extending α', the number of pairs of multilinear shifts m_β and m_γ, each of degree m, such that m_β is disjoint from β, m_γ is disjoint from γ, and $\beta \cdot m_\beta = \gamma \cdot m_\gamma$. Consider

$$\begin{aligned}\binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} &= \tbinom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \cdot \frac{\binom{N-k}{m}}{\binom{N-k}{m}} \\ &= \tbinom{N-k}{m} \cdot \frac{\binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)}}{\binom{N-k}{m}}\end{aligned}$$

Now by an application of Lemma 2.13, we obtain

$$\binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \approx \binom{N-k}{m} \cdot \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)} \tag{4.52}$$

Since by our choice of parameters $D < N/m$, plugging back Equation Equation 4.52 into Equation Equation 4.50, we obtain

$$\begin{aligned}T_3^{=}|_V(\alpha) &\approx \binom{N-k}{m} \cdot \sum_{\substack{\beta,\gamma \in \overline{IMM}|_V(\alpha) \\ \beta \neq \gamma}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)} \\ &\leq \binom{N-k}{m} \cdot \sum_{\beta \in \overline{IMM}|_V(\alpha)} \left(\sum_{\gamma \in \overline{IMM}|_V(\alpha), \gamma \neq \beta} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)} \right) \\ &\leq \binom{N-k}{m} \cdot \sum_{\beta \in \overline{IMM}|_V(\alpha)} \left(\sum_{\gamma \in \overline{IMM}|_V(\alpha), \gamma \neq \beta} (D)^{-\Delta(\beta,\gamma)} \right) \\ &\leq \binom{N-k}{m} \cdot D^k \cdot \sum_{\gamma \in \overline{IMM}|_V(\alpha)} (D)^{-\Delta(\beta,\gamma)}\end{aligned}$$

Now, applying Proposition 4.44, we obtain

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}} [T_3^{=}|_V(\alpha)] \leq \binom{N-k}{m} \cdot D^k \cdot n^{o(r)}.$$

and hence

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}} [T_3^{=}|_V] \leq n^r \cdot \binom{N-k}{m} \cdot D^k \cdot n^{o(r)}. \tag{4.53}$$

Thus, it remains to bound $\mathbb{E}_{V \leftarrow \mathcal{D}}\left[T_3^{\neq}|_V\right]$. For $\alpha, \alpha' \in \overline{IMM}_Y$ such that $\alpha \neq \alpha'$, consider

$$T_3^{\neq}|_V(\alpha,\alpha') = \sum_{\substack{\beta \in \overline{IMM}|_V(\alpha) \\ \gamma \in \overline{IMM}|_V(\alpha')}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)}.$$

For $\beta \in \overline{IMM}|_V(\alpha)$, let

$$T_3^{\neq}|_V(\alpha,\alpha',\beta) = \textstyle\sum_{\gamma \in \overline{IMM}|_V(\alpha')} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)}$$

Now by an application of Equation Equation 4.52, it follows that

$$T_3^{\neq}|_V(\alpha,\alpha',\beta) \approx \binom{N-k}{m} \cdot \sum_{\gamma \in \overline{IMM}|_V(\alpha')} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)}$$

Let $\varepsilon' = 2/\Lambda$ be a constant. We now partition the sum over γ into two parts, depending on whether $\Delta(\beta,\gamma) \geq (1-\varepsilon')k$ or whether $\Delta(\beta,\gamma) < (1-\varepsilon')k$. For $\alpha, \alpha' \in \mathcal{T}(IMM|_V)$ such that $\alpha \neq \alpha'$, and for $\beta \in \overline{IMM}|_V(\alpha)$, let

$$T_{3_{\text{large}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta) = \binom{N-k}{m} \cdot \left(\sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) \geq (1-\varepsilon')k}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)} \right)$$

and

$$T_{3_{\text{small}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta) = \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\varepsilon')k}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)}$$

Thus

$$\begin{aligned}
T_{3_{\text{large}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta) &\leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) \geq (1-\varepsilon')k}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)} \\
&= \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) \geq (1-\varepsilon')k}} \left(\frac{N-m}{N}\right)^{\Delta(\beta,\gamma)} \cdot \left(\frac{m}{N-m}\right)^{\Delta(\beta,\gamma)} \\
&\leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) \geq (1-\varepsilon')k}} \left(\frac{N-m}{N}\right)^{\Delta(\beta,\gamma)} \cdot \left(\frac{m}{N-m}\right)^{(1-\varepsilon')k}
\end{aligned}$$

Here, the last step follows since $\frac{m}{N-m} < 1$.

Now, by our choice of parameters, $\left(\frac{m}{N-m}\right)^k = n^{-r}$ and $D = \frac{N}{N-m}$, we get

$$T_{3_{\text{large}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta) \leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) \geq (1-\varepsilon')k}} D^{-\Delta(\beta,\gamma)} \cdot n^{-(1-\varepsilon')r}$$

From here, by applying Proposition 4.44, we obtain

$$\begin{aligned}
\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}} \left[T_{3_{\text{large}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta) \right] &\leq \binom{N-k}{m} \cdot n^{o(r)} \cdot n^{-(1-\varepsilon')r} && (4.54) \\
&\leq \binom{N-k}{m} \cdot O(n^{(2\varepsilon'-1)r}) && (4.55)
\end{aligned}$$

We will now bound

$$T^{\neq}_{3_{\text{small}\Delta}}|_V(\alpha,\alpha'\beta) = \binom{N-k}{m} \cdot \sum_{\substack{\gamma\in\overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta)<(1-\varepsilon')k}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)}$$

Recall that for $\alpha,\alpha' \in \mathcal{T}$ such that $\alpha \neq \alpha'$, $\Delta(\alpha,\alpha') \geq r'-r$. For $\alpha,\alpha' \in \mathcal{T}(IMM|_V)$ such that $\alpha \neq \alpha'$ and for $\beta \in \overline{IMM}|_V(\alpha)$,

$$\begin{aligned}
T^{\neq}_{3_{\text{small}\Delta}}|_V(\alpha,\alpha'\beta) &\leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma\in\overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta)<(1-\varepsilon')k}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)} \\
&= \binom{N-k}{m} \cdot \sum_{\substack{\gamma\in\overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta)<(1-\varepsilon')k}} \left(\frac{N-m}{N}\right)^{\Delta(\beta,\gamma)} \cdot \left(\frac{m}{N-m}\right)^{\Delta(\beta,\gamma)} \\
&\leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma\in\overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta)<(1-\varepsilon')k}} \left(\frac{N-m}{N}\right)^{\Delta(\beta,\gamma)} \qquad (\text{since } \frac{m}{N-m} < 1) \\
&= \binom{N-k}{m} \cdot \sum_{\substack{\gamma\in\overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta)<(1-\varepsilon')k}} D^{-\Delta(\beta,\gamma)}
\end{aligned}$$

Now, any $\gamma \in \overline{IMM}_X$ can be expressed as $\prod_{i\in[r']} \gamma^{(i)}$, and $D^{-\Delta(\beta,\gamma)} = \prod_{i\in[r']} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}$. We will partition the set $[r']$ according to the number of "agreements" of $\gamma^{(i)}$ and $\beta^{(i)}$.

Let $A(\beta,\gamma) \subseteq [r']$ be the set of all i such that $\Delta(\beta^{(i)},\gamma^{(i)}) < k'$ (i.e. there is some $j \in [k']$ such that $\beta^{(i,j)} = \gamma^{(i,j)}$). Since $\Delta(\gamma,\beta) < (1-\varepsilon')k = (1-\varepsilon')k'r'$, thus $|A(\beta,\gamma)| \geq \varepsilon' r'$. Also, let $B(\alpha,\alpha') \subseteq [r']$ be the set of all $i \in [r']$ such that $\alpha^{(i)} = \alpha'^{(i)}$. Then by Lemma 4.34, for $\alpha \neq \alpha'$, $|B(\alpha,\alpha')| \leq r$.

Claim 4.56. *Let $\alpha,\alpha' \in \mathcal{T}(IMM|_V)$ be such that $\alpha \neq \alpha'$, and let $\beta \in \overline{IMM}|_V(\alpha)$ and $\gamma \in \overline{IMM}|_V(\alpha')$ be such that $\Delta(\beta,\gamma) < (1-\varepsilon')k$. Then for any $i \in A(\beta,\gamma)\setminus B(\alpha,\alpha')$, it holds that $\Delta(\beta^{(i)},\gamma^{(i)}) < k'$, and moreover $\beta^{(i,1)} \neq \gamma^{(i,1)}$. Moreover $|A(\beta,\gamma)\setminus B(\alpha,\alpha')| \geq \varepsilon' r' - r$.*

Proof. The only tricky part is to show that $\beta^{(i,1)} \neq \gamma^{(i,1)}$, and we give a proof of this below. If $\alpha^{(i)} \neq \alpha'^{(i)}$, then this means that the variable in α corresponding to $Y^{(i)}|_V$, is distinct from the variable in α' corresponding to $Y^{(i)}|_V$. Any variable in $Y^{(i)}|_V$ is of the form $y^{(i)}_{1,s}$ for some $s \in [\tilde{n}]$. Suppose that $\alpha^{(i)} = y^{(i)}_{1,s}$ and $\alpha'^{(i)} = y^{(i)}_{1,s'}$, for $s \neq s'$. Then

for $\beta \in \overline{IMM}|_V(\alpha)$, $\beta^{(i,1)}$ is a variable from $X^{(i,1)}$ and must be of the form $x^{(i,1)}_{s,t}$ for some $t \in [\tilde{n}]$ and for $\gamma \in \overline{IMM}|_V(\alpha)$, $\gamma^{(i,1)}$ must be of the form $x^{(i,1)}_{s',t'}$ for some $t' \in [\tilde{n}]$. Since $s \neq s'$, thus $\beta^{(i,1)} \neq \gamma^{(i,1)}$. □

Now for every subset $C \subseteq [r']$ such that $|C| = \varepsilon' r' - r$, Let $M_C(\beta, \alpha')$ be the set of all $\gamma \in \overline{IMM}|_V(\alpha')$ such that for all $i \in C$, $\Delta(\beta^{(i)}, \gamma^{(i)}) < k'$ and $\beta^{(i,1)} \neq \gamma^{(i,1)}$. Thus for every $\alpha, \alpha' \in \mathcal{T}(IMM|_V)$ such that $\alpha \neq \alpha'$, and for every $\beta \in \overline{IMM}|_V(\alpha)$, every $\gamma \in \overline{IMM}|_V(\alpha')$ such that $\Delta(\beta, \gamma) < (1-\varepsilon')k$ gets counted in at least one such set $M_C(\beta, \alpha')$ for some choice of C.

Let $M_C(\beta, \alpha')^{(i)}$ be the set of all $\gamma^{(i)} \in \overline{IMM}|_V(\alpha')^{(i)}$ such that if $i \in C$, then $\Delta(\beta^{(i)}, \gamma^{(i)}) < k'$ and $\beta^{(i,1)} \neq \gamma^{(i,1)}$. If $i \notin C$ then there is no such restriction between β and γ. Thus it is easy to see that $M_C(\beta, \alpha') \subseteq \prod_{i \in [r']} M_C(\beta, \alpha')^{(i)}$.

Now, fixing $\alpha, \alpha' \in \mathcal{T}(IMM|_V)$ such that $\alpha \neq \alpha'$, and $\beta \in \overline{IMM}|_V(\alpha)$, we get that

$$
\begin{aligned}
T^{\neq}_{3_{\text{small}\Delta}}|_V(\alpha, \alpha'\beta) &\leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\varepsilon')k}} D^{-\Delta(\beta,\gamma)} \\
&= \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\varepsilon')k}} \prod_{i \in [r']} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \\
&\leq \binom{N-k}{m} \\
&\cdot \sum_{\substack{C \subset [r'], \\ |C| = \varepsilon' r' - r}} \sum_{\gamma \in M_C(\beta,\alpha')} \left(\prod_{i \in C} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \cdot \prod_{i \in [r'] \setminus C} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right) \\
&\leq \binom{N-k}{m} \cdot \sum_{\substack{C \subset [r'], \\ |C| = \varepsilon' r' - r}} \prod_{i \in C} \left(\sum_{\gamma^{(i)} \in M_C(\beta,\alpha')^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right) \\
&\cdot \prod_{i \in [r'] \setminus C} \left(\sum_{\gamma^{(i)} \in M_C(\beta,\alpha')^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right)
\end{aligned}
$$

Now, observe that $i \in C$, $M_C(\beta, \alpha')^{(i)} \subseteq \mathcal{A}^{(i)}_V(\beta^{(i)})$. Thus, by Lemma 4.47 and

Lemma 4.45, we get that

$$\begin{aligned}\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}\left[T^{\neq}_{3_{\mathrm{small}\Delta}}|_V(\alpha,\alpha'\beta)\right] &\leq \binom{N-k}{m}\cdot\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}\left[\sum_{\substack{\gamma\in\overline{IMM}|_V(\alpha'),\\ \Delta(\gamma,\beta)<(1-\varepsilon')k}} D^{-\Delta(\beta,\gamma)}\right]\\ &\leq \binom{N-k}{m}\cdot\Psi\end{aligned}$$

Here,

$$\begin{aligned}\Psi &= \sum_{\substack{C\subset[r'],\\ |C|=\varepsilon' r'-r}}\prod_{i\in C}\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}\left[\left(\sum_{\gamma^{(i)}\in M_C(\beta,\alpha')^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}\right)\right]\\ &\times \prod_{i\in[r']\setminus C}\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}\left[\left(\sum_{\gamma^{(i)}\in M_C(\beta,\alpha')^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}\right)\right]\end{aligned}$$

Therefore,

$$\begin{aligned}\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}\left[T^{\neq}_{3_{\mathrm{small}\Delta}}|_V(\alpha,\alpha'\beta)\right] &\leq \binom{N-k}{m}\cdot\binom{r'}{\varepsilon' r'-r}\cdot\left(O\left(\frac{1}{n}\right)\right)^{\varepsilon' r'-r}2^{O(r')}\\ &= \binom{N-k}{m}\cdot\left(O\left(\frac{1}{n}\right)\right)^{\varepsilon' r'-r}\cdot n^{o(r)}\end{aligned}$$

Thus since $\varepsilon' r'-r>r$,

$$\mathbb{E}[T^{\neq}_{3_{\mathrm{small}\Delta}}(\alpha,\alpha'\beta)]\leq\binom{N-k}{m}\cdot\left(\frac{1}{n}\right)^{\varepsilon' r'-r}\cdot n^{o(r)}\leq\binom{N-k}{m}\cdot n^{-r+o(r)}.$$

Putting this together with earlier computation showing that

$$\mathbb{E}[T^{\neq}_{3_{\mathrm{large}\Delta}}(\alpha,\alpha'\beta)]\leq\binom{N-k}{m}\cdot O(n^{(2\varepsilon'-1)r}),$$

we conclude that

$$\mathbb{E}[T^{\neq}_{3}(\alpha,\alpha'\beta)]\leq\binom{N-k}{m}\cdot O(n^{(2\varepsilon'-1)r}).$$

Summing over $\beta\in\overline{IMM}|_V(\alpha)$, we get that

$$\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}\left[T^{\neq}_{3}|_V(\alpha,\alpha')\right]\leq\binom{N-k}{m}\cdot D^k\cdot n^{(2\varepsilon'-1)r}\cdot O(1).$$

Summing over $\alpha,\alpha'\in\mathcal{T}(IMM|_V)$ such that $\alpha\neq\alpha'$, we get that

$$\mathop{\mathbb{E}}_{V\leftarrow\mathcal{D}}\left[T^{\neq}_{3}|_V\right]\leq n^{2r}\cdot\binom{N-k}{m}\cdot D^k\cdot n^{(2\varepsilon'-1)r}=n^r\cdot\binom{N-k}{m}\cdot D^k\cdot n^{(2\varepsilon')r}\cdot O(1).$$

Putting this together with the bound in Equation Equation 4.53, we conclude that

$$\mathop{\mathbb{E}}_{V \leftarrow \mathcal{D}}[T_3|_V] \leq n^r \cdot \binom{N-k}{m} \cdot D^k \cdot n^{\frac{4}{\Lambda}r} \cdot O(1).$$

□

Chapter 5

Finer separations between shallow arithmetic circuits[1]

5.1 Introduction

In the results in Chapter 4, we showed a lower bound of $n^{\Omega(\sqrt{d})}$ for a polynomial that has a poly(n)-sized arithmetic circuit. This implies that, in general, the size bound of $n^{O(\sqrt{d})}$ can not be improved to $n^{o(\sqrt{d})}$ for poly(n)-sized arithmetic circuits. However, as far as we know, it was not known if improved depth reductions (depth reductions to homogeneous depth-4 circuits of size $n^{o(\sqrt{d})}$) are conceivable for slightly restricted classes of arithmetic circuits, for instance, arithmetic formulas or constant depth arithmetic circuits.

In this chapter, we study this problem and show that at least for the case when $d = O(\log^2 n)$, one cannot hope to prove such improved depth reduction results, for even extremely restricted classes of arithmetic circuits such as linear size homogeneous depth-5 arithmetic circuits, or polynomial sized non-homogeneous depth-3 arithmetic circuits.

We now state our results, and elaborate on how they compare to the known results.

5.1.1 Our results

We prove the following theorems.

Theorem 5.1. *Let $\mathbb{F}$ be any field. There is a family of polynomials $\{P_n\}$ over $\mathbb{F}$, where P_n is of degree $d = O(\log^2 n)$ on n variables such that P_n can be computed by a homogeneous depth-5 circuit of size n whereas any homogeneous depth-4 circuit computing P_n requires size $n^{\Omega(\sqrt{d})}$.*

[1]The results in this chapter appear in [KS16].

Theorem 5.2. *Let $\mathbb{F}$ be any field of characteristic zero. There is a family of polynomials $\{P_n\}$ over $\mathbb{F}$, where P_n is of degree $d = O(\log^2 n)$ on n variables such that P_n can be computed by a (non-homogeneous) depth-3 circuit of size* $\mathrm{poly}(n)$ *whereas any homogeneous depth-4 circuit computing P_n requires size $n^{\Omega(\sqrt{d})}$.*

5.1.2 Comparison to earlier results

An $n^{\Omega(\sqrt{d})}$ lower bound for homogeneous depth-4 circuits was proved for an explicit polynomial of degree d in n variables in VNP by Kayal, Limaye, Saha and Srinivasan [KLSS14a] and for the iterated matrix product (IMM) by Kumar and Saraf [KS17]. Improvements on this can happen on three fronts – (1) by improving the bound from $n^{\Omega(\sqrt{d})}$ to $n^{\omega(\sqrt{d})}$, or (2) by making the lower bound work for a class more general than homogeneous depth-4 circuits, or (3) by proving the lower bound for a polynomial "simpler" than IMM. This work is of the last category where the polynomial is computed by linear sized homogeneous depth-5 circuits or polynomial sized depth-3 circuits.

We elaborate more on this now.

Depth reduction to depth-4 as a springboard for stronger lower bounds

Let $\mathcal{C}$ be a class of arithmetic circuits. If we had a depth reduction result that showed that all homogeneous polynomials of degree d in n variables that can be computed by an arithmetic circuit $C \in \mathcal{C}$ of size $s(n)$ can also be computed by a homogeneous depth-4 arithmetic circuit of size $s^{o(\sqrt{d})}$, then it follows from the results in [KLSS14a, KS17] that there is an explicit polynomial in VP (or VNP) that cannot be computed by polynomial size arithmetic circuits in $\mathcal{C}$. In this sense, the *efficient* reductions to homogeneous depth-4 circuits is a *springboard* to prove lower bounds for many potentially stronger classes of circuits.

The lower bound for IMM in [KS17] rules out this strategy when $\mathcal{C}$ is the class of algebraic branching programs, since it shows polynomial families (namely IMM) that have linear size ABPs but require homogeneous depth-4 circuits of size $n^{\Omega(\sqrt{d})}$. However the strategy could still, in principle, work for other interesting classes of arithmetic circuits such as arithmetic formulas, constant depth arithmetic circuits or, possibly the simplest

of them all, the class of homogeneous depth-5 arithmetic circuits. Another simple class of circuits for which this strategy could be tried is the class of non-homogeneous depth-3 circuits, where superpolynomial lower bounds are not known when the size of the underlying field is large. Theorem 5.1 and Theorem 5.2 show that the above mentioned classes of arithmetic circuits cannot be reduced to homogeneous depth-4 arithmetic circuits of size $n^{o(\sqrt{d})}$, albeit for an appropriate range of parameters. So, even though quantitatively we do not prove improved lower bounds, qualitatively, we show near optimal separations between complexity classes which are much closer to each other that was earlier known. Unfortunately, we are only able to show such separations when the degree $d = O(\log^2 n)$.

Non-homogeneous depth-3 circuits

Theorem 5.2 shows a separation between non-homogeneous depth-3 circuits and homogeneous depth-4 circuits, in a low degree regime. Intuitively, to prove such a separation, we need a candidate family of hard polynomials which have polynomial sized non-homogeneous depth-3 circuits and are believed to require homogeneous depth-4 circuits of size $n^{\Omega(\sqrt{d})}$. At first glance, it seems unclear what this polynomial should be. The elementary symmetric polynomial of degree d is *not* a good candidate as it can indeed be computed by a homogeneous depth four circuit of size $2^{O(\sqrt{d})}$ [HY11]. However, *a generic affine projection* of the elementary symmetric polynomial, as studied by Shpilka [Shp02], is a natural candidate and is almost complete for this model.

In this chapter, however, we do not directly work with this polynomial but it can be easily inferred that the lower bound applies to a generic affine projection of the elementary symmetric polynomial as well.

Depth hierarchy theorems for arithmetic circuits

Depth hierarchy theorems, which show an exponential, (and near optimal) separation between depth h and depth $h + 1$ circuits [Hås86, RST15] constitute some of the most celebrated results in the theory of lower bounds for bounded depth boolean circuits.

It is natural to ask if such separations can be shown for arithmetic circuits. Unfortunately, superpolynomial lower bounds are not known in general when the depth of the arithmetic circuits is more than four [2]. So, at this point, we can only hope to show such separations between homogeneous depth-5 and homogeneous depth-4 arithmetic circuits. Due to the depth reduction results, the best such separation one can hope to prove for an n variate degree d polynomial would be $n^{\Omega(\sqrt{d})}$. We prove a matching lower bound, as long as the degree d is at most $O(\log^2 n)$. In the arithmetic circuit literature, the question of depth hierarchy theorems has previously been studied by Raz and Yehudayoff [RY09], where they show superpolynomial separation separation between multilinear circuits of product depth d and product depth $d+1$, for $d = O(1)$. In the non-multilinear world, to the best of our knowledge this is the first such attempt. Even in the context of constant depth multilinear circuits, the separation in [RY09] is between depth-4 and depth-6 circuits, and not between depth-4 and depth-5 circuits.

The complexity measure

The proof of Kayal et al. [KLSS14a] and Kumar and Saraf [KS17] rely on the notion of projected shifted partials of a polynomial as a measure of its complexity. This measure can be thought of as a variant of shifted partials which tries to take advantage of the fact that the hard polynomial is multilinear. The measure in this chapter takes advantage of *set-multilinearity* instead of just multilinearity, and such a variant was essentially used in an earlier version of [KLSS14a], where they showed an $n^{O(\log n)}$ lower bound for iterated matrix multiplication and the determinant. Our proofs rely on a slightly different interpretation of the measure, which makes the proofs much more transparent. Intuitively, this measure tries to take advantage of the fact that the hard polynomial (Nisan-Wigderson design polynomials or the IMM) is not just multilinear, but in fact set-multilinear. In the regime where $d \ll n$, set multilinearity is a much more rigid restriction on a polynomial when compared to multilinearity, and in some sense our gain comes from this observation. Our hard polynomial for Theorem 5.1 is also a

[2] For homogeneous depth-5 circuits, such lower bounds are known only over small finite fields.

simple generic balanced depth-5 circuit.

One might wonder if the results in this chapter could have been shown by using the dimension of the projected shifted partial derivatives as the complexity measure. In particular, can we show that the projected shifted partials complexity of a generic depth-5 circuit is sufficiently close to the largest possible value? This would suffice for Theorem 5.1. Although we do not have enough evidence to conjecture one way or the other, intuitively this problem seems tricky since so far the known analyses of the projected shifted partials of a polynomial seems to rely on pairwise distance between the monomials of the hard polynomial, either in the worst case (Nisan-Wigderson polynomial [KLSS14a, KS17]), or in the average case (IMM [KS17]). Clearly, the monomials in a generic depth-5 circuit do not have good distance in the worst case, and to the best of our understanding, the guarantees about distance in the average case seem a bit weaker than what would suffice to simulate the proof in [KS17] for a generic depth-5 circuit. However, this problem of proving lower bounds on the dimension of projected shifted partials of homogeneous depth-5 circuits is of independent interest, since even if the answer is negative and homogeneous depth-5 circuits do not have large enough projected shifted partials complexity, then we could use this as a measure to prove lower bounds for such circuits. So far, such lower bounds are only known over small finite fields [KS15c].

5.2 Preliminaries

5.2.1 Notations

- Throughout the chapter, we use bold-face letters such as $\mathbf{x}$ to denote a sets of variables. Most of the times, the size of this set would be clear from context. We use $\mathbf{x}^{\mathbf{e}}$ to refer to the monomial $x_1^{e_1} \cdots x_n^{e_n}$.
- We use the short-hand $\partial_{\mathbf{x}^{\mathbf{e}}}(P)$ to denote

$$\frac{\partial^{e_1}}{\partial x_1^{e_1}} \left(\frac{\partial^{e_2}}{\partial x_2^{e_2}} \left(\cdots (P) \cdots \right) \right).$$

- For a set of polynomials $\mathcal{P}$ use $\partial^{=k}\mathcal{P}$ to denote the set of all k-th order partial derivatives of polynomials in $\mathcal{P}$, and $\partial^{\leq k}\mathcal{P}$ similarly.

  Also, $\mathbf{x}^{=\ell}\mathcal{P}$ refer to the set of polynomials of the form $\mathbf{x}^{\mathbf{e}} \cdot P$ where $\mathsf{Deg}(\mathbf{x}^{\mathbf{e}}) = \ell$ and $P \in \mathcal{P}$. Similarly $\mathbf{x}^{\leq \ell}\mathcal{P}$.

- For an integer $m > 0$, we use $[m]$ to denote the set $\{1, \ldots, m\}$.

- For a set of vectors (or polynomials) V, their span over $\mathbb{F}$ will be denoted by $\mathsf{Span}(V)$ and their dimension by $\mathsf{Dim}(V)$.

- For a subset $\mathbf{y}$ of variables and a polynomial $P \in \mathbb{F}[\mathbf{x}, \mathbf{y}]$, by $\mathsf{Mult}_{\mathbf{y}}[P]$, we denote the polynomial $P' \in \mathbb{F}[\mathbf{x}, \mathbf{y}]$ which is obtained by projecting P only to its monomials which are multilinear in $\mathbf{y}$.

  Similarly, for a set S of polynomials, $\mathsf{Mult}_{\mathbf{y}}[S]$ denotes the set of polynomials obtained by projecting every polynomial in S to the monomials which are multilinear in $\mathbf{y}$.

5.2.2 The hard polynomial

The hard function for the lower bounds will be a generic balanced $\Pi\Sigma\Pi\Sigma$ circuit with appropriate parameters. We define the polynomial $P_{m,d}$ as

$$P_{m,d} \quad = \quad \prod_{i=1}^{\sqrt{d}} \sum_{j=1}^{m} \prod_{i'=1}^{\sqrt{d}} \sum_{j'=1}^{m} x_{iji'j'}.$$

The polynomial $P_{m,d}$ depends on $m^2 d$ variables. It would be useful to have $L_{iji'} = \sum_{j'} x_{iji'j'}$ so that $P_{m,d} = \prod_i \sum_j \prod_{i'} L_{iji'}$.

Observe that the polynomial $P_{m,d}$ is a set multilinear polynomial for the partition of variables into $\{\mathbf{x}_{i*i'*} : i, i' \in [\sqrt{d}]\}$, where $\mathbf{x}_{i*i'*} = \left\{x_{iji'j'} \ : \ j, j' \in [m]\right\}$. There are d such sets and each is of size m^2.

The range of parameters we will be working with in this chapter when $d = \delta \log^2 n$ for a small enough constant δ. For such small d, it follows from observations in [GKKS13] that the polynomial $P_{m,d}$ is computable by a polynomial sized non-homogeneous depth-3 circuit. More formally, the proof relies on the following lemma which is implicit in [GKKS13].

Lemma 5.3 ([GKKS13]). *Let C be a homogeneous $\Sigma\Pi^{[a]}\Sigma\Pi^{[b]}\Sigma$ circuit of size s over $\mathbb{C}$, the field of complex numbers, which computes an n-variate polynomial P. Then there is an equivalent $\Sigma\Pi\Sigma$ circuit C' of size $s' = \text{poly}(2^a, 2^b, n, s)$ which computes P.*

Using this observation, we have the following lemma which shows that there is a small depth-3 circuit for $P_{m,d}$.

Lemma 5.4. *Let P be an n variate polynomial of degree $d = O(\log^2 n)$ which is computed by a homogeneous $\Sigma\Pi^{[\sqrt{d}]}\Sigma\Pi^{[\sqrt{d}]}\Sigma$ circuit C of size s. Then, P is computable by a $\Sigma\Pi\Sigma$ circuit of size $\text{poly}(n)$.* □

Thus, to prove Theorem 5.1 and Theorem 5.2, it suffices to show an $n^{\Omega(\sqrt{d})}$ lower bound on the size of homogeneous $\Sigma\Pi\Sigma\Pi$ arithmetic circuits computing $P_{m,d}$.

5.2.3 Some useful approximations

Lemma 5.5 ([GKKS14]). *Let n, a, b satisfy $a + b = o(n)$. Then,*

$$\frac{(n+a)!}{(n-b)!} \quad = \quad n^{a+b} \cdot \exp(O((a+b)^2/n)).$$

In particular, if $a + b = o(\sqrt{n})$, then the right hand side is $(1 + o(1)) \cdot n^{a+b}$.

Lemma 5.6. *For all $x, y > 0$,*

$$e^{xy} \quad \geq \quad (1+x)^y \quad \geq \quad e^{\frac{xy}{x+1}}.$$

5.3 Proof of Theorem 5.1

The first step in previous lower bounds for homogeneous depth-4 circuits is using a random restriction to set each variable independently to zero with a certain probability. We shall first analyze the random restriction process on a homogeneous depth-4 circuit and also on the polynomial $P_{m,d}$.

5.3.1 The effect of a random restriction

Our restrictions $\mathcal{R}_p$ will be defined by setting every variable to zero with a probability $1 - p$ and keeping it alive with a probability p.

Lemma 5.7. *Let $\varepsilon > 0$ be any fixed constant and let $p = \frac{1}{n^{\varepsilon}}$. Let C be a $\Sigma\Pi\Sigma\Pi$ circuit of size $n^{\frac{\varepsilon}{2}\sqrt{d}}$. Then with a probability at least $1 - o(1)$ over $\pi \leftarrow \mathcal{R}_p$, every product gate at the lowest level of C (closest to the leaves) that depends on more than $\varepsilon\sqrt{d}$ distinct variables is set to zero in $\pi(C)$.*

Proof. Consider any product gate of support at least $\varepsilon\sqrt{d}$ present at the bottom level of C. The probability that this gate is not set to zero in $\pi(C)$ is at most $\frac{1}{n^{\varepsilon\sqrt{d}}}$. So, by a union bound over all the product gates in C, the probability that some gate of support at least $\varepsilon\sqrt{d}$ survives in $\pi(C)$ is at most $n^{\frac{\varepsilon}{2}\sqrt{d}} \cdot \frac{1}{n^{\varepsilon\sqrt{d}}}$ which is $o(1)$. □

We now analyse the effect of random restrictions on our candidate hard function.

Lemma 5.8. *Let ε be a fixed constant and let $p = \frac{1}{n^{\varepsilon}}$, and let $P_{m,d}$ be the polynomial as defined in Subsection 5.2.2. Then, with probability at least $1 - o(1)$ over $\pi \leftarrow \mathcal{R}_p$, the polynomial $\pi(P_{m,d})$ is of the form*

$$\pi(P_{m,d}) = \prod_{i=1}^{\sqrt{d}} \sum_{j=1}^{m} \prod_{i'=1}^{\sqrt{d}} L'_{iji'}$$

where each $L'_{iji'}$ is a non-zero linear form.

Proof. From our choice of parameters, observe that $n = m^2 d$, and since $d = O(\log^2 n)$, $m > n^{1/4}$. Now, for any fixed linear form $L_{iji'}$, the probability that $\pi(L_{iji'})$ equals zero is equal to $(1-p)^m = (1 - 1/n^{\varepsilon})^m$ which is less than $(1 - 1/n^{\varepsilon})^{n^{2\varepsilon}} = \frac{1}{\omega(n)}$. Therefore, the probability that there exists a linear form $L_{iji'}$ such that $\pi(L_{iji'}) \equiv 0$ is $o(1)$, and the lemma follows. □

At this point, we will deterministically set all but one alive variable in each $L'_{iji'}$ in the above lemma to zero, and obtain the following corollary up to a relabelling of variables.

Corollary 5.9. *Let ε be a fixed constant and $p = \frac{1}{n^{\varepsilon}}$, and let $P_{m,d}$ be the polynomial as defined in Subsection 5.2.2. Then, with probability at least $1 - o(1)$ over $\pi \leftarrow \mathcal{R}_p$, there is a $0,1$ projection of $\pi(P_{m,d})$ which is of the form*

$$P'_{m,d} = \prod_{i=1}^{\sqrt{d}} \sum_{j=1}^{m} \prod_{i'=1}^{\sqrt{d}} x_{iji'},$$

where each $x_{iji'}$ is a distinct variable.

Observe that Lemma 5.7 continues to hold under this additional deterministic restriction, as the bottom support of a depth-4 circuit does not increase under $0,1$ projections. Clearly $P'_{m,d}$ is computable by a homogeneous depth-4 circuit of bottom fan-in $\sqrt{d}$.

In order to complete the proof, it suffices to show that any homogeneous depth-4 circuit of *bottom support* bounded by $\sqrt{d}/10$ that computes $P'_{m,d}$ must have size $n^{\Omega(\sqrt{d})}$. In fact, Kumar and Saraf [KS15d] have shown that any homogeneous depth-4 circuit of bottom fan-in at most $\sqrt{d}/10$ computing $P'_{m,d}$ must require size $n^{\Omega(\sqrt{d})}$ using the measure of dimension of shifted partial derivatives. Thus we need to find a way to lift this lower bound to the class of homogeneous depth-4 circuit of *bottom support* bounded by $\sqrt{d}/10$. To do this, we modify the measure of dimension of shifted partials in order to address small bottom support instead of small bottom fan-in.

5.3.2 The complexity measure

The measure is again the dimension of an appropriate linear space of polynomials.

Definition 5.10 (The complexity measure)**.** *Let $\mathbf{x} = \mathbf{x}_1 \sqcup \cdots \sqcup \mathbf{x}_d$ be a partition of the variables into d sets. For any polynomial $P \in \mathbb{F}[\mathbf{x}]$, define $P' \in \mathbb{F}[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d, y_1, y_2, \ldots, y_d]$ be the the polynomial derived from P by replacing every occurence of the variable $x_{ij} \in \mathbf{x}_i$ by $y_i \cdot x_{ij}$. Then, the complexity measure*

$$\Gamma_{k,\ell}(P) \quad := \quad \mathsf{Dim}_{\mathbb{F}}\left\{\left(\mathbf{x}^{=\ell} \cdot \mathsf{Mult}_{\mathbf{y}}[\partial^{=k}(P')]\right)\right\}. \qquad \Diamond$$

We remark that all the derivatives and shifts in the definition of $\Gamma_{k,\ell}$ are taken with respect to the variables in $\mathbf{x}$. However, the multilinearization is done with respect to the $\mathbf{y}$ variables. As mentioned earlier, this measure was used in a previous version of [KLSS14a] where it was called *dimension of shifted projected partial derivatives.*

Throughout this chapter, we will be using very simple connections between the measure $\Gamma_{k,\ell}$ and the well known notion of shifted partial derivatives of polynomials, defined as

Definition 5.11 (Shifted partial derivatives)**.** *Define $P \in \mathbb{F}[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d]$ be a polynomial. Then, the* dimension of shifted partial derivatives *is defined as*

$$\mathsf{Dim}_{\mathbb{F}} \left\{ \left(\mathbf{x}^{=\ell} \cdot \partial^{=k}(P') \right) \right\}. \qquad \Diamond$$

Observe that if a polynomial P is set-multilinear with respect to the partition of the variables in $\mathbf{x}$ into $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d$, then multilinearization with respect to the $\mathbf{y}$ variables does not kill any of the monomials in the partial derivatives. In particular, for a set multilinear polynomial P, and for every choice of k, ℓ, the quantitity $\Gamma_{k,\ell}(P)$ is exactly equal to the dimension of shifted partial derivatives of the polynomial P where we take derivatives of order k and shifts are of degree ℓ. This observation will be useful for us in the proof and is summarised below.

Observation 5.12. *Let P be a set multilinear polynomial of degree d. Then for every choice of parameters k and ℓ,*

$$\Gamma_{k,\ell}(P) \;=\; \mathsf{Dim}\left(\mathbf{x}^{=\ell} \cdot \partial^{=k}(P) \right).$$

□

Since $P_{m,d}$ is set multilinear with respect to the partition

$$\mathbf{x} \;=\; \bigsqcup_{i,i' \leq \sqrt{d}} \mathbf{x}_{i*i'*}$$

we use this partition for in the definition of $\Gamma_{k,\ell}$. To complete the proof, we use this measure to show that $P'_{m,d}$ cannot be computed by small homogeneous depth-4 circuit of bottom support bounded by $\sqrt{d}/10$.

5.3.3 Upper bound for a small bottom-support depth-4 circuit

Lemma 5.13. *Let C be a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit with bottom support at most s which computes a degree d polynomial in $\mathbb{F}[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d]$. Then, for every k and ℓ,*

$$\Gamma_{k,\ell}(C) \;\leq\; \mathit{Size}(C) \cdot 2^{2d} \cdot \binom{d}{k} \cdot \binom{n + \ell + ks}{n}.$$

Proof. Since the measure $\Gamma_{k,\ell}$ is subadditive, we will prove an upper bound on $\Gamma_{k,\ell}$ for one product term in C. So, let $T = Q_1 \cdot Q_2 \cdots Q_t$, where each Q_i has support at most s. Without loss of generality, we can assume that $t \leq d$ since the circuit C is homogeneous to start with.

Recall that in the first step, we replace every variable x_{ij} by $y_i \cdot x_{ij}$. This transforms $T = Q_1 \cdots Q_t$ into $T = Q'_1 \cdot Q'_2 \cdots Q'_t$. Every monomial $\mathbf{x}^\alpha$ in the $\mathbf{x}$ variables will be transformed to a monomial $\mathbf{y}^{\alpha'} \cdot \mathbf{x}^\alpha$ by this transformation. The key points are that $\mathbf{y}^{\alpha'}$ is only over d variables, and if $\mathbf{x}^\alpha$ is non-multilinear then so is $\mathbf{y}^{\alpha'}$.

Let us now consider the derivative of T with respect to a monomial $\mathbf{x}^\alpha$ of order k.

$$\partial_{\mathbf{x}^\alpha}(T') \in \mathsf{Span}\left\{\partial_{\mathbf{x}^\alpha}(Q'_A) \cdot Q'_{\overline{A}} \ : \ A \subseteq [t], |A| \leq k\right\},$$

where Q'_A is a shorthand for $\prod_{i \in A} Q'_i$.

$$\mathsf{Mult}_{\mathbf{y}}\left[\partial_{\mathbf{x}^\alpha}(T')\right] \in \mathsf{Span}\left\{\mathsf{Mult}_{\mathbf{y}}\left[\partial_{\mathbf{x}^\alpha}(Q'_A) \cdot Q'_{\overline{A}}\right] \ : \ A \subseteq [t], |A| \leq k\right\}.$$

Since we are interested in the multilinear component, it suffices to only focus on multilinear (in $\mathbf{y}$) monomials in both $\partial_{\mathbf{x}^\alpha}(Q'_A)$ and $Q'_{\overline{A}}$. Since Q'_A is a product of at most k polynomials, each of support-size bounded by s, the only monomials $\mathbf{x}^\beta$ that can contribute a non-multilinear $\mathbf{y}$-part can have degree at most ks. Therefore,

$$\begin{aligned}
\mathsf{Mult}_{\mathbf{y}}\left[\partial_{\mathbf{x}^\alpha}(Q'_A)\right] &\in \mathsf{Span}\left\{\mathbf{y}^\beta \cdot \mathbf{x}^\gamma \ : \ \mathsf{Deg}(\mathbf{x}^\gamma) \leq ks \ , \ \mathbf{y}^\beta \text{ multilinear}\right\} \\
\mathsf{Mult}_{\mathbf{y}} Q'_{\overline{A}} &= \sum_{\beta'} \mathbf{y}^{\beta'} \cdot Q'_{\overline{A},\beta'} \\
\implies \mathsf{Mult}_{\mathbf{y}}\left[\partial_{\mathbf{x}^\alpha}(Q'_A) \cdot Q'_{\overline{A}}\right] &\in \mathsf{Span}\Big\{\mathbf{y}^\beta \mathbf{y}^{\beta'} \cdot \mathbf{x}^\gamma \cdot Q'_{\overline{A},\beta'} \ : \\
&\qquad \mathsf{Deg}(\mathbf{x}^\gamma) \leq ks \ , \ \mathbf{y}^\beta \mathbf{y}^{\beta'} \text{ multilinear}\Big\}.
\end{aligned}$$

Taking the union over all shifts and all derivatives, we get

$$\begin{aligned}
\mathbf{x}^{=\ell} \cdot \mathsf{Mult}_{\mathbf{y}}[\partial^{=k}(T')] \subseteq \mathsf{Span}\Big\{&\mathbf{y}^\beta \mathbf{y}^{\beta'} \cdot \mathbf{x}^\gamma \cdot Q'_{\overline{A},\beta'} \ : \ A \subseteq [t] \ , \ |A| \leq k \ , \\
&\text{degree } (\mathbf{x}^\gamma) \leq \ell + ks \ , \ \mathbf{y}^\beta \mathbf{y}^{\beta'} \text{ is multilinear } \Big\}.
\end{aligned}$$

For any k, ℓ, it follows that

$$\Gamma_{k,\ell}(T') \leq 2^{2d} \cdot \binom{d}{k} \cdot \binom{n+\ell+ks}{n}.$$

Using subadditivity, we obtain the lemma. □

5.3.4 Lower bound for the measure on $P'_{m,d}$

The final technical ingredient of our proof will be a lower bound on the dimension of shifted partials of the polynomial $P'_{m,d}$. The bound follows from the calculations in [KS15d], but we provide the calculation here for completeness.

Lemma 5.14. *Recall the polynomial*

$$P'_{m,d} \;=\; \prod_{i=1}^{\sqrt{d}} \sum_{j=1}^{m} \prod_{i'=1}^{\sqrt{d}} x_{iji'}$$

where each $x_{iji'}$ is a distinct variable. For $k = \sqrt{d}$ and any ℓ, we have

$$\mathsf{Dim}\left(\mathbf{x}^{=\ell} \cdot \partial^{=k}(P)\right) \;\geq\; \frac{1}{4} \cdot \left(\frac{n+\ell}{\ell}\right)^{\frac{1}{2}\cdot(d-\sqrt{d})} \cdot \binom{n+\ell-1}{n}.$$

Proof. To show that the shifted partials complexity of P is large, we will follow the outline in [KS15d]. We consider the following subset $\mathcal{S}$ of monomials of degree equal to $k = \sqrt{d}$:

$$\mathcal{S} = \{x_{1a_11} \cdot x_{2a_21} \cdots x_{ka_k1} : a_1, a_2, \ldots, a_k \in [m]\}.$$

Firstly, note that for any monomial $\mathbf{x}^\alpha = x_{1a_11} \cdots x_{ka_k1} \in \mathcal{S}$, the derivative $\partial_{\mathbf{x}^\alpha}(P)$ is just the monomial

$$(x_{1a_12} \cdots x_{1a_1k}) \cdots (x_{1a_k2} \cdots x_{1a_kk})\,.$$

Thus, it suffices to get a lower bound of distinct monomials obtained as shifts of such derivatives. To assist this calculation, we pick a subset $\mathcal{S}'$ of the set $\mathcal{S}$ such that the distance between any two monomials in $\mathcal{S}'$ is 'large', and the size of $\mathcal{S}'$ is also 'large'. This can be done by picking the monomials which correspond to a good code of length k over the alphabet $\Sigma = \{1, 2, \ldots, m\}$. To this end, we pick a Reed-Somolon code of relative distance $1/2$ and rate $1/2$. This can be done as long as m is a prime power and $\sqrt{d} \leq m$. Let $\mathcal{S}'$ be a such set of size $m^{k/2}$ where any pair of monomials in $\mathcal{S}'$ differ on at least $\sqrt{d}/2$ locations.

When we take derivatives of P with respect to monomials in the set $\mathcal{S}'$, two monomials obtained from distinct elements of $\mathcal{S}'$ have distance at least $\Delta = \sqrt{d}(\sqrt{d}-1)/2 = (d-\sqrt{d})/2$. So, each of the shifted partial derivatives obtained by shifting the derivatives

of P by monomials of degree ℓ is just a monomial, and a lower bound on the number of distinct monomials obtained in this way gives us a lower bound on $\mathsf{Dim}\left(\mathbf{x}^{=\ell} \cdot \partial^{=k}(P)\right)$. In fact, we shall choose an even smaller set $\mathcal{S}''$ to ensure the following bounds work out.

By the inclusion-exclusion approach of Chillara and Mukhopadhyay [CM14a], for any set $\mathcal{S}'' \subset \mathcal{S}'$ we get the following:

$$\mathsf{Dim}\left(\mathbf{x}^{=\ell} \cdot \partial^{=k}(P)\right) \;\geq\; \left|\mathcal{S}''\right| \cdot \binom{n+\ell-1}{n} \;-\; \frac{\left|\mathcal{S}''\right|^2}{2} \cdot \binom{n+\ell-\Delta-1}{n}.$$

If we pick our parameters, such that the first term above is at least twice the second term, then we would be done. For this, we need

$$\left|\mathcal{S}''\right| \;\leq\; \frac{\binom{n+\ell-1}{n}}{\binom{n+\ell-\Delta-1}{n}}.$$

For our choice of parameters, $\ell, n \gg d^2$, the ratio $\frac{\binom{n+\ell-1}{n}}{\binom{n+\ell-\Delta-1}{n}}$ can be approximated by $\left(\frac{n+\ell}{\ell}\right)^{\Delta}$ within a factor $1 \pm o(1)$ by Lemma 5.5. So, it suffices if our choice of parameters satisfies (omitting floors)

$$\left|\mathcal{S}''\right| \;=\; \frac{1}{2} \cdot \left(\frac{n+\ell}{\ell}\right)^{\Delta}.$$

Plugging in Δ and the size of $\mathcal{S}''$ in the inclusion-exclusion bound, we get

$$\mathsf{Dim}\left(\mathbf{x}^{=\ell} \cdot \partial^{=k}(P)\right) \;\geq\; \frac{1}{4} \cdot \left(\frac{n+\ell}{\ell}\right)^{(d-\sqrt{d})/2} \cdot \binom{n+\ell-1}{n}. \qquad \square$$

5.3.5 Putting it together

Theorem 5.15 (Theorem 5.1 restated)**.** *Let C be a homogeneous depth-4 arithmetic circuit which computes the polynomial $P_{m,d}$ for $d = 0.0001 \log^2 n$. Then, the size of C is at least* $\exp(\Omega(\sqrt{d} \log n))$.

Proof. Assume on the contrary that the polynomial $P_{m,d}$ can be computed by C, a homogeneous depth-4 circuit of size at most $\exp(0.001\sqrt{d} \log n)$. If we apply a random restriction that sets every variable to zero independently with probability $1/n^{0.1}$, by Lemma 5.7 (with $\varepsilon = 0.1$), the circuit reduces to C', a homogeneous depth-4 circuit with bottom support bounded by $\sqrt{d}/10$ with probability $1 - o(1)$.

On the other hand by Corollary 5.9, the polynomial $P_{m,d}$ under such a random restriction still retains $P'_{m,d}$ as a projection with high probability. Fix a restriction that

satisfies both these properties and we now have a homogeneous depth-4 circuit C'' with bottom support bounded by $\sqrt{d}/10$ and size at most $\exp(0.001\sqrt{d}\log n)$ that computes $P'_{m,d}$.

Let $k = \sqrt{d}$ and $\ell = \frac{n\sqrt{d}}{\log n}$. By Lemma 5.13, we have

$$\Gamma_{k,\ell}(C'') \quad \leq \quad \text{Size}(C'') \cdot 2^{2d} \cdot \binom{n+\ell+(0.1)d}{n}.$$

On the other hand, by Lemma 5.14 and Observation 5.12,

$$\Gamma_{k,\ell}(P'_{m,d}) \quad \geq \quad \frac{1}{4} \cdot \left(\frac{n+\ell}{\ell}\right)^{(d-\sqrt{d})/2} \cdot \binom{n+\ell-1}{n}.$$

Together, this implies that

$$\text{Size}(C'') \geq \frac{1}{4} \cdot \frac{\binom{n+\ell-1}{n} \cdot \left(\frac{n+\ell}{\ell}\right)^{(d-\sqrt{d})/2}}{2^{2d} \cdot \binom{n+\ell+(0.1)d}{n}}.$$

For our regime of parameters, $\sqrt{d} = 0.01 \log n$ and hence $2^{2d} = n^{0.02\sqrt{d}} = \exp(0.02\sqrt{d}\log n)$. Simplifying the ratio of binomial coefficients using (Lemma 5.5), and using $\frac{d-\sqrt{d}}{2} > \frac{d}{3}$, we get

$$\begin{aligned}
\text{Size}(C'') &\geq \frac{1}{\exp(0.02\sqrt{d}\log n)} \cdot \left(1+\frac{n}{\ell}\right)^{d/3} \\
&\geq \frac{1}{\exp(0.02\sqrt{d}\log n)} \cdot \exp\left(\frac{(nd/3\ell)}{(n/\ell)+1}\right) \quad \text{(By Lemma 5.6)} \\
&> \exp\left(0.1\sqrt{d}\log n\right),
\end{aligned}$$

which contradicts the assumption on the size of C. Hence $\text{Size}(C) \geq \exp(0.001\sqrt{d}\log n)$.

□

Chapter 6

Exponential lower bounds for depth-5 circuits over small finite fields[1]

6.1 Introduction

The results in Chapter 4 [KS17] show that the reduction from general arithmetic circuits to depth-4 circuits with support $O(\sqrt{d})$ cannot be improved, as they give an example of a polynomial in VP for which any depth-4 circuits of support $O(\sqrt{d})$ must be of size $n^{\Omega(\sqrt{d})}$. Further, with the current upper-bounds for the projected shifted partials on such depth-4 circuits, the best we can hope to prove using this measure is an $n^{\Omega(\sqrt{d})}$ lower bound. Hence, it might be insufficient for general arithmetic circuits lower bounds but it could well be the case that we might be able to prove stronger lower bounds for constant depth arithmetic circuits, or arithmetic formulas by variants of this family of measures.

Hence, as a start, the problem of proving lower bounds for homogeneous depth five circuits, seems like the next natural question to explore. This already seems to introduce new challenges as the proofs of lower bounds for homogeneous depth-4 circuits seem to break down for homogeneous depth-5 circuits. In this chapter, we pursue this line of enquiry, and prove exponential lower bounds for homogeneous depth-5 circuits over small finite fields. Before stating our results, we first discuss prior results on this question, and the challenges involved in extending the proofs of lower bounds for homogeneous depth four circuits, in the next section.

[1]The results in this chapter appear in [KS15b].

Lower bounds for depth-5 circuits

Prior to this work, the only known lower bounds for depth-5 circuits that we are aware of are the results of Raz [Raz10b], which show superlinear lower bounds for bounded depth circuits over large enough fields, the results of Kalorkoti [Kal85] which show quadratic lower bounds for arithmetic formulas and the results of Bera and Chakrabarti [BC15] and Kayal and Saha [KS15a] which show exponential lower bounds for homogeneous depth-5 circuits if the bottom fan-in is bounded.

Given that we have lower bounds for homogeneous depth-4 circuits, it seems natural to try and apply these techniques to prove lower bounds for homogeneous depth-5 circuits. Unfortunately, the obvious attempts to generalize the proofs in [KLSS14a, KS17] seem to fail for homogeneous depth-5 circuits. We now elaborate on this.

On extending the depth-4 lower bound proofs to depth-5 circuits

To understand these issues, we first need a birds-eye view of the major steps in the proofs of lower bounds for depth-4 circuits [KLSS14a, KS17]. These proofs have two major components.

- **Reduction to depth-4 circuits with bounded bottom support :** In the first step, the circuit C and the polynomial are hit with a random restriction, in which each variable is kept alive independently with some small probability p. The observation is that a bottom level product gate in C of support (the number of distinct variable inputs) at least s survives with probability at most p^s. Therefore, the probability that some bottom product of support at least s in C survives is at most $\text{Size}(C) \cdot p^s$. Now, if the size of C is small (say $\varepsilon \cdot 1/p^s$), then this probability is quite small, so with a high probability C reduces to a homogeneous depth-4 circuit with bounded bottom support.

- **Lower bounds for depth-4 circuits with bounded bottom support :** The goal in the second step is to show that the polynomial obtained after random restrictions still remains hard for homogeneous depth-4 circuits with bottom support at most s.

The key point in step 1 is that if Size(C) is not too large, then we can assume that with a high probability over the random restrictions, all the high support product gates are set to 0. This is where things are not quite the same for depth-5 circuits. When we express a homogeneous depth-5 circuit as a homogeneous depth-4 circuit by expanding the product of linear forms at level four, we might increase the number of monomials a lot (potentially to all possible monomials). Now, the random restriction step no longer works and we do not have a reduction to homogeneous depth-4 circuits with bounded bottom support. If the bottom fan-in of C is bounded, then this strategy does indeed generalize. Bera and Chakrabarti [BC15] and Kayal and Saha [KS15a] show exponential lower bounds for such cases.

It is not clear to us how fundamental this obstruction is, but our key insight is a strategy for proving lower bounds for homogeneous depth-4 circuits that avoids the random restriction step. Morally speaking, we *do* proceed by a 'reduction' from a depth-5 circuit to a depth-4 circuit, but the meaning of a 'reduction' here is more subtle and largely remains implicit.

Our Contribution

We give an exponential lower bound for homogeneous depth-5 circuits over any fixed finite field $\mathbb{F}_q$. To the best of our understanding, this is the first such lower bound for depth-5 circuits over any field apart from $\mathbb{F}_2$[2]. Stated precisely, we prove the following theorem.

Theorem 6.1. *There is an explicit family of polynomials* $\{P_d : d \in \mathbb{N}\}$*, with* $\mathsf{Deg}(P_d) = d$*, in the class* VNP *such that for any finite field* $\mathbb{F}_q$*, any homogeneous depth-*5 *circuit computing* P_d *must have size* $\exp(\Omega_q(\sqrt{d}))$*.*

The polynomial P_d is from the Nisan-Wigderson family of polynomials (introduced by [KSS14], Definition 6.3) with carefully chosen parameters.

Our proof also extends to non-homogeneous depth-5 circuits where the layer of multiplication gates closer to the output have fan-in bounded by $O(\sqrt{d})$ (with no restriction

[2]For $\mathbb{F}_2$, exponential lower bounds easily follow from the lower bounds of Razborov [Raz87]

on the fan-in of the other multiplication layer).

Theorem 6.2. *There is an explicit family of polynomials* $\{P_d : d \in \mathbb{N}\}$*, with* $\mathsf{Deg}(P_d) = d$*, in the class* VNP *such that for any finite field* $\mathbb{F}_q$*, any* $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi\Sigma$ *circuit computing* P_d *must have size* $\exp(\Omega_q(\sqrt{d}))$*.*

It is worth mentioning that for characteristic zero fields, it suffices to prove an $\exp(\omega(d^{1/3}\log d))$ lower bound for an explicit polynomial computed by such $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi\Sigma$ circuits to separate VP from VNP (by combining the depth reductions of [AV08, Koi12, Tav15] and [GKKS13]). We elaborate on this in Subsection 6.7.4. Such a phenomenon also happens for non-homogeneous depth three circuits, where over finite fields, we know quite strong lower bounds while much weaker ones would imply $\mathsf{VNP} \neq \mathsf{VP}$ over fields of characteristic zero.

The key technical ingredient of our proof is to look at the space of shifted partial derivatives and the projected shifted partial derivatives of a polynomial. We study them as a space of functions from $\mathbb{F}_q^n \to \mathbb{F}_q$ as opposed to as a space of formal polynomials, as has been the case for the results obtained so far. This perspective allows us the freedom to confine our attention to the evaluations of the shifted partial derivatives of a polynomial on certain well chosen subsets of $\mathbb{F}_q^n$, and this turns out to be critical to our cause. This leads to a new family of complexity measures which could have applications to other lower bound questions as well. Our proof also involves a tighter analysis of the lower bound of Kumar and Saraf [KS17] (for homogeneous depth-4 circuits) which may be interesting in its own right.

We now give an overview of our proof.

6.2 An overview of the proof

The proof would consist of the following main steps:

1. Define a function $\Gamma : \mathbb{F}_q[\mathbf{x}] \to \mathbb{N}$. Intuitively, we think of $\Gamma(P)$ to be a measure of the *complexity* of P.

2. For all homogeneous depth-5 circuits C of size at most $\exp(\delta\sqrt{d})$, prove an upper bound on $\Gamma(C)$.

3. For the target hard polynomial P, show that $\Gamma(P)$ is much larger than the upper bound proved in step 2.

The complexity measure

At a high level, the proof of lower bounds in [NW97, GKKS14, KSS14, FLMS14, KS15d, KLSS14a, KS17] associate a linear space polynomials to every polynomial in $\mathbb{F}_q[\mathbf{x}]$ and use the dimension of this space over $\mathbb{F}_q$ as a measure of complexity of the polynomial. The mapping from polynomials to linear space of polynomials undergoes subtle changes as we go from the proof of lower bounds for homogeneous depth-3 circuits [NW97] to lower bounds for homogeneous depth-4 circuits [KLSS14a, KS17].

In this chapter, we follow this outline and associate to every polynomial, the space of its shifted partial derivatives as defined in [GKKS14]. However, instead of working with this space of polynomials as it is, we study their evaluation vectors over a subset of $\mathbb{F}_q^n$ (similar to [GK98, GR00], where they worked with partial derivatives of a polynomial). The key gain that we have from this change in outlook is that as evaluation vectors, we can choose to confine our attention to evaluations on certain properly chosen subsets of $\mathbb{F}_q^n$. For formal polynomials, it is not clear what should be the correct analog of this approximation. The necessity and the utility of this will be more clear as we go along.

High rank products of linear forms

Consider a polynomial Q which is a product of τ linearly independent linear forms $L_1, L_2, \ldots, L_\tau$.

$$Q = \prod_{i=1}^{\tau} L_i$$

It is not hard to see that

$$\Pr_{\mathbf{a}\in\mathbb{F}_q^n}[Q(\mathbf{a}) \neq 0] \leq \left(1 - \frac{1}{q}\right)^{\tau}$$

In other words, products of linear forms of rank τ vanish on all but a $o(1)$ fraction of the entire space if $\tau = \omega(1)$. If the size of a depth-5 circuit is not too large as a function of τ (say, at most $\exp(\delta\tau)$ for a small enough $\delta > 0$), then by a union bound, all the products of rank at least τ at the fourth level vanish everywhere apart from a $o(1)$ fraction of the points in $\mathbb{F}_q^n$.

In summary, we just argued that a depth-5 circuit C over $\mathbb{F}_q$ of size at most $\exp(\delta\tau)$ can be approximated by a sub-circuit C' of C which is obtained from C by dropping all products of linear forms of rank at least τ from the bottom level.

Low rank products of linear forms

A second simple observation (Lemma 6.7) shows that for every product of linear forms of rank at most τ, there is a polynomial of degree at most $(q-1)\tau$, such that they agree at all points in $\mathbb{F}_q^n$. Thus, the circuit C' is equal, as a function from $\mathbb{F}_q^n \to \mathbb{F}_q$ to a depth-4 circuit C'' of bottom fan-in at most $(q-1)\tau$. Moreover, the formal degree and the top fan-in of C'' are upper bounded by the formal degree and top fan-in of C, respectively.

Putting things together

This implies that for every homogeneous depth-5 circuit C computing a polynomial of degree d of size at most $\exp(\delta\tau)$ for some τ, there exists a depth-4 circuit C'' of formal degree at most d and top fan-in at most $\exp(\delta\tau)$ such that

$$\Pr_{\mathbf{a}\in\mathbb{F}_q^n}[C(\mathbf{a}) \neq C''(\mathbf{a})] \leq o(1).$$

Therefore, a polynomial P which can be computed by C can be *approximated* by C'' in the pointwise sense. Since we know lower bounds on the top fan-in of homogeneous (and low formal degree) depth-4 circuits with bounded bottom fan-in [GKKS14, KSS14], it seems that we only have a small way to go. Unfortunately, we do not quite know

how to make this idea work. The key technical obstacle here is that it seems to be hard to say much about the partial derivatives of C by looking at partial derivatives of C''. As a pathological case, the polynomial $\prod_{i\in[n]} x_i$ has a partial derivative span of size 2^n but is well approximated by the 0 polynomial over $\mathbb{F}_2$.

If we had started with a depth-3 circuit instead of a depth-5 circuit, then such a strategy is indeed known to work [GR00]. Observe that in this case it is enough to show that that there is an explicit polynomial which cannot be approximated well by a low degree polynomial over $\mathbb{F}_q$. In [GR00], the authors show this by an adaptation of a similar result of Smolenksy [Smo87] over $\mathbb{F}_2$.

A strengthening of the strategy

The key additional observation that helps us make things work is the fact that not only do high rank product gates at level four of C vanish almost everywhere on $\mathbb{F}_q^n$, but they vanish with a high multiplicity. As we show in Corollary 6.10, if the size of C is not *too large*, all the product gates of rank at least τ vanish with a multiplicity $\Omega(\tau)$ at $1-o(1)$ fraction of points on $\mathbb{F}_q^n$[3].

Therefore, not only can C agree with C' almost everywhere, all the partial derivatives of C of order at most $k = \Omega(\tau)$ agree with all the partial derivatives of C' almost everywhere. This already hints at the fact that if we are to take advantage of this then we should be looking at the evaluation of these derivatives, since our only guarantee is about their evaluations.

In [GK98], exponential lower bounds were proved for non-homogeneous depth-3 circuits using a very similar strategy. However, adapting the method for shifted partials and projected shifted partials seems to be a challenge.

In Section 6.4, we show that the dimension of the span of evaluation vectors of shifted partial derivatives of C, when restricted to a properly chosen subset S of $\mathbb{F}_q^n$, is small if the size of the circuit C we started with was small.

As a final step of the proof, we show that with respect to this complexity measure,

[3]In the rest of this discussion, we will think of τ as $\Theta(\sqrt{d})$

our target hard polynomial (from the so-called *Nisan-Wigderson* family, defined below) has a large complexity.

Definition 6.3 (Nisan-Wigderson polynomial families)**.** *Let d, m, e be arbitrary parameters with m being a power of a prime, and $d, e \leq m$. Since m is a power of a prime, let us identify the set $[m]$ with the field $\mathbb{F}_m$ of m elements. Note that since $d \leq m$, we have that $[d] \subseteq \mathbb{F}_m$. The Nisan-Wigderson polynomial with parameters d, m, e, denoted by $\mathsf{NW}_{d,m,e}$ is defined as*

$$\mathsf{NW}_{d,m,e}(\mathbf{x}) \quad = \quad \sum_{\substack{p(t) \in \mathbb{F}_m[t] \\ \mathsf{Deg}(p) < e}} x_{1,p(1)} \dots x_{d,p(d)}$$

That is, for every univariate polynomial $p(t) \in \mathbb{F}_m[t]$ of degree less thatn e, we add one monomial that encodes the 'graph' of p on the points $[d]$. This is a homogeneous, multilinear polynomial of degree d over dm variables with exactly m^e monomials. $\Diamond$

This step of the proof builds on a tighter analysis of the lower bound on the dimension of the span of *projected shifted partial derivatives* of the Nisan-Wigderson polynomials in [KS17]. We show that if the set S is carefully chosen, then we do not incur much loss in the dimension by going from looking at shifted partial derivatives as formal polynomials to looking at them as functions over a small subset of the finite field. We provide the details in Section 6.5.

One important technicality is the dependency between various parameters involved. For our proof, the choice of k would be $O_q(\tau)$ and would depend on q. The lower bound of [KS17] would then choose specific parameters for the $\mathsf{NW}_{d,m,e}$. This would mean that for every q, we get a *different* polynomial for which we show a lower bound. We remedy the order of quantifiers and start by fixing specific parameters for $\mathsf{NW}_{d,m,e}$. Then, depending on q, we choose a restriction of $\mathsf{NW}_{d,m,e}$ that would be identical to $\mathsf{NW}_{d',m,e}$ by setting some variables to 0/1. We then apply the [KS17] argument for this restriction to obtain our lower bound for $\mathsf{NW}_{d',m,e}$ which also yields a lower bound for $\mathsf{NW}_{d,m,e}$. The details are in Subsection 6.6.1.

6.3 Notation

- Throughout the chapter, we shall use bold-face letters such as $\mathbf{x}$ to denote a set $\{x_1, \ldots, x_n\}$. Most of the times, the size of this set would be clear from context. We shall also abuse this notation to use $\mathbf{x}^{\mathbf{e}}$ to refer to the monomial $x_1^{e_1} \cdots x_n^{e_n}$.

- For an integer $m > 0$, we shall use $[m]$ to denote the set $\{1, \ldots, m\}$.

- We shall use the short-hand $\partial_{\mathbf{x}^{\mathbf{e}}}(P)$ to denote

$$\frac{\partial^{e_1}}{\partial x_1^{e_1}} \left(\frac{\partial^{e_2}}{\partial x_2^{e_2}} \left(\cdots (P) \cdots \right) \right).$$

- For a set of polynomials $\mathcal{P}$ shall use $\partial^{=k}\mathcal{P}$ to denote the set of all k-th order partial derivatives of polynomials in $\mathcal{P}$, and $\partial^{\leq k}\mathcal{P}$ similarly.

  Also, $\mathbf{x}^{=\ell}\mathcal{P}$ shall refer to the set of polynomials of the form $\mathbf{x}^{\mathbf{e}} \cdot P$ where $\mathsf{Deg}(\mathbf{x}^{\mathbf{e}}) = \ell$ and $P \in \mathcal{P}$. Similarly $\mathbf{x}^{\leq \ell}\mathcal{P}$.

- For a polynomial $P \in \mathbb{F}_q[\mathbf{x}]$ and for a set $S \subseteq \mathbb{F}_q^n$, we shall denote by $\mathsf{Eval}_S(P)$ the vector of the evaluation of P on points in S (in some natural predefined order like say the lexicographic order). For a set of vectors V, their span over $\mathbb{F}_q$ will be denoted by $\mathsf{Span}(V)$ and their dimension by $\mathsf{Dim}(V)$.

- We shall use $\mathcal{H}$ to denote the set $\{0,1\}^n \subset \mathbb{F}_q^n$.

The complexity measure

We now define the complexity measure that we shall be using to prove the lower bound. The measure will depend on a carefully chosen set $S \subset \mathbb{F}_q^n$.

Definition 6.4 (The complexity measure)**.** *Let k, ℓ be some parameters and let $S \subset \mathbb{F}_q^n$. For any polynomial P, define $\Gamma_{k,\ell,S}(P)$ as*

$$\Gamma_{k,\ell,S}(P) \quad := \quad \mathsf{Dim}\left\{ \mathsf{Eval}_S\left(\mathbf{x}^{=\ell}\partial^{=k}(P) \right) \right\}. \qquad \Diamond$$

6.4 Complexity measure on a depth-5 circuit

A depth-5 circuit computes a polynomial of the form

$$C \quad = \quad \sum_a \prod_b \sum_c \prod_d L_{abcd} \tag{6.5}$$

where each L_{abcd} are linear polynomials.

Definition 6.6 (Terms of a circuit, and rank)**.** *For a depth-5 circuit such as* (6.5)*, we shall denote by* $\mathrm{Terms}(C)$ *the set*

$$\mathrm{Terms}(C) \quad := \quad \left\{ \prod_d L_{abcd} \right\}_{a,b,c}$$

which are all products of linear polynomials computed by the bottommost product gates.

For any term $T = \prod_d L_d$*, define* $\mathsf{Rank}(T)$ *to be* $\mathsf{Dim}\{L_d\}_d$*, which the maximum number of independent linear polynomials among the factors of* T*. For a depth-5 circuit* C*, we shall use* $\mathsf{Rank}(C)$ *to denote* $\max_{T \in \mathrm{Terms}(C)} \mathsf{Rank}(T)$*.*

For a parameter τ*, we shall use* $\mathrm{Terms}_{>\tau}(C)$ *to refer to terms* $T \in \mathrm{Terms}(C)$ *with* $\mathsf{Rank}(T) > \tau$*.* ♢

Low rank gates are low-degree polynomials

The following Lemma, present implicitly in [GK98, GR00], is a very useful transformation of gates of low-rank (and possibly large degree) when working over a finite field.

Lemma 6.7 ([GK98, GR00])**.** *Let* Q *be a product of linear polynomials of rank at most* τ*. Then, there is a polynomial* $\tilde{Q}$ *of degree at most* $(q-1) \cdot \tau$ *such that* $\tilde{Q}(\mathbf{a}) = Q(\mathbf{a})$ *for all* $\mathbf{a} \in \mathbb{F}_q^n$*.*

Proof. Without loss of generality, we shall assume that the rank is equal to τ, as the degree upper bound will only be better for a smaller rank and let $L_1, \ldots, L_\tau$ be linearly independent. Let

$$Q = \prod_{i=[\tau]} L_i \cdot \prod_{j \notin [\tau]} L_j$$

Here, each linear form in the second product term is in the linear span of the linear forms $\{L_i : i \in [\tau]\}$, and so can be expressed as their linear combination. Therefore, Q can be expressed as a polynomial in $\{L_i : i \in [\tau]\}$. Let $Q = f(L_1, L_2, \ldots, L_\tau)$. Since we are working over $\mathbb{F}_q$, it follows that for every choice of L_i and for every $\mathbf{a} \in \mathbb{F}_q^n$, we have $L_i^q(\mathbf{a}) = L_i(\mathbf{a})$. So, for every $\mathbf{a} \in \mathbb{F}_q^n$,

$$f(L_1, L_2, \ldots, L_\tau)(\mathbf{a}) = [f(L_1, L_2, \ldots, L_\tau) \mod \langle \{L_i^q - L_i \ : \ i = 1, \ldots, \tau\} \rangle](\mathbf{a})$$

The lemma immediately follows by setting

$$\tilde{Q} := f(L_1, L_2, \ldots, L_\tau) \mod \langle \{L_i^q - L_i \ : \ i = 1, \ldots, \tau\} \rangle .$$

□

High rank gates are almost always zero

Let us assume that $\text{size}(C) \leq 2^{\sqrt{d}/100}$. We shall fix a threshold τ and call all terms T with $\mathsf{Rank}(T) > \tau$ as "high rank terms" and the rest as "low rank terms". Under a random evaluation in $\mathbb{F}_q^n$, every non-zero linear polynomial takes value zero with probability $1/q$. Thus, if we have a term that is a product of *many* independent linear polynomials, then with very high probability *many* of them will be set to zero, i.e. the term will vanish with high multiplicity at most points. This is formalized by the following definition and the lemma after it.

Definition 6.8 (Multiplicity at a point)**.** *For any polynomial P and a point $\mathbf{a} \in \mathbb{F}_q^n$, we shall say that $\mathbf{a}$ vanishes* with multiplicity *t on P if $Q(\mathbf{a}) = 0$ for all $Q \in \partial^{\leq t-1}(P)$. In other words, $\mathbf{a}$ is a root of P and all its derivatives up to order $t-1$.*

We shall denote by $\mathsf{Mult}(P, \mathbf{a})$ the maximum t such that $\mathbf{a}$ vanishes on $\partial^{\leq t-1}(P)$. ◊

It is easy to see that if P is a product of linear polynomials, then $\mathbf{a}$ vanishes with multiplicity t on P if $\mathbf{a}$ vanishes on at least t factors of P.

Observation 6.9. *Let $T = \prod_{i=1}^{d} L_i$ be a term of rank at least r. Then, for every $\delta > 0$,*

$$\Pr_{\mathbf{a} \in \mathbb{F}_q^n} \left[\mathsf{Mult}(T, \mathbf{a}) \leq (1-\delta)\frac{r}{q} \right] \leq \exp\left(-\frac{\delta^2 r}{2q} \right).$$

Proof. Without loss of generality, let $L_1, \ldots, L_r$ be linearly independent. Then, the evaluations of $L_1, \ldots, L_r$ at a point $\mathbf{a} \in \mathbb{F}_q^n$ are also linearly independent and $\Pr_{\mathbf{a}}[L_i(\mathbf{a}) = 0] = (1/q)$ for $i = 1, \ldots, r$.

For $i = 1, \ldots, r$, let Y_i be the indicator random variable that is one if $L_i(\mathbf{a}) = 0$ and zero otherwise. Let $Y = \sum_{i \in [r]} Y_i$. Clearly, by linearity of expectations

$$\mathbb{E}[Y] = \sum_{i \in [r]} \mathbb{E}[Y_i] = \frac{r}{q}.$$

Since the events Y_i are linearly independent, by the Chernoff Bound, we know that for every $\delta > 0$

$$\Pr\left[Y \leq (1-\delta)\frac{r}{q}\right] \quad \leq \quad \exp\left(-\frac{\delta^2 r}{2q}\right). \qquad \square$$

The following corollary is a simple union bound on all high-rank gates in a small circuit.

Corollary 6.10. *Let C be a depth-5 circuit over $\mathbb{F}_q$ such that $\text{size}(C) \leq 2^{\sqrt{d}/100}$. Let $\tau = \frac{q\sqrt{d}}{6}$ so that*

$$\exp\left(\frac{\tau}{8 \cdot q}\right) \quad > \quad 2^{\sqrt{d}/50}.$$

Then,

$$\Pr_{\mathbf{a} \in \mathbb{F}_q^n}\left[\exists T \in \text{Terms}_{>\tau}(C) \ : \ \mathsf{Mult}(T, \mathbf{a}) \ \leq \ \frac{\tau}{2q}\right] \quad \leq \quad 2^{-(\sqrt{d}/100)} \qquad \square$$

We shall set our parameter τ as in the above corollary and our parameter $k = \tau/2q^3$.

6.4.1 Upper bound on complexity measure

For a circuit C of size at most $2^{\sqrt{d}/100}$, let $\mathcal{E}$ refer to the "bad set" of points $\mathbf{a}$ such that there is some $T \in \text{Terms}_{>\tau}(C)$ for which $\mathsf{Mult}(T, \mathbf{a}) \leq k = \tau/2q^3$. By the above corollary, we know that

$$|\mathcal{E}| = \delta \cdot q^n \quad \text{for some } \delta = \exp(-O(\sqrt{d})).$$

Let S be any subset of $\mathbb{F}_q^n \setminus \mathcal{E}$ that is contained in a "translate of a hypercube", that is there exists some $\mathbf{c} \in \mathbb{F}_q^n$ such that

$$S \subset (\mathbf{c} + \mathcal{H}) \setminus \mathcal{E}.$$

The following lemma allows us to "multilinearize" any polynomial as long as we are only interested in evaluations on a translate of a hypercube.

Lemma 6.11 (Multilinearization)**.** *Fix a translate of a hypercube* $\mathbf{c} + \mathcal{H}$*. Then for every polynomial* $Q \in \mathbb{F}_q[\mathbf{x}]$*, there is a unique multilinear polynomial* Q' *such that* $\mathsf{Deg}(Q') \leq \mathsf{Deg}(Q)$ *and* $Q'(\mathbf{a}) = Q(\mathbf{a})$ *for every* $\mathbf{a} \in \mathbf{c} + \mathcal{H}$*.*

Proof. If $\mathbf{a} \in \mathbf{c} + \mathcal{H}$, then for each $i \in [n]$ we have a_i to be either c_i or $c_i + 1$. Thus, it suffices to replace each x_i^2 by a linear polynomial in x_i that maps c_i to c_i^2 and $c_i + 1$ to $(c_i + 1)^2$. This is achieved by $x_i^2 \mapsto c_i^2 + (x_i - c_i)(2c_i + 1)$. By repeated applications of this reduction, we obtain a multilinear polynomial Q' of degree at most $\mathsf{Deg}(Q)$ that agrees on all points on $\mathbf{c} + \mathcal{H}$.

Another way to state this is by looking at $Q \bmod \mathcal{I}_{\mathbf{c}}$ where $\mathcal{I}_{\mathbf{c}}$ is the ideal defined by

$$\mathcal{I}_{\mathbf{c}} \quad := \quad \left\langle \left\{ x_i^2 - (c_i^2 + (x_i - c_i)(2c_i + 1)) \ : \ i = 1, \ldots, n \right\} \right\rangle .$$

It is easy to check that $\mathcal{I}_{\mathbf{c}}$ vanishes on $\mathbf{c} + \mathcal{H}$, and any Q can be reduced to a multilinear polynomial modulo $\mathcal{I}_{\mathbf{c}}$.

The uniqueness of Q' follows from the fact that no non-zero multilinear polynomial can vanish on all of $\mathbf{c} + \mathcal{H}$. □

The main lemma of this theorem would be the following bound on the complexity measure on a depth-5 circuit.

Lemma 6.12 (Upper bound on circuit)**.** *Let* C *be a depth-*5 *circuit, of formal degree at most* $2d$ *and* $\mathrm{size}(C) \leq 2^{\sqrt{d}/100}$*, that computes an* n*-variate degree* d *polynomial. Let* τ *and* k *be chosen as above, and* ℓ *be a parameter satisfying* $\ell + k\tau q < n/2$*. If* S *is any subset of* $\mathbb{F}_q^n \setminus \mathcal{E}$ *that is contained in a translate of a hypercube, then*

$$\Gamma_{k,\ell,S}(C) \quad \leq \quad 2^{\sqrt{d}/100} \cdot \binom{\frac{4d}{\tau} + 1}{k} \cdot \binom{n}{\ell + k\tau q} \cdot \mathrm{poly}(n).$$

Proof. Suppose $C = R_1 + \cdots + R_s$, where $s \leq 2^{\sqrt{d}/100}$ and each R_i is a product of depth-3 circuits with $\mathsf{Deg}(R_i) \leq 2d$. Since $\Gamma_{k,\ell,S}$ is clearly sub-additive (i.e. $\Gamma_{k,\ell,S}(f + g) \leq \Gamma_{k,\ell,S}(f) + \Gamma_{k,\ell,S}(g)$ for any f, g), it suffices to show that for each R_i we have

$$\Gamma_{k,\ell,S}(R_i) \quad \leq \quad \binom{\frac{4d}{\tau} + 1}{k} \cdot \binom{n}{\ell + k\tau q} \cdot \mathrm{poly}(n).$$

For each such R_i, define the $R_i^{\leq\tau}$ as the polynomial obtained by "deleting" all terms $T \in \text{Terms}_{>\tau}(R_i)$. That is,

$$\text{if} \quad R_i \quad = \quad \prod_a \sum_b T_{ab} \quad \text{then} \quad R_i^{\leq\tau} \quad = \quad \prod_a \sum_{b:Rank(T_{ab})\leq\tau} T_{ab}.$$

The lemma follows from the following two claims whose proofs shall be deferred to the end of this section.

Claim 6.13. *For every $i \in [r]$*

$$\Gamma_{k,\ell,S}(R_i) \quad = \quad \Gamma_{k,\ell,S}(R_i^{\leq\tau})$$

Claim 6.14. *For every $i \in [r]$*

$$\Gamma_{k,\ell,S}(R_i^{<\tau}) \quad \leq \quad \binom{\frac{4d}{\tau}+1}{k} \cdot \binom{n}{\ell + k\tau q} \cdot \text{poly}(n)$$

The lemma readily follows from Claim 6.13 and Claim 6.14. □

Proof of Claim 6.13. For brevity, we shall drop some indices and work with $R = Q_1 \cdots Q_m$. Let $T \in \text{Terms}_{>\tau}(C)$. We shall show if $R' = (Q_1 - T)Q_2 \cdots Q_m$, then for any k-th order partial derivative $\partial_{\mathbf{x}^\alpha}$,

$$\mathsf{Eval}_S(\partial_{\mathbf{x}^\alpha}(R)) \quad = \mathsf{Eval}_S(\partial_{\mathbf{x}^\alpha}(R')).$$

Indeed, consider the difference $R - R' = T \cdot Q_2 \cdots Q_m$. By the chain rule, every term in $\partial_{\mathbf{x}^\alpha}(R - R')$ is divisible by some k'-th order partial derivative of T with $k' \leq k$. By the choice of S, we know that every $\mathbf{a} \in S$ satisfies $\mathsf{Mult}(T, \mathbf{a}) > k$, and hence $\mathbf{a}$ vanishes on $\partial^{\leq k}(T)$ for any $T \in \text{Terms}_{>\tau}(C)$. Thus, it follows that $\mathsf{Eval}_S(\partial_{\mathbf{x}^\alpha}(R - R'))$ is just the zero vector.

Repeating this argument, we can prune away all terms in $\text{Terms}_{>\tau}(C)$ to get that $\mathsf{Eval}_S(\partial_\alpha(R)) = \mathsf{Eval}_S(\partial_{\mathbf{x}^\alpha}(R^{\leq\tau}))$ where $\mathsf{Deg}(\mathbf{x}^\alpha) = k$. Thus, $\Gamma_{k,\ell,S}(R) = \Gamma_{k,\ell,S}(R^{<\tau})$. □

Proof of Claim 6.14. Let $R^{\leq\tau} = Q_1 \cdots Q_d$, with each Q_i being a $\Sigma\Pi\Sigma$ circuit. Some of these Q_is could have degree more than τ although their rank is bounded by τ. Without

loss of generality, let $Q_1, \cdots, Q_m$ be all the Q_is with $\mathsf{Deg}(Q_i) > \tau$, and $Q_{m+1}, \ldots, Q_d$ have $\mathsf{Deg}(Q_i) \leq \tau$.

We shall modify the "low-degree" Q_is by multiplying together any two of them of degree less than $\tau/2$. This ensures that at most one of the Q_is may have degree less than $\tau/2$ and all the Q_is have degree at most τ. The sizes of some of the low-degree Q_is do increase in the process but this would not be critical as the degree of any such term is still bounded by τ. The main point is that now we have an expression of the form

$$R^{\leq \tau} = Q_1 \cdots Q_m \cdot Q'_1 \cdots Q'_r$$

where each Q_i is a $\Sigma\Pi\Sigma$ circuit of rank at most $\tau - 1$ and $\mathsf{Deg}(Q_i) \geq \tau$, and all but one of the Q'_i satisfies $\tau \geq \mathsf{Deg}(Q'_i) \geq \tau/2$. As $\mathsf{Deg}(R^{\leq \tau}) \leq 2d$, it follows that $m + r \leq \frac{4d}{\tau} + 1$.

As a notational convenience, for any set H let $Q_H := \prod_{i \in H} Q_i$. Let us look at any partial derivative $\partial_{\mathbf{x}^\alpha}$ of order k applied on R. By the chain-rule, any such partial derivative can be written seen as a natural linear combination of terms.

$$\begin{aligned}
\partial_{\mathbf{x}^\alpha}(R) \quad &\in \quad \mathsf{Span}\Big\{ Q_{\overline{A}} \cdot \partial_{\mathbf{x}^\alpha}(Q_A) \partial_{\mathbf{x}^\beta}(Q_A) \cdot \partial_{\mathbf{x}^\gamma}(Q'_B) \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}} \ : \\
&\qquad\qquad \left.\begin{array}{c} \mathbf{x}^\alpha = \mathbf{x}^\beta \cdot \mathbf{x}^\gamma \ , \ A \subseteq [m] \ , \\ B \subseteq [r] \ , \ |A| + |B| = k \end{array}\right\} \\
&\in \quad \mathsf{Span}\Big\{ \partial_{\mathbf{x}^\beta}(Q_A) \cdot \mathbf{x}^{\leq k\tau} \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}} \ : \\
&\qquad\qquad \left.\begin{array}{c} \mathbf{x}^\alpha = \mathbf{x}^\beta \cdot \mathbf{x}^\gamma \ , \ A \subseteq [m] \ , \\ B \subseteq [r] \ , \ |A| + |B| = k \end{array}\right\} \\
\implies \mathbf{x}^{=\ell} \partial_{\mathbf{x}^\alpha}(R) \quad &\subseteq \quad \mathsf{Span}\Big\{ \partial_{\mathbf{x}^\beta}(Q_A) \cdot \mathbf{x}^{\leq \ell + k\tau} \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}} \ : \\
&\qquad\qquad \left.\begin{array}{c} \mathbf{x}^\alpha = \mathbf{x}^\beta \cdot \mathbf{x}^\gamma \ , \ A \subseteq [m] \ , \\ B \subseteq [r] \ , \ |A| + |B| = k \end{array}\right\} \\
\implies \mathsf{Eval}_S(\mathbf{x}^{=\ell} \partial_{\mathbf{x}^\alpha}(R)) \quad &\subseteq \quad \mathsf{Span}\Big\{ \mathsf{Eval}_S\left(\partial_{\mathbf{x}^\beta}(Q_A) \cdot \mathbf{x}^{\leq \ell + k\tau} \cdot Q_{\overline{A}} \cdot Q'_{\overline{B}} \right) \ : \\
&\qquad\qquad \left.\begin{array}{c} \mathbf{x}^\alpha = \mathbf{x}^\beta \cdot \mathbf{x}^\gamma \ , \ A \subseteq [m] \ , \\ B \subseteq [r] \ , \ |A| + |B| = k \end{array}\right\}
\end{aligned}$$

If we focus on the term $\partial_{\mathbf{x}^\beta}(Q_A)$, since Q_A is a product of $\Sigma\Pi\Sigma$ circuits of rank at most τ, we have that $\partial_{\mathbf{x}^\beta}(Q_A)$ is a linear combination of terms $T_1 \cdots T_{|A|}$ where each T_i is

a product of linear polynomials and has rank at most τ. Using Lemma 6.7 on each of these T_is,

$$\mathsf{Eval}_S(\partial_{\mathbf{x}^\beta}(Q_A)) \in \mathsf{Span}\left\{\mathsf{Eval}_S(\mathbf{x}^{\leq(q-1)\tau|A|})\right\}.$$

Therefore,

$$\begin{aligned}
\mathsf{Eval}_S(\mathbf{x}^{=\ell}\partial_{\mathbf{x}^\alpha}(R)) \quad &\subseteq \quad \mathsf{Span}\Big\{\mathsf{Eval}_S\left(\partial_{\mathbf{x}^\beta}(Q_A)\cdot\mathbf{x}^{\leq\ell+k\tau}\cdot Q_{\overline{A}}\cdot Q'_{\overline{B}}\right) \;:\\
&\qquad\qquad \begin{array}{c}\mathbf{x}^\alpha=\mathbf{x}^\beta\cdot\mathbf{x}^\gamma\;,\; A\subseteq[m]\;,\\ B\subseteq[r]\;,\;|A|+|B|=k\end{array}\;\Big\}\\
&\subseteq \quad \mathsf{Span}\Big\{\mathsf{Eval}_S\left(\mathbf{x}^{\leq\ell+k\tau+(q-1)k\tau}\cdot Q_{\overline{A}}\cdot Q'_{\overline{B}}\right) \;:\\
&\qquad\qquad \begin{array}{c}\mathbf{x}^\alpha=\mathbf{x}^\beta\cdot\mathbf{x}^\gamma\;,\; A\subseteq[m]\;,\\ B\subseteq[r]\;,\;|A|+|B|=k\end{array}\;\Big\}.
\end{aligned}$$

Finally, Lemma 6.11 shows for every polynomial f, there is a multilinear polynomial of degree at most $\mathsf{Deg}(f)$ that agrees with f on all evaluations on a translate of a hypercube. Therefore, in the above span, we may assume that we are only shifting by multilinear monomials of degree $\ell+qk\tau$ as this doesn't change the evaluations $S\subseteq\mathbf{c}+\{0,1\}^n$. Hence,

$$\mathsf{Eval}_S(\mathbf{x}^{=\ell}\partial_{\mathbf{x}^\alpha}(R)) \subseteq \mathsf{Span}\left\{\mathsf{Eval}_S\left(\mathbf{x}_{\text{mult}}^{\leq\ell+qk\tau}\cdot Q_{\overline{A}}\cdot Q'_{\overline{B}}\right) \;:\; \begin{array}{c}\mathbf{x}^\alpha=\mathbf{x}^\beta\cdot\mathbf{x}^\gamma\;,\; A\subseteq[m]\;,\\ B\subseteq[r]\;,\;|A|+|B|=k\end{array}\right\}.$$

Therefore, using the fact that $m+r\leq(4d/\tau)+1$, we get the bound

$$\Gamma_{k,\ell,S}(R) := \mathsf{Dim}\left\{\mathsf{Eval}_S(\mathbf{x}^{=\ell}\partial^{=k}(R))\right\} \quad\leq\quad \binom{\frac{4d}{\tau}+1}{k}\cdot\binom{n}{\ell+qk\tau}\cdot n$$

where the first term corresponds to the number of choices for the subsets A and B, and the last two terms correspond to the number of multilinear monomials of degree at most $\ell+qk\tau$. □

Remark 6.15. Observe that, even if the circuit C is of the form

$$C \quad=\quad \sum_a \prod_{b\in[m]} \sum_c \prod_d L_{abcd}$$

such that $\text{Size}(C)\leq 2^{\sqrt{d}/100}$ and $m=O(\frac{d}{\tau})$, then the upper bound in Lemma 6.12

continues to hold.[4] In particular, the formal degree of C could be much larger than d but if the product fan-in at level two of C is small, then

$$\Gamma_{k,\ell,S}(C) \quad \leq \quad 2^{\sqrt{d}/100} \cdot \binom{O(\frac{d}{\tau})}{k} \cdot \binom{n}{\ell + k\tau q} \cdot \mathrm{poly}(n) \qquad \Diamond$$

We later use this observation to complete the proof of Theorem 6.2 in Section 6.6.

6.5 Lower bound for the complexity measure for an explicit polynomial

Let $\mathcal{E}$ be an arbitrary subset of $\mathbb{F}_q^n$ of size at most $\delta \cdot q^n$. We will choosing a specific set S that shall be a subset of a translate of the hypercube and disjoint from $\mathcal{E}$. We will fix the precise definition of S shortly once we motivate the requirements.

The polynomial for which was shall prove our lower bound would be from the Nisan-Wigderson family. We would have to set our parameters carefully but for now we shall just be intentionally vague and refer to the polynomial as just NW and fix parameters at a later point.

Associated with our measure $\Gamma_{k,\ell,S}(\mathsf{NW})$ is a natural matrix that we shall call $\Lambda(\mathsf{NW})$:

> The rows of $\Lambda(\mathsf{NW})$ are indexed by a derivative $\partial_{\mathbf{x}^\alpha} \in \partial^{=k}$ of order k, and a monomial $\mathbf{x}^\beta$ of degree equal to ℓ. The columns are indexed by all points $\mathbf{a} \in S$. The entry in $(\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}, \mathbf{a})$ is the evaluation of $\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}(\mathsf{NW})$ at the point $\mathbf{a}$.

In other words, the matrix is just the vectors $\mathsf{Eval}_S(\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}(\mathsf{NW}))$ listed as different rows for each choice of $\mathbf{x}^\alpha$ and $\mathbf{x}^\beta$. Therefore,

$$\Lambda(\mathsf{NW}) = \Gamma_{k,\ell,S}(\mathsf{NW}) \tag{6.16}$$

Recall from Lemma 6.11 (multilinearization), as long as we only care about evaluations on a translate of a hypercube, we can assume each row is the evaluations of the

[4] Essentially, in the proof of Claim 6.14, we already have an expression of the form $R^{\leq \tau} = Q_1 \cdots Q_m$ with $m = O\left(\frac{d}{\tau}\right)$ and the rest of the proof proceeds as expected.

multilinearization of $\mathbf{x}^\alpha \cdot \partial_{\mathbf{x}^\beta}(\mathsf{NW})$. This does not change the evaluation on any point $\mathbf{a} \in S \subseteq \mathbf{c} + \mathcal{H}$.

Now any such matrix of evaluations can be naturally factorized as a coefficient matrix and an evaluation matrix.

$C_{k,\ell}(\mathsf{NW})$: Each row is indexed by a derivative $\partial_{\mathbf{x}^\alpha}$ of order k, and a monomial $\mathbf{x}^\beta$ of degree ℓ, and each column is indexed by a multilinear monomial m of degree *at most* $\ell + d - k$ over n variables, and the $(\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}, m)$ entry is the coefficient of monomial m in the multilinearization of $\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}(\mathsf{NW})$ with respect to $\mathbf{c} + \mathcal{H}$ (Lemma 6.11).

$V_t(S)$: Rows are indexed by multilinear monomials of degree at most $t = \ell + d - k$ over n variables, columns are indexed by $\mathbf{a} \in S$ and $(m, \mathbf{a})$ entry is the evaluation monomial m at $\mathbf{a}$.

Clearly, $\Lambda(\mathsf{NW}) = C_{k,\ell}(\mathsf{NW}) \cdot V_t(S)$. Thus if we can get a good lower bound on the ranks of the matrices $C_{k,\ell}(\mathsf{NW})$ and $V_t(S)$ for a suitable set S, we would then be able to lower bound the rank of $\Lambda(\mathsf{NW})$. This is formalized by the following simple linear algebraic fact.

Lemma 6.17 (Rank of products of matrices)**.** *If A, B and C are matrices such that $A = B \cdot C$, then* $\mathsf{Rank}(A) \geq \mathsf{Rank}(B) + \mathsf{Rank}(C) - (\#$ *rows of* $C)$.

6.5.1 Rank of $C_{k,\ell}(\mathsf{NW})$

Let us focus on the matrix $C_{k,\ell}(\mathsf{NW})$ and restrict ourselves a submatrix $C'_{k,\ell}(\mathsf{NW})$ to only those columns whose degree is *exactly* $t = \ell + d - k$, and rows indexed by $(\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha})$ with $\mathbf{x}^\beta$ being a multilinear monomial of degree exactly ℓ.

Since our polynomial NW is multilinear, if we were to read off any row of $C'_{k,\ell}(\mathsf{NW})$, this is just the list of coefficients of all multilinear monomials of $(\mathbf{x}^\beta \cdot \partial_{\mathbf{x}^\alpha}(\mathsf{NW}))$. This is because the multilinearization operation in Lemma 6.11 maps any non-multilinear monomial to a multilinear polynomial of strictly smaller degree and hence these monomials are not included in the columns of $C'_{k,\ell}$.

The key point here is that the matrix $C'_{k,\ell}(\mathsf{NW})$ is just the matrix of *projected shifted partial derivatives* of NW. The results of Kayal et. al [KLSS14a] and Kumar and Saraf [KS17] give a lower bound on the rank of this matrix for a suitable choice of the polynomial, but the lower bound of Kumar and Saraf [KS17] is more relevant as it is true over any field (unlike [KLSS14a] that works only over characteristic zero fields).

Using a tight analysis of the argument in [KS17], that we present in Section 6.8 we obtain the following lemma for the Nisan-Wigderson polynomial for very carefully chosen parameters.

Lemma 6.18 (Tight analysis of [KS17])**.** *For every d and $k = O(\sqrt{d})$ there exists parameters m, e, ε such that $m = \Theta(d^2)$, $n = md$ and $\varepsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$ with*

$$\begin{aligned} m^k &\geq (1+\varepsilon)^{2(d-k)} \\ m^{e-k} &= \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot \mathrm{poly}(m). \end{aligned}$$

For any $\{d, m, e, k, \varepsilon\}$ satisfying the above constraints and for $\ell = \frac{n}{2}(1-\varepsilon)$, over any field $\mathbb{F}$, we have

$$\mathsf{Rank}(C_{k,\ell}(\mathsf{NW}_{d,m,e})) \quad \geq \quad \mathsf{Rank}(C'_{k,\ell}(\mathsf{NW}_{d,m,e})) \quad \geq \quad \binom{n}{\ell+d-k} \cdot \exp(-O(\log^2 d)).$$

6.5.2 Rank of $V_t(S)$

Let $\mathcal{H}_{\leq t}$ refer to elements of $\{0,1\}^n$ of Hamming weight at most t. Our first step would be to choose our set S carefully so that we can maximize the rank of $V_t(S)$.

Observation 6.19. *Let $\mathcal{E}$ be a subset of $\mathbb{F}_q^n$ of size at most $\delta \cdot q^n$. Then for any $0 \leq t \leq n$, there is a vector $\mathbf{c} \in \mathbb{F}_q^n$ such that*

$$|(\mathbf{c} + \mathcal{H}_{\leq t}) \cap \mathcal{E}| \quad \leq \quad \delta \cdot |H_{\leq t}|\,.$$

Proof. Let $\mathbb{1}_{\mathcal{E}}(\mathbf{a})$ be the indicator function that is 1 if $y \in \mathcal{E}$, and 0 otherwise. Then,

$$\delta \quad \geq \quad \mathop{\mathbb{E}}_{y \in \mathbb{F}_q^n}[\mathbb{1}_{\mathcal{E}}(\mathbf{a})] \quad = \quad \mathop{\mathbb{E}}_{c \in \mathbb{F}_q^n}\left[\mathop{\mathbb{E}}_{y \in \mathcal{H}_{\leq t}}[\mathbb{1}_{\mathcal{E}}(\mathbf{c}+\mathbf{a})]\right].$$

Thus, there exists some $\mathbf{c} \in \mathbb{F}_q^n$ that still maintains the inequality. □

Our set would be $S = (\mathbf{c} + \mathcal{H}_{\leq t}) \setminus \mathcal{E}$, which has the property that $|S \cap (\mathbf{c} + \mathcal{H}_{\leq t})| \geq (1-\delta) \cdot |\mathcal{H}_{\leq t}|$ by the above observation, and $S \cap \mathcal{E} = \emptyset$.

Let $V_t(S - \mathbf{c})$ be the matrix whose rows are indexed by the polynomials $m(\mathbf{x} - \mathbf{c})$, where m is a multilinear monomials in variables $\mathbf{x}$ of degree at most t. The columns of $V_t(S - \mathbf{c})$ are indexed by S. We have the following observation.

Observation 6.20. $\mathsf{Rank}(V_t(S)) = \mathsf{Rank}(V_t(S - \mathbf{c}))$.

Proof. For any multilinear monomial m of degree at most t, the polynomial $m(\mathbf{x} - \mathbf{c})$ is multilinear and has degree at most t. Thus clearly, the row-space of $V_t(S - \mathbf{c})$ is contained in the row-space of $V_t(S)$. The converse also holds trivially as the translation is invertible. □

We now prove our next lemma which shows a lower bound on the rank of $V_t(S - \mathbf{c})$.

Lemma 6.21. *For any set $S \subseteq \{0,1\}^n \subset \mathbb{F}_q^n$ and any $0 \leq t \leq n$,*

$$\mathsf{Rank}(V_t(S - \mathbf{c})) \quad = \quad |S|$$

Proof. Since $S \subseteq \mathbf{c} + \mathcal{H}_{\leq t}$, the set $S' := S - \mathbf{c} \subset \mathcal{H}_{\leq t}$. Thus the matrix $V_t(S - \mathbf{c})$ is simply the matrix $V_t(S')$. For any $\mathbf{a} \in \{0,1\}^n$, let $m_{\mathbf{a}}$ refer to the 'characteristic' monomial $\prod_{i:a_i=1} x_i$, and let $m_{\mathbf{0}} = 1$.

Consider the sub-matrix of $V_t(S')$ by restricting to rows indexed by $\{m_{\mathbf{a}} \ : \ \mathbf{a} \in S'\}$. By rearranging the rows and columns in the increasing order of the weight of $\mathbf{a}$, it is evident that the sub-matrix is upper-triangular with ones on the diagonal. It therefore follows that the rank of $V_t(S')$ (which is just $V_t(S - \mathbf{c})$) is at least $|S'| = |S|$. □

Combining Observation 6.20 and Lemma 6.21, we have our required bound on the rank of $V_t(S)$.

Lemma 6.22. *Let $\mathcal{E}$ be an arbitrary subset of $\mathbb{F}_q^n$ of size at most $\delta \cdot q^n$. Then, there exists a set $S \subset \mathbb{F}_q^n \setminus \mathcal{E}$ such that $S \subseteq \mathbf{c} + \mathcal{H}$ for some $\mathbf{c} \in \mathbb{F}_q^n$ for which*

$$\mathsf{Rank}(V_t(S)) \quad \geq \quad (1-\delta) \cdot |\mathcal{H}_{\leq t}| \quad = \quad (1-\delta) \cdot (\#\ \textit{rows of}\ V_t(S))$$

Putting them together

Lemma 6.23 (Rank bound for $\Lambda(\mathsf{NW}_{d,m,e})$)**.** *Let $\mathcal{E}$ be an arbitrary subset of $\mathbb{F}_q^n$ of size at most $\delta \cdot q^n$, with $\delta = \exp(-\omega(\log^2 d))$. Then, there exists a set $S \subset \mathbb{F}_q^n \setminus \mathcal{E}$ such that $S \subseteq \mathbf{c} + \mathcal{H}$ for some $\mathbf{c} \in \mathbb{F}_q^n$ for which*

$$\mathsf{Rank}(\Lambda(\mathsf{NW}_{d,m,e})) \quad \geq \quad \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d))$$

where the parameters d, m, e, k, ℓ are chosen as in Lemma 6.18.

Proof. Consider the set S chosen in Lemma 6.22 (for $t = \ell + d - k$). By Lemma 6.22,

$$\mathsf{Rank}(V_t(S)) - (\#\text{ rows of } V_t(S)) \quad \leq \quad (-\delta)\,|\mathcal{H}_{\leq t}| \quad \leq \quad (-\delta) \cdot n \cdot \binom{n}{\ell + d - k}$$

Lemma 6.18 shows that rank of $C_{k,\ell}(\mathsf{NW}_{d,m,e})$ can be lower bounded by

$$\mathsf{Rank}(C_{k,\ell}(\mathsf{NW}_{d,m,e})) \quad \geq \quad \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d))$$

Thus, since $\Lambda(\mathsf{NW}_{d,m,e}) = C_{k,\ell}(\mathsf{NW}_{d,m,e}) \cdot V_t(S)$ with $t = \ell + d - k$, Lemma 6.17 implies that

$$\begin{aligned}
\mathsf{Rank}(\Lambda(\mathsf{NW}_{d,m,e})) &\geq \mathsf{Rank}(C_{k,\ell}(\mathsf{NW}_{d,m,e})) + \mathsf{Rank}(V_t(S)) - (\#\text{ rows of } V_t(S)) \\
&\geq \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d)) - \delta \cdot n \cdot \binom{n}{\ell + d - k} \\
&\geq \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d)) \qquad \text{as } \delta = \exp(-\omega(\log^2 d)).\square
\end{aligned}$$

Combining this with Equation 6.16, we get the following lemma.

Lemma 6.24 (Rank bound for $\Lambda(\mathsf{NW}_{d,m,e})$)**.** *Let $\mathcal{E}$ be an arbitrary subset of $\mathbb{F}_q^n$ of size at most $\delta \cdot q^n$, with $\delta = \exp(-\omega(\log^2 d))$. Then, there exists a set $S \subset \mathbb{F}_q^n \setminus \mathcal{E}$ such that $S \subseteq \mathbf{c} + \mathcal{H}$ for some $\mathbf{c} \in \mathbb{F}_q^n$ for which*

$$\mathsf{Rank}(\Gamma_{k,\ell,S}(\mathsf{NW}_{d,m,e})) \quad \geq \quad \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d))$$

6.6 Wrapping up the proof

Theorem 6.25. *Let $\mathbb{F}_q$ be the finite field of cardinality q. Let C be a depth-5 circuit of formal degree at most $2d$ which computes the polynomial $\mathsf{NW}_{d,m,e}$ with parameters as*

in Lemma 6.18. Then

$$Size(C) > 2^{\sqrt{d}/100}.$$

Further, the same lower bound holds even if C was a circuit of the form

$$C \quad = \quad \sum_{i} \prod_{j \in [m]} \sum_{k} \prod_{\ell} L_{ijk\ell}$$

with $m = O(\sqrt{d})$.

Proof. We shall prove the above theorem for homogeneous depth-5 circuits. The lower bound for such non-homogeneous circuits would also follow directly since such circuits also have the same upper-bound on the complexity measure (Remark 6.15).

Assume on the contrary that there is a circuit C computing $\mathsf{NW}_{d,m,e}$ with $\text{Size}(C) \leq 2^{\sqrt{d}/100}$. Let τ be as defined in Corollary 6.10 and let $k = \tau/2q^3$. Let $\mathcal{E} = \mathcal{E}(\mathcal{C})$ be the set as defined in Subsection 6.4.1. We know that

$$|\mathcal{E}| \leq \delta \cdot q^n$$

for some $\delta = \exp(-O(\sqrt{d}))$. Let $\ell = \frac{n}{2}(1-\varepsilon)$ where $\varepsilon = \frac{\log d}{c\sqrt{d}}$ is chosen as in Lemma 6.18. Since $n = d^3$, clearly we have satisfy $\ell + k\tau q < n/2$. Let $S \subset \mathbb{F}_q^n \setminus \mathcal{E}$ be the set guaranteed by Lemma 6.24. From Lemma 6.24, we know that

$$\Gamma_{k,\ell,S}(\mathsf{NW}) \geq \binom{n}{\ell + d - k} \cdot \exp(-O(\log^2 d))$$

This may be simplified using Lemma 6.37 to

$$\Gamma_{k,\ell,S}(\mathsf{NW}) \geq \binom{n}{\ell} \cdot (1+\varepsilon)^{2d-2k} \cdot \exp(-O(d\varepsilon^2)) \cdot \exp(-O(\log^2 d))$$

Also, from Lemma 6.12, we know that

$$\Gamma_{k,\ell,S}(C) \leq 2^{\sqrt{d}/100} \cdot \binom{\frac{4d}{\tau} + 1}{k} \cdot \binom{n}{\ell + qk\tau} \cdot \text{poly}(n)$$

Again, using Lemma 6.37, we get

$$\Gamma_{k,\ell,S}(C) \leq 2^{\sqrt{d}/100} \cdot \binom{\frac{4d}{\tau} + 1}{k} \cdot \binom{n}{\ell} \cdot (1+\varepsilon)^{2qk\tau} \cdot \exp(O(-qk\tau \cdot \varepsilon^2)) \cdot \text{poly}(n)$$

Since C computes $\mathsf{NW}_{d,m,e}$, so it must be the case that

$$2^{\sqrt{d}/100} \cdot \mathrm{poly}(n) \quad \geq \quad (1+\varepsilon)^{(d-k)+(d-k-2qk\tau)} \cdot \exp(-O_q(\log^2 d))$$

Since $k = \tau/2q^3$, so $2qk\tau = \tau^2/q^2$. From our choice of τ in Corollary 6.10, $\tau = \frac{q\sqrt{d}}{6}$. So

$$2qk\tau = \tau^2/q^2 = d/36$$

Therefore,

$$2^{\sqrt{d}/100} \cdot \mathrm{poly}(n) \geq (1+\varepsilon)^{(d-k)} \cdot \exp(-O_q(\log^2 d))$$

But this is a contradiction since $(1+\varepsilon)^{(d-k)} = \exp(\Omega(\sqrt{d}\log d))$ by our choice of parameters. Therefore, the size of C is at least $2^{\sqrt{d}/100}$. □

In fact, the above proof gives more. It shows that if we have a depth-5 circuit computing $\mathsf{NW}_{d,m,e}$ over $\mathbb{F}_q$, then either the number of high-rank terms is at least $2^{\sqrt{d}/50}$ or the top fan-in is $\exp(\Omega(\sqrt{d}\log d))$.

6.6.1 Getting the right order of quantifiers

In our proof so far, we first fix the field $\mathbb{F}_q$ that we are working over and the parameters of our polynomial are then chosen based on q. Thus, as q varies, the polynomial for which we show the lower bound also seems to vary. The ideal scenario would be to construct a fixed polynomial family so that for every q we get a lower bound of $\exp(\Omega_q(\sqrt{d}))$. We do that now, and this would complete the proof of Theorem 6.1.

We shall be dealing with a lot of parameters and constraints. The following is essentially the "zone" in which we can prove strong lower bounds (Lemma 6.18).

Definition 6.26 (Goldilocks Zone)**.** *We shall say that parameters m, d, e, k, ε with $k = \Theta(\sqrt{d})$ and $\varepsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$ lie in the* Goldilocks Zone *if they satisfy*

$$\begin{aligned} m^k &\geq (1+\varepsilon)^{2(d-k)} \\ m^{e-k} &= \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot \mathrm{poly}(m). \end{aligned}$$

◊

Recall that for Lemma 6.18, and consequently Theorem 6.25, the parameters m, d, e, k must lie in the Goldilocks zone. The crucial point is that this is a field dependent condition since k (and hence everything else) explicitly depends on q. In the next lemma, we show that we can start with a fixed polynomial such that for every finite field $\mathbb{F}_q$ of fixed size, there exists a $0,1$ projection which lies in the Goldilocks zone.

Lemma 6.27. *Consider the* $\mathsf{NW}_{d,m,e}$ *polynomial with* $m = \Theta(d^2)$ *and* e *chosen so that*

$$m^e \quad = \quad 2^d \cdot \mathrm{poly}(m).$$

Suppose $k = \Theta(\sqrt{d})$ *and* $\varepsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$ *satisfy the constraint* $m^k > (1+\varepsilon)^{2(d-k)}$*. Then, there exists a* $d' \in [d - O(\sqrt{d}\log d), d]$ *such that* $\mathsf{NW}_{d',m,e}$ *is a* $0/1$ *projection of* $\mathsf{NW}_{d,m,e}$ *and the parameters* $\{d', m, e, k, \varepsilon\}$ *fall in the Goldilocks Zone.*

Proof. Since $m^e = 2^d \cdot \mathrm{poly}(m)$, $m^k > (1+\varepsilon)^{2(d-k)}$ and $(1+\varepsilon)^d = \exp(\Theta(\sqrt{d}\log d))$, we have

$$m^{e-k} \quad = \quad \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot \exp(-\Theta(\sqrt{d}\log d)).$$

Since the slack in m^{e-k} is just $\exp(\Theta(\sqrt{d}\log d))$ (when compared to the desired value in Definition 6.26), there exists some $d' \in [d - O(\sqrt{d}\log d), d]$ such that

$$m^{e-k} \quad = \quad \left(\frac{2}{1+\varepsilon}\right)^{d'-k} \cdot \mathrm{poly}(m).$$

Further, since $m^k > (1+\varepsilon)^{2(d-k)}$, it follows that $m^k > (1+\varepsilon)^{2(d'-k)}$ as $d' < d$. Hence the parameters $\{d', m, e, k, \varepsilon\}$ indeed fall in the Goldilocks Zone (Definition 6.26).

It suffices to show that $\mathsf{NW}_{d',m,e}$ is a projection of $\mathsf{NW}_{d,m,e}$. This is readily seen as setting the variables $x_{ij} = 1$ for all $i \in [d - d']$ and $j \in [m]$ yields $\mathsf{NW}_{d',m,e}$ up to relabelling variables. □

With this, we can finally prove our main theorems.

Theorem 6.28 (Theorem 6.1 restated)**.** *Consider the polynomial* $\mathsf{NW}_{d,m,e}$ *with parameters chosen such that* $m = \Theta(d^2)$ *and* $m^e = 2^d \cdot \mathrm{poly}(m)$*. Then, for any fixed finite field* $\mathbb{F}_q$*, any homogeneous depth-*5 *circuit over* $\mathbb{F}_q$ *computing* $\mathsf{NW}_{d,m,e}$ *must have size at least* $2^{\sqrt{d}/200}$*.*

Proof. Fix a field $\mathbb{F}_q$ and let $k = \sqrt{d}/12q^3$.

Suppose on the contrary that there is indeed a homogeneous depth-5 circuit C computing $\mathsf{NW}_{d,m,e}$. Then, by Lemma 6.27, this also implies there is a projection C' that computes the $\mathsf{NW}_{d',m,e}$ such that there is an $d - O(\sqrt{d}\log d) \leq d' \leq d$ and there is an $\varepsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$ for which $\{d', m, e, k, \varepsilon\}$ fall in the Goldilocks Zone (Definition 6.26). Now C' is a circuit formal degree $d \leq d' + O(\sqrt{d}\log d) \leq 2d'$ that computes the polynomial $\mathsf{NW}_{d',m,e}$. By Theorem 6.25, this implies that

$$\text{size}(C) \geq \text{size}(C') > 2^{\sqrt{d'}/100} > 2^{\sqrt{d}/200}. \qquad \square$$

The proof of this theorem also follows along the same lines.

Theorem 6.29 (Theorem 6.2 restated)**.** *Consider the polynomial $\mathsf{NW}_{d,m,e}$ with parameters chosen such that $m = \Theta(d^2)$ and $m^e = 2^d \cdot \text{poly}(m)$. Then, for any fixed finite field $\mathbb{F}_q$, any depth-5 circuit over $\mathbb{F}_q$ of the form*

$$C \quad = \quad \sum_i \prod_{j\in[m]} \sum_k \prod_\ell L_{ijk\ell}$$

where each $L_{ijk\ell}$ is a linear polynomial and $m = O(\sqrt{d})$ that computes $\mathsf{NW}_{d,m,e}$ must have size at least $2^{\sqrt{d}/200}$. $\square$

6.7 Discussion

6.7.1 Connections between arithmetic circuits over $\mathbb{F}_q$ and $\mathsf{AC}^0[\text{mod}q]$

Although constant depth arithmetic circuits over $\mathbb{F}_q$ appear to be similar to the class $\mathsf{AC}^0[\text{mod}q]$, they are surprisingly very different with respect to functions computed by them. A striking example, due to Agrawal, Allender and Datta [AAD00], is that arithmetic circuits over $\mathbb{F}_3$ can "compute" both the Mod3 function, as well as the Mod2 function via

$$\text{Mod2}(x_1, \ldots, x_n) \quad = \quad \left(2 + \prod_{i=1}^{n}(1 + x_i)\right)^2.$$

However, it is true that functions computed by arithmetic circuits over $\mathbb{F}_{p^k}$ have strong connections with $\mathsf{AC}^0[\bmod p(p^k - 1)]$ but unless we are working over $\mathbb{F}_2$ it seems difficult

lift a lower bound for $\mathrm{AC}^0[\mathrm{mod}p]$ to arithmetic circuits over $\mathbb{F}_p$. For more on this, see [AAD00].

The only exception we know of is the result of Grigoriev and Razborov [GR00] where they lift Smolensky's [Smo87] lower bound for $\mathrm{AC}^0[\mathrm{mod}\,p]$ to depth-3 arithmetic circuits over $\mathbb{F}_p$, and this crucially uses the fact that depth-3 arithmetic circuits can be point-wise approximated by a "sparse polynomial". But in general, constant depth arithmetic circuits over $\mathbb{F}_p$ and boolean circuits in $\mathrm{AC}^0[\mathrm{mod}p]$ seem to be two very different classes.

6.7.2 Finer separations for bounded depth circuits ?

In [KS15d], it was shown that homogeneous depth-4 circuits are exponentially more powerful than homogeneous depth-4 circuits with bounded bottom fan-in. A natural question to ask is whether such separations can be shown between homogeneous depth-4 circuit and homogeneous depth-5 circuits. One of the first strategies to attempt for this question would be to try and show that there is a homogeneous depth-5 circuit such that its projected shifted partial derivative complexity is quite large. The results in this chapter show that the measure can not to be too close to the largest possible value, in particular it needs to be at least a factor $\exp(\Omega(\sqrt{d}))$ away from the largest possible value. If this is bound is tight, then such a separation between homogeneous depth-5 circuits and homogeneous depth-4 circuits can still be shown using projected shifted partial derivatives. However, it is not clear if this is the case. As mentioned before, even the known lower bounds on the dimension of the projected shifted partials for the IMM seem a factor $\exp\left(\Omega(\sqrt{d}\log d)\right)$ away from the largest possible value.

6.7.3 The tightness of the results and relevance to VP vs VNP

For homogeneous depth-4 circuits, we know $\exp\left(\Omega(\sqrt{d}\log d)\right)$ lower bounds [KLSS14a, KS17] and any asymptotic improvement in the exponent would imply general arithmetic circuit lower bounds. In this sense, the lower bounds for homogeneous depth-4 circuits are tight, over all fields. It is natural to ask, if the bounds in this chapter are tight in this sense? The answer to this question is far from obvious to us. In particular, it is

not clear if we can use computational advantage of having linear forms at the bottom level of the circuit to get a better depth reduction from VP to homogeneous depth-5 circuits, when compared to depth reduction to homogeneous depth-4 circuits.

6.7.4 Lower bounds over fields of characteristic zero ?

One might wonder if the techniques in this chapter could be potentially adapted to work for depth-5 circuits over fields of characteristic zero. As in the work of Grigoriev and Karpinski [GK98], our proof (Lemma 6.12 in particular) strongly relies on the fact that we are working over a fixed finite field, so it clearly seems hard to generalize over large fields (even when the characteristic is small). In addition to this obvious technical obstruction to generalizing the proof in this chapter, there seems to be another reason why the proof strategy in this chapter could be hard to replicate over fields of characteristic zero, namely, an analog of Theorem 6.2 over fields of characteristic zero would imply that $\mathsf{VP} \neq \mathsf{VNP}$. The reason is that over characteristic zero fields, one can obtain better depth reductions to non-homogeneous depth-5 circuits by combining [AV08, Koi12, Tav15] with [GKKS13]. Although this is reasonably well known, we give a formal proof here for completeness.

The following lemma is a simple generalization of the proof of depth reduction to depth-4 circuits by Tavenas [Tav15].

Lemma 6.30 (Depth reduction to homogeneous depth six circuits)**.** *Let P be a polynomial of degree d in* $\mathrm{poly}(d)$ *variables which can be computed by an arithmetic circuit C of size* $\mathrm{poly}(d)$*. Then, there is a homogeneous depth-6 circuit C' which computes P and satisfies*

- *Size(C)* $\leq \exp\left(O(d^{1/3}\log d)\right)$*, and*
- *The fan-in of all the product gates in C' is bounded by $O(d^{1/3})$.*

Now, we start with the circuit C' as guaranteed by the lemma above, and for each of the product gates at the second level, look at its inputs. Each such input is a $\Sigma\Pi^{O(d^{1/3})}\Sigma\Pi^{O(d^{1/3})}$ circuit (depth-4 circuit with all product fan-ins being at most

$O(d^{1/3})$) of size at most $\exp(O(d^{1/3}\log d))$. We now apply the depth reduction to depth-3 by Gupta et al. [GKKS13] to each one of these depth-4 circuits. As a result, each of these depth-4 circuits get reduced to a depth-3 circuit, with at most a factor of $\exp(O(d^{1/3}))$ blow up in size. Plugging these depth-3 circuits back into C', we obtain a depth-5 circuit C'' such that

- Size(C) $\leq \exp(O(d^{1/3}\log d))$, and
- The fan-in of all the product gates at level two of C'' is bounded by $O(d^{1/3})$.

Recall that the depth reduction in[GKKS13] only works over fields of characteristic zero. This yields the following depth reduction to non-homogeneous depth-5 circuits.

Lemma 6.31 (Depth reduction to non-homogeneous depth-5 circuits)**.** *Let $\mathbb{F}$ be a field of characteristic zero. Let P be a polynomial of degree d in* poly(d) *variables over $\mathbb{F}$ which can be computed by an arithmetic circuit C of size* poly(d)*. Then, there is a depth-5 circuit C'' which computes P and satisfies*

- *Size(C)* $\leq \exp(O(d^{1/3}\log d))$*, and*
- *The fan-in of all the product gates at level two of C' is bounded by $O(d^{1/3})$.*

Now, observe that an analogue of Theorem 6.2 over fields of characteristic zero, would imply an $\exp(\Omega(d^{1/2}))$ lower bound for the depth-5 circuits obtained in Lemma 6.31, and hence imply $\mathsf{VP} \neq \mathsf{VNP}$.

6.8 Tight analysis of the [KS17] lower bound

We recall the measure of *projected shifted partial derivatives* that was used in [KLSS14a] and [KS17].

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad = \quad \mathsf{Dim}\left\{\mathrm{mult}\left(\mathbf{x}^{=\ell}\partial^{=k}(P)\right)\right\}$$

where mult(f) is just the polynomial f restricted to just its multilinear monomials. As mentioned before, this $\Gamma^{\mathrm{PSD}}_{k,\ell}(P)$ is precisely $\mathsf{Rank}(C'_{k,\ell}(P))$ as defined in Subsection 6.5.1.

The goal of this section would be to prove Lemma 6.18 that we restate below.

Lemma. *For every d and $k = O(\sqrt{d})$ there exists parameters m, e, ε such that $m = \Theta(d^2)$, $n = md$ and $\varepsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$ with*

$$\begin{aligned} m^k &\geq (1+\varepsilon)^{2(d-k)} \\ m^{e-k} &= \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot \mathrm{poly}(m). \end{aligned}$$

For any $\{d, m, e, k, \varepsilon\}$ satisfying the above constraints, the polynomial $\mathsf{NW}_{d,m,e}$, if $\ell = \frac{n}{2}(1-\varepsilon)$, then over any field $\mathbb{F}$, we have

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(\mathsf{NW}_{d,m,e}) \quad \geq \quad \binom{n}{\ell+d-k} \cdot \exp(-O(\log^2 d)).$$

The rest of this section would just be a proof of this lemma.

Before we proceed to the lower bound on $\Gamma^{\mathrm{PSD}}_{k,\ell}(\mathsf{NW}_{d,m,e})$, let us first show that we can indeed find parameters that satisfy the above constraints. Fix m to be the smallest power of 2 greater than d^2 to get $m = \Theta(d^2)$. Next, we shall fix the constant c in $\varepsilon = \frac{\log d}{c\sqrt{d}}$ so that

$$m^k \quad \geq \quad (1+\varepsilon)^{2(d-k)}$$

This is always possible by choosing c to be large enough as $(1+\varepsilon)^{d-k} = \exp(O(\sqrt{d}\log d))$ and that is also the order of m^k.

Once we have done that, we shall fix e so as to ensure that

$$m^{e-k} \quad = \quad \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot \mathrm{poly}(m)$$

This is always possible because choosing $e = k$ makes the LHS $<$ RHS and choosing $e = m$ makes LHS $>$ RHS. Hence, there must be an integer e such that LHS and RHS are within a multiplicative factor of m.

All lower bounds on the dimension of shifted partial derivatives of a polynomial P was obtained by finding a *large* set of *distinct leading monomials*. In [KS17], they take this approach but require a very careful analysis. The key difference in this setting is the following:

If β is the leading monomial of a polynomial P, then for any monomial γ, we also have that $\beta \cdot \gamma$ is the leading monomial of γP.

However, the leading monomial of $\text{mult}(\gamma P)$ could be $\beta' \cdot \gamma$ for some $\beta' \neq \beta$ (as higher monomials could be made non-multilinear during the shift by γ).

The multilinear projection makes the task of counting leading monomials much harder and [KS17] come up with an indirect way to count them. Throughout this discussion, let $\text{LM}(f)$ refer to the leading monomial of f in some natural ordering, say the lexicographic order.

Leading monomials after multilinear projections

Let P the polynomial of degree d for which we are trying to lower bound $\Gamma^{\text{PSD}}_{k,\ell}(P)$. For every monomial multilinear monomial α of degree k, and a monomial $\beta \in \partial_\alpha(P)$, define the set $A(\alpha, \beta)$ as

$$A(\alpha,\beta) \quad = \quad \left\{ \gamma \ : \ \begin{array}{c} \mathsf{Deg}(\gamma) = \ell + d - k \text{ and there is a } \gamma' \text{ of degree } \ell \\ \text{such that } \gamma = \text{LM}(\text{mult}(\gamma' \cdot \partial_\alpha(P))) = \gamma' \cdot \beta \end{array} \right\}$$

In other words, we want the number of distinct monomials that are contributed by β, which are also distinct leading monomials obtained from $\partial_\alpha(P)$ that are divisible by β. We then have

$$\Gamma^{\text{PSD}}_{k,\ell}(P) \quad \geq \quad \left| \bigcup_{\alpha,\beta} A(\alpha,\beta) \right|$$

Choice of derivatives: Instead of looking at all derivatives in $\partial^{=k}$, we shall restrict ourselves to just a subset of derivatives. Restricting the above union to a subset $\Delta \subset \mathbf{x}^{=k}$ still continues to remain a lower bound for $\Gamma^{\text{PSD}}_{k,\ell}(P)$. Keeping in mind that we are dealing with $P = \mathsf{NW}_{d,m,e}$ and that $m^k > (1+\varepsilon)^{2(d-k)}$. We shall choose Δ to be a set of size exactly $(1+\varepsilon)^{2(d-k)}$ which consists of monomials of the form $x_{1a_1} \cdots x_{ka_k}$ with each $a_i \leq m$. This shall become relevant later.

$$\Gamma^{\text{PSD}}_{k,\ell}(P) \quad \geq \quad \left| \bigcup_{\substack{\alpha \in \Delta \\ \beta \in \mathbf{x}^{=\ell}}} A(\alpha,\beta) \right| \tag{6.32}$$

We shall need the following lemma from [KS17] that is a strengthening of the standard Inclusion-Exclusion principle.

Lemma 6.33 (Stronger Inclusion-Exclusion [KS17])**.** *Let $A_1, \ldots, A_r$ be sets such that there is some $\lambda > 1$ such that*

$$\sum_{i \neq j} |A_i \cap A_j| \quad \leq \quad \sum_i \lambda \cdot |A_i|$$

Then,

$$\left|\bigcup_i A_i\right| \quad \geq \quad \left(\frac{1}{4\lambda}\right) \cdot \left(\sum_i |A_i|\right)$$

Corollary 6.34. *Considers sets $A_1, \ldots, A_r$ and let $S_1 = \sum_i |A_i|$ and $S_2 = \sum_{i \neq j} |A_i \cap A_j|$. Then,*

$$\left|\bigcup A_i\right| \quad \geq \quad \frac{S_1}{4} \cdot \min\left(1, \frac{S_1}{S_2}\right)$$

Estimating $|\bigcup A(\alpha, \beta)|$ via Inclusion-Exclusion

$$\left|\bigcup_{\alpha,\beta} A(\alpha,\beta)\right| \quad \geq \quad \sum_{\alpha,\beta} |A(\alpha,\beta)| \quad - \quad \sum_{(\alpha,\beta)\neq(\alpha',\beta')} \left|A(\alpha,\beta) \cap A(\alpha',\beta')\right|$$

Let us first address the term $\sum |A(\alpha, \beta)|$. As mentioned earlier, it is not an easy task to get a good handle on the set $A(\alpha, \beta)$ for polynomial such as NW, for any reasonable monomial ordering. However, [KS17] circumvent this difficult by using an indirect approach to estimate this term.

For any derivative α and $\beta \in \partial_\alpha(P)$, define the set $S(\alpha, \beta)$ as the following set of multilinear monomials of degree ℓ that is disjoint from β.

$$S(\alpha,\beta) \quad = \quad \left\{ \gamma \ : \ \begin{array}{c} \gamma \text{ is multilinear, has} \\ \text{degree } \ell \text{ and } \gcd(\beta,\gamma) = 1 \end{array} \right\}$$

This on the other hand is independent of any monomial ordering, and is also easy to calculate:

$$\text{For every } \alpha, \beta \qquad |S(\alpha,\beta)| \quad = \quad \binom{n-d+k}{\ell}.$$

Lemma 6.35 ([KS17])**.** *For any α,*

$$\sum_{\beta} |A(\alpha,\beta)| \quad \geq \quad \left|\bigcup_{\beta} S(\alpha,\beta)\right|$$

Proof. Consider any $\gamma \in \bigcup_\beta S(\alpha,\beta)$. By definition, there is at least one non-multilinear monomial in $\gamma \cdot \partial_\alpha(P)$. Thus, in particular LM(mult($\gamma \cdot \partial_\alpha(P)$) is non-zero and equal to some $\gamma \cdot \beta$ for some monomial $\beta \in \partial_\alpha(P)$. This also implies that $\gamma' = \gamma \cdot \beta \in A(\alpha,\beta)$. This yields an injective map ϕ

$$\phi : \bigcup_{\beta} S(\alpha,\beta) \quad \rightarrowtail \quad \left\{(\beta,\gamma') \ : \ \beta \in \partial_\alpha(P) \ , \ \gamma' \in A(\alpha,\beta)\right\}$$

Since the size of the RHS is precisely $\sum_\beta |A(\alpha,\beta)|$, the lemma follows. □

Thus, by another use of Inclusion-Exclusion on the $S(\alpha,\beta)$'s, we get

$$\begin{aligned}
\left|\bigcup_{\alpha,\beta} A(\alpha,\beta)\right| \ &\geq \ \sum_{\alpha,\beta} |A(\alpha,\beta)| \quad - \sum_{(\alpha,\beta)\neq(\alpha',\beta')} \left|A(\alpha,\beta) \cap A(\alpha',\beta')\right| \\
&\geq \ \sum_{\alpha} \left(\sum_{\beta} |S(\alpha,\beta)|\right) \quad - \quad \sum_{\alpha} \left(\sum_{\beta\neq\beta'} \left|S(\alpha,\beta) \cap S(\alpha,\beta')\right|\right) \\
&\qquad - \sum_{(\alpha,\beta)\neq(\alpha',\beta')} \left|A(\alpha,\beta) \cap A(\alpha',\beta')\right|
\end{aligned}$$

Let us call the three terms in the RHS of the last equation as T_1, T_2 and T_3 respectively. Since we know the size of each $S(\alpha,\beta)$ exactly, the value of T_1 is easily obtained.

Lemma 6.36 ([KS17])**.**

$$T_1(\alpha) \quad := \quad \sum_{\beta} |S(\alpha,\beta)| \quad = \quad (\#\ \textit{mons in a deriv}) \cdot \binom{n-d+k}{\ell}$$

We shall be simplifying such binomial coefficients very often.

Lemma 6.37. *Let n and ℓ be parameters such that $\ell = \frac{n}{2}(1-\varepsilon)$ for some $\varepsilon = o(1)$. For any a, b such that $a, b = O(\sqrt{n})$,*

$$\binom{n-a}{\ell-b} \quad = \quad \binom{n}{\ell} \cdot 2^{-a} \cdot (1+\varepsilon)^{a-2b} \cdot \exp(O(b \cdot \varepsilon^2))$$

Proof. The proof of the above lemma would repeatedly use the fact that $n! = (n-a)! \cdot n^a \cdot \mathrm{poly}(n)$ whenever $a = O(\sqrt{n})$ (see [GKKS14, Lemma 3.4]).

$$\begin{aligned}
\frac{\binom{n-a}{\ell-b}}{\binom{n}{\ell}} &= \frac{(n-a)!}{n!} \cdot \frac{\ell!}{(\ell-b)!} \cdot \frac{(n-\ell)!}{(n-\ell-a+b)!} \\
&\overset{\text{poly}}{\approx} \frac{1}{n^a} \cdot \ell^b \cdot \frac{(n-\ell)^a}{(n-\ell)^b} \\
&= \frac{\left(\frac{n}{2}\right)^a (1+\varepsilon)^a}{n^a} \cdot \frac{(1-\varepsilon)^b}{(1+\varepsilon)^b} \\
&= 2^{-a} \cdot (1+\varepsilon)^{a-2b} \cdot \exp(O(b \cdot \varepsilon^2))
\end{aligned}$$

□

Since our of parameters would be $\varepsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$, the bound on T_1 can be simplified as

$$\begin{aligned}
T_1(\alpha) &= (\#\text{ mons in a deriv}) \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{d-k} \cdot \exp(-O(\log^2 d)) \\
&= m^{e-k} \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{d-k} \cdot \exp(-O(\log^2 d)) \\
&= \binom{n}{\ell} \cdot \exp(-O(\log^2 d))
\end{aligned}$$

where we used the fact that every non-zero k-th order derivative of $\mathsf{NW}_{d,m,e}$ has exactly m^{e-k} monomials and our setting of parameters.

Remark. *To avoid writing this factor of* $\exp(O(\log^2 d))$*, we shall use* $\approx$ *of* $\gtrsim$ *or* $\lesssim$ *to indicate that a factor* $\exp(O(\log^2 d))$ *is omitted.* ◊

We now move on to the calculation of T_2. This is the first place where the choice of the polynomial and parameters becomes crucial.

Lemma 6.38 ([KS17])**.** *For the polynomial* $P = \mathsf{NW}_{d,m,e}$*, if* $n = md$ *and* $\ell = \frac{n}{2}(1-\varepsilon)$ *for* $\varepsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$*, for any* $\alpha \in \Delta$

$$T_2(\alpha) := \sum_{\beta \neq \beta'} \left|S(\alpha,\beta) \cap S(\alpha,\beta')\right| \lesssim m^{2(e-k)} \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{2d-2k}$$

Proof. Recall that $S(\alpha,\beta) \cap S(\alpha,\beta')$ is just set of all multilinear monomials γ of degree ℓ that are disjoint from both β and β'. Hence, for any pair of multilinear degree $(d-k)$ monomials $\beta \neq \beta' \in \partial_\alpha(P)$ such that $\mathsf{Deg}(\gcd(\beta,\beta')) = t$,

$$\left|S(\alpha,\beta) \cap S(\alpha,\beta')\right| = \binom{n-2d+2k+t}{\ell}$$

Thus, if we can count the number of pairs (β, β') that agree on exactly t places, we can obtain $T_2(\alpha)$. Note that for $\mathsf{NW}_{d,m,e}$, any two $\beta, \beta' \in \partial_\alpha(\mathsf{NW}_{d,m,e})$ can agree on at most $e-k$ places. Further, the number of pairs that agree in exactly $0 \le t \le e-k$ places is at most

$$m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-k-t}$$

as there are m^{e-k} choices for β, and $\binom{d-k}{t}$ choices for places where they may agree, and $(m-1)^{e-k-t}$ choices for β' that agree with β on those t places. Thus,

$$\begin{aligned}
T_2(\alpha) &\le \sum_{t=0}^{e-k} m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-k-t} \cdot \binom{n-2d+2k+t}{\ell} \\
&\approx \sum_{t=0}^{e-k} m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-k-t} \cdot \binom{n}{\ell} \frac{1}{2^{2d-2k-t}} \cdot (1+\varepsilon)^{2d-2k-t} \\
&\le m^{2(e-k)} \binom{n}{\ell} \left(\frac{1+\varepsilon}{2}\right)^{2d-2k} \cdot \sum_{t=0}^{e-k} \binom{d-k}{t} \left(\frac{2}{(1+\varepsilon)m}\right)^t \\
&\le m^{2(e-k)} \binom{n}{\ell} \left(\frac{1+\varepsilon}{2}\right)^{2d-2k} \cdot \left(1 + \frac{2}{(1+\varepsilon)m}\right)^{d-k} \\
&= m^{2(e-k)} \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{2d-2k} \cdot O(1) \qquad \text{if } m = \Omega(d) \qquad \square
\end{aligned}$$

Combining this with Lemma 6.36 and using Inclusion-Exclusion (Corollary 6.34), we get that for every $\alpha \in \Delta$,

$$\begin{aligned}
\left|\bigcup_\beta S(\alpha,\beta)\right| &\gtrsim T_1(\alpha) \cdot \min\left(1, \frac{T_1(\alpha)}{T_2(\alpha)}\right) \\
&\approx T_1(\alpha) \cdot \min\left(1, \frac{\left(\frac{2}{1+\varepsilon}\right)^{d-k}}{m^{e-k}}\right) \\
&\approx T_1(\alpha)
\end{aligned}$$

by our choice of parameters. Note that e needs to tailored very precisely to force the above condition! If e is chosen too large or small, we get nothing from this whole exercise!

Thus by Lemma 6.35 and Lemma 6.36, we get

$$\sum_{\substack{\alpha\in\Delta \\ \beta\in\partial_\alpha(P)}} |A(\alpha,\beta)| \quad \ge \quad |\Delta| \cdot \left|\bigcup_\beta S(\alpha,\beta)\right| \quad \ge \quad |\Delta| \cdot T_1(\alpha) \quad \approx \quad |\Delta| \cdot \binom{n}{\ell} \qquad (6.39)$$

Upper bounding $\sum |A(\alpha,\beta) \cap A(\alpha',\beta')|$

We are still left with the task of upper bounding

$$T_3 \quad = \quad \sum_{(\alpha,\beta)\neq(\alpha',\beta')} \left|A(\alpha,\beta) \cap A(\alpha',\beta')\right|$$

As mentioned earlier, we really do not have a good handle on the set $A(\alpha,\beta)$, and certainly not on the intersection of two such sets. Once again, we shall use a proxy that is easier to estimate to upper bound T_3.

The set $A(\alpha,\beta) \cap A(\alpha',\beta')$ consists of multilinear monomials γ of degree $\ell + d - k$ such that there exists multilinear monomials γ', γ'' of degree ℓ satisfying

$$\begin{aligned} \gamma &= \gamma'\beta \quad = \quad \gamma''\beta', \\ \gamma'\beta &= \mathrm{LM}(\mathrm{mult}(\gamma'\partial_\alpha(P))) \\ \text{and} \quad \gamma''\beta' &= \mathrm{LM}(\mathrm{mult}(\gamma''\partial_{\alpha'}(P))) \end{aligned}$$

This in particular implies that γ must be divisible by both β and β'.

Observation 6.40. *If* $\mathsf{Deg}(\gcd(\beta,\beta')) = t$*, then*

$$\left|A(\alpha,\beta) \cap A(\alpha',\beta')\right| \quad \leq \quad \binom{n - 2d + 2k + t}{\ell - d + k + t}$$

Proof. Every monomial $\gamma \in A(\alpha,\beta) \cap A(\alpha',\beta')$ must be divisible by β and β'. Since $|\beta \cup \beta'| = 2d - 2k - t$, the number of choices of γ is precisely

$$\binom{n - (2d - 2k - t)}{(\ell + d - k) - (2d - 2k - t)} \quad = \quad \binom{n - 2d + 2k + t}{\ell - d + k + t} \qquad \square$$

One needs a similar argument as in the case of T_2 to figure out how many pairs $(\alpha,\beta) \neq (\alpha',\beta')$ are there with $\mathsf{Deg}(\gcd(\beta,\beta')) = t$ and sum them up accordingly.

Lemma 6.41 ([KS17])**.** *For the polynomial* $\mathsf{NW}_{d,m,e}$*, and* $n = md$ *and* $\ell = \frac{n}{2}(1-\varepsilon)$ *for* $\varepsilon = \Theta\left(\frac{\log d}{\sqrt{d}}\right)$*,*

$$\sum_{(\alpha,\beta)\neq(\alpha',\beta')} \left|A(\alpha,\beta) \cap A(\alpha',\beta')\right| \quad \lesssim \quad |\Delta|^2 \cdot \left(\frac{m^{e-k}}{2^{d-k}}\right)^2 \cdot \binom{n}{\ell}.$$

Proof. Fix a pair of derivatives α, α'. Let

$$T_3(\alpha, \alpha') \quad := \quad \sum_{\substack{\beta \in \partial_\alpha(P) \\ \beta' \in \partial_{\alpha'}(P) \\ (\alpha,\beta) \neq (\alpha',\beta')}} \left| A(\alpha, \beta) \cap A(\alpha', \beta') \right|$$

As before, we shall first count the number of pairs of monomials $\beta \in \partial_\alpha P$ and $\beta' \in \partial_{\alpha'} P$ such that $\gcd(\beta, \beta') = t$. Note that since α may differ from α', we could potentially have $\gcd(\beta_1, \beta_2) = e$. Once again, this is easily seen to be at most

$$m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-k-t}.$$

Therefore, using Observation 6.40,

$$\begin{aligned}
T_3(\alpha, \alpha') &\leq \sum_{t=0}^{e} m^{e-k} \cdot (m-1)^{e-k-t} \binom{d-k}{t} \binom{n-2d+2k+t}{\ell-d+k+t} \\
&\approx \sum_{t=0}^{e} m^{e-k} \cdot (m-1)^{e-k-t} \binom{d-k}{t} \cdot \binom{n}{\ell} \left(\frac{1}{2}\right)^{2d-2k-t} (1+\varepsilon)^t \\
&\leq \frac{m^{2(e-k)}}{2^{2(d-k)}} \cdot \binom{n}{\ell} \cdot \left(1 + \frac{2(1+\varepsilon)}{m}\right)^{d-k} \\
&\approx \left(\frac{m^{e-k}}{2^{d-k}}\right)^2 \cdot \binom{n}{\ell} \\
\implies T_3 &\lesssim |\Delta|^2 \cdot \left(\frac{m^{e-k}}{2^{d-k}}\right)^2 \cdot \binom{n}{\ell}
\end{aligned}$$

□

Recalling that we have chosen our parameters so that

$$\frac{m^{e-k}}{2^{d-k}} \quad \approx \quad \left(\frac{1}{1+\varepsilon}\right)^{d-k} \quad \text{and} \quad |\Delta| = (1+\varepsilon)^{2(d-k)},$$

the above equation reduces to

$$T_3 \quad = \quad \sum_{(\alpha,\beta) \neq (\alpha',\beta')} \left| A(\alpha, \beta) \cap A(\alpha', \beta') \right| \quad \lesssim \quad |\Delta| \cdot \binom{n}{\ell}.$$

Combining with (6.39), we obtain the required bound for $|\bigcup A(\alpha, \beta)|$ via Inclusion-Exclusion (Corollary 6.34).

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(\mathsf{NW}_{d,m,e}) \quad \geq \quad \left| \bigcup_{\alpha,\beta} A(\alpha, \beta) \right| \quad \gtrsim \quad \binom{n}{\ell} \cdot (1+\varepsilon)^{2d-2k}$$

The only thing left to observe is that by Lemma 6.37,

$$\binom{n}{\ell+d-k} \quad \approx \quad \binom{n}{\ell} \cdot (1+\varepsilon)^{2d-2k}$$

and that completes the proof of Lemma 6.18. □

Chapter 7

Conclusion and open problems

The main questions which motivated the research in this thesis were to prove strong superpolynomial lower bounds for homogeneous arithmetic circuits of depth-4 and depth-5. We also hoped to understand if the upper bounds obtained for depth reduction for such low depth circuits could be further improved. To an extent, we make concrete progress on both these directions. However, a number of exciting questions continue to remain open. The most natural problem of interest here is to improve the state of art of lower bounds for general arithmetic circuits, arithmetic formula and algebraic branching programs. In addition to these, there are a number of concrete and interesting questions closely related to the results in this thesis which remain open. We conclude with a list of such problems and research directions.

Lower bounds for homogeneous depth-4 circuits for elementary symmetric polynomials

The elementary symmetric polynomial of degree d in n variables, $S_{d,n}$ is defined as

$$S_{d,n} = \sum_{T \subseteq [n], |T|=d} \prod_{i \in T} x_i$$

These are a natural family of multilinear polynomials, and have been studied in a number of prior works in the context of arithmetic circuit lower bounds[NW97, SW01, Shp02]. It is known that $S_{d,n}$ can be computed by a homogeneous depth-4 circuit of size $2^{O(\sqrt{d}) \cdot \mathrm{poly}(n)}$ over fields of characteristic zero. However, we do not know any superpolynomial lower bounds for homogeneous depth-4 circuits computing $S_{d,n}$. In fact, the problem is interesting even for homogeneous depth-4 circuits with bounded bottom fan-in.

Depth-5 circuits vs depth-4 circuits

The results in Chapter 5 lead to the following open questions.

- One question of great interest would be to show the lower bounds in Chapter 5 when the degree of the polynomials is larger. The other proofs of lower bounds for homogeneous depth-4 circuits [KLSS14a, KS17] tolerate degrees as high as $n^{1/2}$. We conjecture that the results in this chapter are true even when the degree d and the number of variables n are polynomially related.

- Is the dimension of projected shifted partials of a generic homogeneous depth-5 circuit close to the largest possible value? This could offer one approach to resolving the first open problem.

- If the answer to the second problem above is negative, then we might be able to use projected shifted partials as a complexity measure to prove new lower bounds for homogeneous depth-5 arithmetic circuits. Hence, even proving non-trivial *upper bounds* on the projected shifted partials complexity of homogeneous depth-5 circuits would be very interesting.

Lower bounds for depth-5 circuits for iterated matrix multiplication

Given the results in Chapter 6, one might wonder if the lower bounds in there work for a polynomial in VP. One natural candidate polynomial for which one might hope to show such a lower bound would be the iterated matrix multiplication polynomial (IMM). It was shown in [KS17] that IMM has a large complexity with respect to the measure of projected shifted partial derivatives. Unfortunately, the bounds in [KS17] only show that the dimension of the space of projected shifted partial derivatives of the IMM (degree d in $d^{O(1)}$ variables) are a factor $\exp(\delta\sqrt{d}\log d)$ close to the maximum possible value for some constant δ. This slack seems to be insufficient for our proofs in this chapter to work as in the proof of Lemma 6.24, we would have to rely on the fact that for the polynomial NW, the projected shifted partials complexity was at most a quasi-polynomial factor away from the largest possible.

Lower bounds for non-homogeneous depth-3 circuits

In contrast to the homogeneous case, non homogeneous computation is much less understood even at depth-3. Over fields of growing size, the best known lower bounds for depth-3 circuits is just cubic [KST16b]. It would be extremely interesting to prove superpolynomial lower bounds for this problem.

One approach to this question would be by proving better lower bounds for homogeneous depth-5 circuits over fields of characteristic zero. It is known that strong enough superpolynomial lower bounds for homogeneous depth-5 circuits over fields of characteristic zero would imply would imply superpolynomial lower bounds for non-homogeneous depth-3 circuits over such fields.

Lower bounds for arithmetic circuits of larger depth

Even for depth-5 circuits, we only know superpolynomial lower bounds over fields of constant size. It is therefore natural to try and prove such lower bounds over all fields. For circuits of depth larger than 5, our state of understanding is even worse, as we only know slightly superlinear lower bounds. Improving these bounds, even for constant depth circuits would be an extremely interesting line of research.

Limitations of partial derivative based techniques

Many of the known lower bounds for arithmetic circuits (including all the results in this thesis) rely on using a variant of the method of partial derivatives to measure the *complexity* of a polynomial. Abstractly, these proofs associate to every polynomial a matrix whose entries are linear functions in the coefficient vector of the polynomial. The intuition is that if the polynomial has a *simple circuit*, the rank of this matrix is not too large, whereas there exist explicit polynomials for which this matrix is of high rank. A natural question here is to understand if such techniques are strong enough for proving stronger lower bounds. A better understanding of this question, even under certain believable pseudorandomness assumptions, would provide some much-needed insight in the search for appropriate complexity measures.

References

[AAD00] Manindra Agrawal, Eric Allender, and Samir Datta. On TC^0, AC^0, and Arithmetic Circuits. *Journal of Computer and System Sciences*, 60(2):395–421, 2000.

[AKS04] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. PRIMES is in P. *Annals of Mathematics*, 160(2):781–793, 2004.

[AV08] Manindra Agrawal and V. Vinay. Arithmetic Circuits: A Chasm at Depth Four. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008)*, pages 67–75, 2008. Pre-print available at `eccc:TR08-062`.

[BC15] Suman K. Bera and Amit Chakrabarti. A Depth-Five Lower Bound for Iterated Matrix Multiplication. In *Conference on Computational Complexity (CCC)*, pages 183–197, 2015.

[Bjö14] Andreas Björklund. Determinant Sums for Undirected Hamiltonicity. *SIAM Journal of Computing*, 43(1):280–299, 2014.

[BS83] Walter Baur and Volker Strassen. The Complexity of Partial Derivatives. *Theoretical Computer Science*, 22:317–330, 1983.

[CM14a] Suryajith Chillara and Partha Mukhopadhyay. Depth-4 Lower Bounds, Determinantal Complexity : A Unified Approach. *Proceedings of the 31st Symposium on Theoretical Aspects of Computer Science (STACS 2014)*, 2014. Pre-print available at `arXiv:1308.1640`.

[CM14b] Suryajith Chillara and Partha Mukhopadhyay. On the Limits of Depth Reduction at Depth 3 Over Small Finite Fields. In *Mathematical Foundations of Computer Science (MFCS)*, pages 177–188, 2014. Pre-print available at `arXiv:1401.0189`.

[FKS16] Michael A. Forbes, Mrinal Kumar, and Ramprasad Saptharishi. Functional Lower Bounds for Arithmetic Circuits and Connections to Boolean Circuit Complexity. In *Proceedings of the 31st Annual Computational Complexity Conference (CCC 2016)*, 2016.

[FLMS14] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC 2014)*, pages 128–135, 2014. Pre-print available at `eccc:TR13-100`.

[GK98] Dima Grigoriev and Marek Karpinski. An Exponential Lower Bound for Depth 3 Arithmetic Circuits. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC 1998)*, pages 577–582, 1998.

[GKKS13] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic Circuits: A Chasm at Depth Three. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2013)*, pages 578–587, 2013. Pre-print available at `eccc:TR13-026`.

[GKKS14] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the Chasm at Depth Four. *Journal of the ACM*, 61(6):33:1–33:16, 2014. Preliminary version in the *28th Annual IEEE Conference on Computational Complexity (CCC 2013)*. Pre-print available at `eccc:TR12-098`.

[GKL12] Ankit Gupta, Neeraj Kayal, and Satyanarayana V. Lokam. Reconstruction of depth-4 multilinear circuits with top fan-in 2. In *STOC2012*, pages 625–642, 2012.

[GR00] Dima Grigoriev and Alexander A. Razborov. Exponential Lower Bounds for Depth 3 Arithmetic Circuits in Algebras of Functions over Finite Fields. *Appl. Algebra Eng. Commun. Comput.*, 10(6):465–487, 2000. Preliminary version in the *39th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1998)*.

[Hås86] Johan Håstad. Almost Optimal Lower Bounds for Small Depth Circuits. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC 1986)*, pages 6–20, 1986.

[HY11] Pavel Hrubes and Amir Yehudayoff. Homogeneous Formulas and Symmetric Polynomials. *Computational Complexity*, 20(3):559–578, 2011.

[JS82] Mark Jerrum and Marc Snir. Some Exact Complexity Results for Straight-Line Computations over Semirings. *Journal of the ACM*, 29(3):874–897, 1982.

[Kal85] Kyriakos Kalorkoti. A Lower Bound for the Formula Size of Rational Functions. *SIAM Journal of Computing*, 14(3):678–687, 1985.

[Kay12] Neeraj Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. In *Electronic Colloquium on Computational Complexity (ECCC)TR12-081*, 2012.

[KLSS14a] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An Exponential Lower Bound for Homogeneous Depth Four Arithmetic Circuits. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2014)*, 2014. Pre-print available at `eccc:TR14-005`.

[KLSS14b] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. Super-polynomial lower bounds for depth-4 homogeneous arithmetic formulas. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC 2014)*, pages 119–127, 2014.

[KMSV10] Zohar Shay Karnin, Partha Mukhopadhyay, Amir Shpilka, and Ilya Volkovich. Deterministic identity testing of depth-4 multilinear circuits with bounded top fan-in. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC 2010)*, pages 649–658, 2010.

[Koi12] Pascal Koiran. Arithmetic Circuits: The Chasm at Depth Four Gets Wider. *Theoretical Computer Science*, 448:56–65, 2012. Pre-print available at `arXiv:1006.4700`.

[KS13] 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.

[KS14] Mrinal Kumar and Shubhangi Saraf. Superpolynomial Lower Bounds for General Homogeneous Depth 4 Arithmetic Circuits. In *Proceedings of the 41st International Colloquium on Automata, Languages and Programming (ICALP 2014)*, 2014.

[KS15a] Neeraj Kayal and Chandan Saha. Lower Bounds for Depth Three Arithmetic Circuits with Small Bottom Fanin. In *Proceedings of the 30th Annual Computational Complexity Conference (CCC 2015)*, pages 158–208, 2015. Pre-print available at `eccc:TR14-089`.

[KS15b] Mrinal Kumar and Ramprasad Saptharishi. An exponential lower bound for homogeneous depth-5 circuits over finite fields. *CoRR*, abs/1507.00177, 2015. `arXiv:1507.00177`.

[KS15c] Mrinal Kumar and Ramprasad Saptharishi. An exponential lower bound for homogeneous depth-5 circuits over finite fields. *Electronic Colloquium on Computational Complexity (ECCC)*, 2015. `eccc:TR15-109`.

[KS15d] Mrinal Kumar and Shubhangi Saraf. The limits of depth reduction for arithmetic formulas: it's all about the top fan-in. *SIAM J. Comput.*, 44(6):1601–1625, 2015. Preliminary version in the *46th Annual ACM Symposium on Theory of Computing (STOC 2014)*. Pre-print available at `eccc:TR13-068`.

[KS16] Mrinal Kumar and Ramprasad Saptharishi. Finer separations for shallow arithmetic circuits. In *Proceedings of the 36th International Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2016)*, 2016.

[KS17] Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. *SIAM J. Comput.*, 46(1):336–387, 2017. Preliminary version in the *55th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2014)*. Pre-print available at `eccc:TR14-045`.

[KSS14] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A superpolynomial lower bound for regular arithmetic formulas. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC 2014)*, pages 146–153, 2014. Pre-print available at `eccc:TR13-091`.

[KST16a] Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. An almost Cubic Lower Bound for Depth Three Arithmetic Circuits. In *Proceedings of the 43rd International Colloquium on Automata, Languages and Programming (ICALP 2016)*, 2016.

[KST16b] Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. An almost Cubic Lower Bound for Depth Three Arithmetic Circuits. Technical report, Electronic Colloquium on Computational Complexity (ECCC), `eccc:TR16-006`, 2016.

[MVV87] Ketan Mulmuley, Umesh V. Vazirani, and Vijay V. Vazirani. Matching Is as Easy as Matrix Inversion. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC 1987)*, pages 345–354, 1987.

[NW97] Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997. Available on `citeseer:10.1.1.90.2644`.

[Raz87] Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

[Raz06] Ran Raz. Separation of Multilinear Circuit and Formula Size. *Theory of Computing*, 2(1):121–135, 2006. Preliminary version in the *45th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2004)*. Preprint available at `eccc:TR04-042`.

[Raz09] Ran Raz. Multi-Linear Formulas for Permanent and Determinant are of Super-Polynomial Size. *Journal of the ACM*, 56(2), 2009. Preliminary version in the *36th Annual ACM Symposium on Theory of Computing (STOC 2004)*. Pre-print available at `eccc:TR03-067`.

[Raz10a] Ran Raz. Elusive Functions and Lower Bounds for Arithmetic Circuits. *Theory of Computing*, 6(1):135–177, 2010. Preliminary version in the *40th Annual ACM Symposium on Theory of Computing (STOC 2008)*. Pre-print available at `eccc:TR08-001`.

[Raz10b] Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC 2010)*, pages 659–666, 2010. Pre-print available at `eccc:TR10-002`.

[RS60] I. S. Reed and G. Solomon. Polynomial Codes Over Certain Finite Fields. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):300–304, 1960.

[RST15] Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. An Average-Case Depth Hierarchy Theorem for Boolean Circuits. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2015)*, 2015. Preliminary version at `arXiv:1504.03398`.

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

2009. Preliminary version in the *23rd Annual IEEE Conference on Computational Complexity (CCC 2008)*. Pre-print available at `eccc:TR08-006`.

[Sap15] Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity. Github survey, 2015.

[Shp02] Amir Shpilka. Affine Projections of Symmetric Polynomials. *Journal of Computer and System Sciences*, 65(4):639–659, December 2002.

[Smo87] Roman Smolensky. Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC 1987)*, pages 77–82, 1987.

[Str73] V. Strassen. Die berechnungskomplexiät von elementarsymmetrischen funktionen und von interpolationskoeffzienten. *Numerische Mathematik*, 20:238–251, 1973.

[SV11] Shubhangi Saraf and Ilya Volkovich. Black-box identity testing of depth-4 multilinear circuits. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC 2011)*, pages 421–430, 2011.

[SW01] Amir Shpilka and Avi Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Computational Complexity*, 10(1):1–27, 2001. Preliminary version in the *14th Annual IEEE Conference on Computational Complexity (CCC 1999)*.

[SY10] Amir Shpilka and Amir Yehudayoff. Arithmetic Circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5:207–388, March 2010.

[Tav15] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. *Inf. Comput.*, 240:2–11, 2015. Preliminary version in the *38th Internationl Symposium on the Mathematical Foundations of Computer Science (MFCS 2013)*.

[Val79] Leslie G. Valiant. Completeness Classes in Algebra. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing (STOC 1979)*, pages 249–261, 1979.

[VSBR83] Leslie G. Valiant, Sven Skyum, S. Berkowitz, and Charles Rackoff. Fast Parallel Computation of Polynomials Using Few Processors. *SIAM Journal of Computing*, 12(4):641–644, 1983. Preliminary version in the *6th Internationl Symposium on the Mathematical Foundations of Computer Science (MFCS 1981)*.