Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 45 (2014)

On the power of homogeneous depth 4 arithmetic circuits[*]

Mrinal Kumar[†] Shubhangi Saraf[‡]

Abstract

We prove exponential lower bounds on the size of homogeneous depth 4 arithmetic circuits computing an explicit polynomial in VP. Our results hold for the *Iterated Matrix Multiplication* polynomial - in particular we show that any homogeneous depth 4 circuit computing the $(1, 1)$ entry in the product of n generic matrices of dimension $n^{O(1)}$ must have size $n^{\Omega(\sqrt{n})}$.

Our results strengthen previous works in two significant ways.

1. Our lower bounds hold for a polynomial in VP. Prior to our work, Kayal et al [KLSS14] proved an exponential lower bound for homogeneous depth 4 circuits (over fields of characteristic zero) computing a poly in VNP. The best known lower bounds for a depth 4 homogeneous circuit computing a poly in VP was the bound of $n^{\Omega(\log n)}$ by [KLSS, KLSS14].

   Our exponential lower bounds also give the first exponential separation between general arithmetic circuits and homogeneous depth 4 arithmetic circuits. In particular they imply that the depth reduction results of Koiran [Koi12] and Tavenas [Tav13] are tight even for reductions to general homogeneous depth 4 circuits (without the restriction of bounded bottom fanin).

2. Our lower bound holds over all fields. The lower bound of [KLSS14] worked only over fields of characteristic zero. Prior to our work, the best lower bound for homogeneous depth 4 circuits over fields of positive characteristic was $n^{\Omega(\log n)}$ [KLSS, KLSS14].

[*] A preliminary version of this paper appeared in the proceedings of FOCS 2014

[†] Department of Computer Science, Rutgers University. Email: mrinal.kumar@rutgers.edu.

[‡] Department of Computer Science and Department of Mathematics, Rutgers University. Email: shubhangi.saraf@gmail.com. Research supported by NSF grant CCF-1350572.

ISSN 1433-8092

Contents

1 Introduction

In a seminal work [Val79], Valiant defined the classes VP and VNP as the algebraic analogs of the classes P and NP. The problem of separating VNP from VP has since been one of the most important open problems in algebraic complexity theory. Although the problem has received a great deal of attention in the following years, the best lower bounds known for general arithmetic circuits are barely super linear [Str73, BS83]. The absence of progress on the general problem has led to much attention being devoted to proving lower bounds for *restricted classes* of arithmetic circuits. Arithmetic circuits of small depth are one such class that has been intensively studied.

Depth Reduction: In a very interesting direction of research, Valiant et al [VSBR83] showed that every polynomial of degree n in $\mathsf{poly}(n)$ variables, which can be computed by a $\mathsf{poly}(n)$ sized arithmetic circuit, can also be computed by a $\mathsf{poly}(n)$ sized arithmetic circuit of *depth* $O(\log^2 n)$. In other words, arbitrary depth circuits in VP can be reduced to circuits of depth $O(\log^2 n)$ with only a polynomial blowup in size. Thus, in order to separate VNP from VP, it would suffice to show a super-polynomial lower bound for just circuits of depth $O(\log^2 n)$. In an intriguing line of recent works in this direction, Agrawal-Vinay [AV08], Koiran [Koi12] and Tavenas [Tav13] built upon the results of Valiant et al [VSBR83] and showed that much stronger depth reductions are possible. In order to separate VNP form VP, it would suffice to prove strong enough ($n^{\omega(\sqrt{n})}$) lower bounds for just *homogeneous depth 4 circuits*.

Lower bounds for homogeneous bounded depth circuits: In an extremely influential work, Nisan and Wigderson [NW95] proved the first super-polynomial (and in fact exponential) lower bound for the class of homogeneous depth 3 circuits. This work used the *dimension of the space of partial derivatives* as a measure of complexity of a polyomial, and used this measure to prove the lower bounds. For several years thereafter, there were no improved lower bounds - even for the case of depth 4 homogeneous circuits, the best lower bounds were just mildly super-linear [Raz10]. This is contrary to what is known for Boolean circuits, where we know exponential lower bounds for bounded depth circuits. This seemed surprising until the depth reduction results of Agrawal-Vinay [AV08] and later Koiran [Koi12] and Tavenas [Tav13], which demontrated that in some sense, homogeneous depth 4 circuits *capture* the inherent complexity of general arithmetic circuits.

In a breakthrough result in 2012, Gupta, Kamath, Kayal and Saptharishi [GKKS13a], made the first major progress on the problem of obtaining lower bounds for bounded depth circuits, by proving $2^{\Omega(\sqrt{n})}$ lower bounds for an explicit polynomial of degree n in $n^{O(1)}$ variables computed by a homogeneous depth 4 circuit, where the fan-in of the product gates at the bottom level of the depth 4 circuits is bounded by $\sqrt{n}$. For ease of exposition, let us denote the class of depth 4 circuits with bottom fanin $\sqrt{n}$ by $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits. The lower bounds of [GKKS13a] were later improved to $2^{\Omega(\sqrt{n}\log n)}$ in a follow up work of Kayal, Saha, Saptharishi [KSS13]. These results were all the more remarkable in the light of the results of Koiran [Koi12] and Tavenas [Tav13] who had in fact showed that $2^{\omega(\sqrt{n}\log n)}$ lower bounds even for homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits would suffice to separate VP from VNP. Thus, any asymptotic improvement in the exponent, in either the upper bound on depth reduction or the lower bound of [KSS13] would separate VNP from VP. Both papers [GKKS13a, KSS13] used the notion of the dimension of *shifted partial derivatives* as a complexity measure, a refinement of the Nisan-Wigderson complexity measure of dimension of partial derivatives.

The most tantalizing questions left open by these works was to improve either the depth reduction or the lower bounds. In [FLMS13], the lower bounds of [KSS13] were strengthened by showing that they also held for a polynomial in VP. These were further extended in [KS], where the same exponential ($n^{\Omega(\sqrt{n})}$) lower bounds were also shown to hold for very simple polynomial sized formulas of just depth 4 (if one requires them to be computed by homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$

circuits). On one hand, these results give us extremely strong lower bounds for an interesting class of depth 4 homogeneous circuits. On the other hand, since these lower bounds also hold for polynomials in VP and for homogeneous formulas [FLMS13, KS], it follows that the depth reduction results of Koiran [Koi12] and Tavenas [Tav13] to the class of homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits are tight and cannot be improved even for homoegeneous formulas.

Although these results represent a lot of exciting progress on the problem of proving lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits, and these results seemed possibly to be on the brink of proving lower bounds for general arithmetic circuits, they still seemed to give almost no nontrivial results for general homogeneous depth 4 circuits with no bound on bottom fanin (homogeneous $\Sigma\Pi\Sigma\Pi$ circuits). Moreover, it was shown in [KS] that general homogeneous $\Sigma\Pi\Sigma\Pi$ circuits are exponentially more powerful than homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits[1]. Till very recently, the only lower bounds we knew for general homogeneous depth 4 circuits were the slightly super-linear lower bounds by Raz using the notion of elusive functions [Raz10] (these worked even for non-homogeneous circuits).

Lower bounds for general homogeneous depth 4 circuits: Recently, the first super-polynomial lower bounds for general homogeneous depth 4 ($\Sigma\Pi\Sigma\Pi$) circuits were proved independently by the authors of this paper [KS13] who showed a lower bound of $n^{\Omega(\log\log n)}$ for a polynomial in VNP and Kayal, Limaye, Saha and Srinivasan [KLSS], who showed a lower bound of $n^{\Omega(\log n)}$ for a 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 [GKKS13a] and [KSS13] who proved similar exponential lower bounds for the more restricted class of homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits. The result by [KLSS14] 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 paper build upon and strengthen the results of Kayal et al [KLSS14]. Before we describe our results we first highlight some important questions left open by [KLSS14] and place them in the context of several of the other recent results in this area.

- **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 [GKKS13b], 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 [GR98] [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.

[1] It was demonstrated that even very simple homogeneous $\Sigma\Pi\Sigma\Pi$ circuits of polynomial size might need $n^{\Omega(\sqrt{n})}$ sized homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits to compute the same polynomial.

[2] albeit with loss of homogeneity.

[3] Recently, Chillara and Mukhopadhyay [CM14] showed $2^{\Omega(n\log n)}$ lower bounds for depth 3 circuits over small finite fields for a polynomial in VP.

The lower bounds of [KLSS14] 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 [KLSS14] 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 [KLSS, KLSS14]. Recall that when one introduces the restriction on bounded bottom fanin, then stronger exponential lower bounds are indeed known [FLMS13, KS]. This fact is also related to the next bullet point below.
- **Tightness of depth reduction:** The result of [FLMS13] (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 [Tav13] (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 [KSS13] 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 [KS], 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 [KLSS14], 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 [KLSS14] result, this would also suffice in separating VP from VNP!
- **Shifted partial derivatives and variants:** The results of [KS13, KLSS, KLSS14] 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 any any limitations to these methods.

1.1 Our results

In this paper, 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 [KLSS14]. Our results extend those of [KLSS14] 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 1.1. *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 1.1 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 [KLSS14] show their lower bounds. We give a formal definition in Section 3. The main difference in our proof of the above result from the proof in [KLSS14] is that our proof of the lower bound on the complexity of the polynomial is completely combinatorial, while the proof in [KLSS14], 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 in Section 2.

Theorem 1.2 (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 $n^{\Omega(\sqrt{n})}$.*

As an immediate corollary of the result above, we conclude that the depth reduction results of Koiran [Koi12] and Tavenas [Tav13] are tight even when one wants to depth reduce to the class of general homogeneous depth 4 circuits.

Corollary 1.3 (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 [Tav13] is tight, even when the bottom fan-in is unbounded.*

The polynomial in Theorem 1.2 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 1.4. *There exists a constant $\epsilon > 0$ such that any homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing the polynomial Det_n has size at least $2^{\Omega(n^{\epsilon})}$.*

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

1.2 Organisation of the paper

In Section 2, we provide a broad overview of the proofs of Theorem 1.1 and Theorem 1.2. In Section 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. We lay down our strategy for obtaining a lower bound on the complexity of the polynomials of interest in Section 5. Finally in Sections 6 and 7, we prove Theorem 1.1 and in Sections 8 and 9, we prove Theorem 1.2. We conclude with some open problems in Section 10.

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* ($\geq n^{\epsilon\sqrt{n}}$) to start with, we have nothing to prove. Else, the size of C is small ($< n^{\epsilon\sqrt{n}}$).
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.
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 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 3) in terms of its size. This step is very similar to that in [KLSS14], 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 [KLSS14], 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 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 Section 5.1 and give formal proofs in Sections 8 and 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 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 [KLSS14]. 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 paper these calculations use much more sophisticated arguments.

Comparison to [KLSS14] Although the complexity measure and the random restrictions in this paper are the same as the one used in [KLSS14], 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 paper, 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).

3 Preliminaries

Arithmetic Circuits: An arithmetic circuit over a field $\mathbb{F}$ and a set of variables $x_1, x_2, \ldots, x_N$ is a directed acyclic graph with internal nodes labelled by the field operations and the leaf nodes labelled by input variables or field elements. By the *size* of the circuit, we mean the total number of nodes in the underlying graph and by the *depth* of the circuit, we mean the length of the longest path from the output node to a leaf node. A circuit is said to be *homogeneous* if the polynomial computed at every node is a homogeneous polynomial. By a $\Sigma\Pi\Sigma\Pi$ circuit or a depth 4 circuit, we mean a circuit of depth 4 with the top layer and the third layer only have sum gates and the second and the bottom layer have only product gates. A homogeneous polynomial P of degree n in N variables, which is computed by a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit can be written as

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

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

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

Support of a polynomial: By the support of a polynomial P, denoted by $\text{Supp}(P)$, 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 1, is said to be a $\Sigma\Pi\Sigma\Pi^{\{s\}}$ circuit if every product gate at the bottom level has support at most s (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 homoegeneous $\Sigma\Pi\Sigma\Pi^{\{s\}}$ 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 introduced in [Kay12] and was subsequently used as a complexity measure in proving several recent lower bound results [FLMS13, GKKS13a, KSS13, KS13, KS]. In this paper, we use a variant of the method which first introduced in [KLSS14].

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 3.1 (($\mathcal{M}, m$)-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*

$$\langle \partial_{\mathcal{M}}(P) \rangle_m \overset{def}{=} \mathbb{F}\text{-}span\{\sigma(\prod_{i \in S} x_i \cdot g) : g \in \partial_{\mathcal{M}}(P), S \subseteq [N], |S| = m\} \tag{2}$$

In this paper, 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 paper, 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 3.2 (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 will now define the family of polynomials $NW_{n,D}$ in VNP which were used for the first time in the context of lower bounds in [KSS13]. The key motivation for this definition is that over any finite field, any two distinct low degree polynomials do not agree at too many points, and hence we use this property to construct a polynomial with monomials that have large distance. Let $\mathbb{F}_n$ be a finite field of size n^4 and let F_{n^2} be its quadratic extension. 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 paper, 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 3.3. *A monomial θ is said to be an extension of a monomial $\tilde{\theta}$, if θ divides $\tilde{\theta}$.*

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 use the notation $\mathsf{Lead\text{-}Mon}(P)$ to indicate the leading monomial of P under this monomial ordering.

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

[4] We are assuming for simplicity that n is a prime power, but the definitions can be easily adapted for when n is not.

Definition 3.4 (Monomial distance)**.** *Let m_1 and m_2 be two monomials over a set of variables. Let S_1 and S_2 be the multiset of variables in m_1 and m_2 respectively, then the distance $\Delta(m_1, m_2)$ between m_1 and m_2 is the min$\{|S_1| - |S_1 \cap S_2|, |S_2| - |S_1 \cap S_2|\}$ where the cardinalities are the order of the multisets.*

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

Observation 3.5. *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 $|\mathrm{Supp}(\alpha) \cap \mathrm{Supp}(\beta)| = t$. When $t = 0$, we say that α and β are disjoint.

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

Lemma 3.6 ([GKKS13a])**.** *Let $a(n), f(n), g(n) : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be integer valued functions such that $(f + g) = o(a)$. Then,*

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

In this paper, we invoke Lemma 3.6 only in situations where $(f + g)^2$ will be $O(a)$. In this case, the error term will be bounded by an absolute constant. Hence, up to multiplication by constants, $\frac{(a+f)!}{(a-g)!} = a^{(f+g)}$. We will use the symbol $\approx$ to indicate equality up to multiplication by constants.

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

Lemma 3.7. *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,

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

The proof is given in Appendix A.

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

Lemma 3.8 (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, $\left|\bigcup_{i \in [l]} W_i\right| \geq \frac{1}{4\lambda} \sum_{i \in [l]} |W_i|$.

The proof appears in Appendix B.

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 [KLSS14]. We include a proof for completeness.

Lemma 4.1. *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,*

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

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 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 $\{\sigma(\beta\cdot\prod_{j\in[l]\setminus S} P_j) : m \leq \mathrm{Supp}(\beta) \leq m+rs\}$. Note that the set of polynomials $\{\sigma(\beta\cdot\prod_{j\in[l]\setminus S} P_j) : m \leq \mathrm{Supp}(\beta) \leq m+rs\}$ 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 $\{\sigma(\beta\cdot\prod_{j\in[l]\setminus S} P_j) : m \leq \mathrm{Supp}(\beta) \leq m+rs\}$. 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 $\{\sigma(\beta\cdot\prod_{j\in[l]\setminus S} P_j) : m \leq \mathrm{degree}(\beta) = \mathrm{Supp}(\beta) \leq m+rs\}$. The dimension of the space $\{\sigma(\beta\cdot\prod_{j\in[l]\setminus S} P_j) : m \leq \mathrm{degree}(\beta) = \mathrm{Supp}(\beta) \leq m+rs\}$ 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 $rs\cdot\binom{N}{m+rs}$ 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 $rs\cdot\binom{N}{m+rs}$. 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 $rs\cdot\binom{N}{m+rs}\cdot\binom{\lceil\frac{2u}{s}\rceil+r}{r}$ 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.

□

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

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

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 5.1. *Let P be a polynomial in N variables and let $\mathcal{M}$ be any set of monomials on these variables. Let $m \le N$ 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 $\beta \in \mathrm{Supp}(\partial_\alpha(P))$, $A_m(\alpha,\beta)$ are a subset of leading monomials of polynomials in $\mathbb{F}$-span $\{\sigma(\gamma\cdot\partial_{\mathcal{M}}(P)) : \mathrm{Supp}(\gamma) = \deg(\gamma) = m\}$. 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 5.2. *Let P be a polynomial in N variables and let $\mathcal{M}$ be any set of monomials on these variables. Let $m \le N$ 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 5.3. *Let P be a polynomial in N variables and let $\alpha\in\mathcal{M}$ be a monomial on these variables such that $\partial_\alpha(P)$ is not identically zero. 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 $Z=\{(\beta,\gamma):\beta\in\mathrm{Supp}(\partial_\alpha(P)),\gamma\in A_m(\alpha,\beta)\}$ and $W=\bigcup_{\beta\in\mathrm{Supp}(\partial_\alpha(P))}S_m(\alpha,\beta)$. 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 $\beta\in\mathrm{Supp}(\partial_\alpha(P))$, such that $\gamma\in S_m(\alpha,\beta)$. This implies that $\gamma\cdot\beta\in\mathrm{Supp}(\sigma(\gamma\cdot\partial_\alpha(P)))$. In particular, $\sigma(\gamma\cdot\partial_\alpha(P))$ is not the identically zero polynomial. So, there exists a $\beta'\in\mathrm{Supp}(\partial_\alpha(P))$ 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. □

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

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

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

and

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

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 5.2 and Lemma 5.3, 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 3.7 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 3.8, 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 3.8, 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 3.8 and Lemma 5.3 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 3.8 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 5.1, $\Phi_{\mathcal{G},m}(P)$ is large.

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.

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 $p = n^{-\epsilon}$, where ϵ is an absolute constant such that $0 \leq \epsilon \leq 0.01$. 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 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.

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 $0 < \epsilon < 0.01$.
7. $p = n^{-\epsilon}$. (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$n^{2d} = 1/4 \cdot n^{-2}\frac{\binom{N-k}{m}}{\binom{N-2k}{m-k}}$.
10. $D = \frac{\epsilon n}{2} + d$. (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 $p = n^{-\epsilon}$)

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

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 6.1. *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 $\epsilon > 0$ 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 $2^{\frac{\epsilon}{2}\sqrt{n}\log n}$ computing the polynomial NW. Then, with probability at least $1 - o(1)$ over $V \leftarrow \mathcal{D}$*

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

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

Now, by Lemma 4.1, the complexity of the circuit is at most $\text{Size}(C) \cdot \binom{\lceil \frac{2n}{s} \rceil + r}{r} \cdot \binom{N}{m+rs}$, with probability at least $1 - o(1)$. □

Observe that we have just argued that if the circuit was of size at most $2^{\frac{\epsilon}{2}\sqrt{n}\log n}$, 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 $2^{\frac{\epsilon}{2}\sqrt{n}\log n}$ to start with, in which case, we have nothing to prove.

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 3.7 and the bound on the variance of the number of monomials alive at the end of random restrictions obtained in [KLSS14].

Recall that $D = \frac{n.\epsilon}{2} + d$ 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 $\text{Supp}(NW)$. This implies that for every $\alpha \in \mathcal{M}^{[r]}$, $\partial_\alpha(NW)$ is non zero. In fact, it consists 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 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 $NW|_V$,

$$\Phi_{\mathcal{M},m}(NW|_V) \geq T_1|_V - T_2|_V - T_3|_V \tag{3}$$

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

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[\Phi_{\mathcal{M},m}(NW|_V)] \geq \mathbb{E}_{V \leftarrow \mathcal{D}}[T_1|_V] - \mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V] - \mathbb{E}_{V \leftarrow \mathcal{D}}[T_3|_V] \tag{4}$$

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 6.2.

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

Lemma 6.3.

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

Lemma 6.4.

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

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 7.

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

Lemma 6.5. *For any choice of parameters $m, r, d, \epsilon, n, N, k$ such that*

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

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 6.2, Lemma 6.3 and Lemma 6.4, 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 6.10). We make the following crucial observation.

Lemma 6.6. *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 $\bigcup_{\alpha\in\mathcal{M}^{[r]}} \mathrm{Supp}(\partial_\alpha(NW|_V))$ in the following claim.

Claim 6.7. *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 $\bigcup_{\alpha\in\mathcal{M}^{[r]}} \mathrm{Supp}(\partial_\alpha(NW|_V))$ to the set $\mathrm{Supp}(NW|_V)$. 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 $\mathrm{Supp}(NW|_V)$, there is exactly one monomial $\alpha \in \mathcal{M}^{[r]}$ such that β is an extension of α. Or, in other words, for each $\beta \in \mathrm{Supp}(NW|_V)$, there is exactly one $\alpha \in \mathcal{M}^{[r]}$ such that $\partial_\alpha(\beta) \in \mathrm{Supp}(\partial_\alpha(NW|_V))$. 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 $\beta \in \mathrm{Supp}(NW|_V)$ 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 $|\bigcup_{\alpha\in\mathcal{M}^{[r]}} \mathrm{Supp}(\partial_\alpha(NW|_V))|$ and $\binom{N-k}{m}$. The lemma follows from the claim above. □

We will now use Lemma 3.7 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 $g(P) = |\mathrm{Supp}(P)| \cdot \binom{N-k}{m}$ and define $f(P) = \Phi_{\mathcal{M}^{[r]},m}(P)$. Lemma 6.6 implies that for every restriction $P = NW|_V$ of NW, $f(P) \leq g(P)$. Lemma 6.5 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 [KLSS14] tells us that g takes values very close to its expected value with a high probability.

Lemma 6.8 ([KLSS14])**.** *$Pr_{V\leftarrow\mathcal{D}}[|g(NW|_V) - \mathbb{E}_{V'\leftarrow\mathcal{D}}[g]| \geq 0.1 \cdot \mathbb{E}_{V'\leftarrow\mathcal{D}}[g]] \leq 0.01$.*

The functions f and g now satisfy the hypothesis of Lemma 3.7. Therefore, we get the following lemma.

Lemma 6.9. *$Pr_{V\leftarrow\mathcal{D}}[f(NW|_V) \geq 0.01 \cdot \mathbb{E}_{V'\leftarrow\mathcal{D}}[g]] \geq 0.1$.*

Therefore, the following lemma is true.

Lemma 6.10. *For any choice of parameters $m, r, d, \epsilon, n, N, k$ such that*

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

the following is true

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

6.6 Wrapping up the proof

We now complete the proof of the lower bound for the case of NW polynomial which implies Theorem 1.1.

Theorem 6.11. *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 $\epsilon < 0.01$. 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 $n^{\frac{\epsilon}{2}\sqrt{n}}$, then we are done. Else, the size of C is at most $n^{\frac{\epsilon}{2}\sqrt{n}}$. Lemma 6.1 implies that with probability at least $1 - o(1)$ the complexity of the circuit is at most $\text{Size}(C)\binom{\lceil \frac{2n}{s} \rceil + r}{r}\binom{N}{m+rs}$.

We will first show that for the choice of paramters made above, the hypotheses of Lemma 6.5 hold.

Claim 6.12. *For $m, r, d, \epsilon, n, N, k$ as chosen above,*

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

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.

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

In other words, we would like to show that

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

Now,

$$
\begin{aligned}
& n^{2d-2r+\epsilon r} \cdot 4n \cdot \frac{\binom{N-2k}{m}}{\binom{N-k}{m}} \\
=& n^{-2r+\epsilon r} \cdot \frac{1}{n} \cdot \frac{\binom{N-2k}{m}}{\binom{N-2k}{m-k}} && \text{substituting value of } n^{2d} \\
=& n^{-2r+\epsilon r} \cdot \frac{1}{n} \cdot \frac{(N-m-k)!}{(N-m-2k)!} \times \frac{(m-k)!}{m!} \\
\approx& n^{-2r+\epsilon r} \cdot \frac{1}{n} \cdot \left(\frac{N-m}{m}\right)^{k} && \text{By Lemma 3.6} \\
=& n^{-2r+\epsilon r} \cdot \frac{1}{n} \cdot \left(\frac{1+\frac{\ln n}{5s}}{1-\frac{\ln n}{5s}}\right)^{k} && \text{substituting choice of } m \\
\leq& n^{-2r+\epsilon r} \cdot \frac{1}{n} \cdot e^{2.01k\frac{\ln n}{5s}} && \text{for large enough } n \\
=& n^{-2r+\epsilon r} \cdot \frac{1}{n} \cdot n^{2.01k/5s}
\end{aligned}
$$

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

Thus by the claim above and Lemma 6.10, we conclude that with

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

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

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

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

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

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

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

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

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

□

7 Calculations for $NW_{n,D}$

In this sections, we provide the proofs of Lemma 6.2, Lemma 6.3 and Lemma 6.4.

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

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

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 upto $w = D - r$ 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

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

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 7.1. *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 $\mathrm{Supp}(NW)$. 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 $\mathrm{Supp}(\gamma) \setminus \mathrm{Supp}(\beta)$ are alive. This happens with probability equal to $p^{|\mathrm{Supp}(\gamma)\setminus\mathrm{Supp}(\beta)|} = p^{k-w}$. 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

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

Claim 7.2. *Let α,β be monomials such that $\alpha\in\mathcal{M}^{[r]}$ and $\beta\in M(\alpha)$. Then*

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

Proof. Claim 7.1 implies that

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

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 3.6

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

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 $p = n^{-\epsilon}$ 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

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

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

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

□

Putting this value back into the equality

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

we obtain

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

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,

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

Plugging back the values of p and D, we get Lemma 6.3.

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.

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[T_3|_V] = \sum_{\substack{\alpha_1,\alpha_2\in\mathcal{M}^{[r]}\\ \beta_1\in M(\alpha_1)\\ \beta_2\in M(\alpha_2)\\ (\alpha_1,\beta_1)\neq(\alpha_2,\beta2)}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_1,\alpha_2,\beta_1,\beta_2}\cdot|A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|]$$

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

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[T_3|_V] = \sum_{\substack{0\le w_1\le r, 0\le w_2\le k\\ w_1+w_2\le D}} \sum_{\substack{\alpha_1,\alpha_2\in\mathcal{M}^{[r]}\\ \beta_1\in M(\alpha_1)\\ \beta_2\in M(\alpha_2)\\ |\mathrm{Supp}(\alpha_1)\cap\mathrm{Supp}(\alpha_2)|=w_1\\ |\mathrm{Supp}(\beta_1)\cap\mathrm{Supp}(\beta2)|=w_2}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_1,\alpha_2,\beta_1,\beta_2}\cdot|A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|]$$

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

$$\begin{aligned}\mathbb{E}_{V\leftarrow\mathcal{D}}[T_3|_V] = &\sum_{\substack{\alpha_1\in\mathcal{M}^{[r]}\\ \beta_1\in M(\alpha)}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_1,\beta_1}]\\ &\times\sum_{\substack{0\le w_1\le r, 0\le w_2\le k\\ w_1+w_2\le D}} \sum_{\substack{\alpha_2\in\mathcal{M}^{[r]}\\ \beta_2\in M(\alpha_2)\\ |\mathrm{Supp}(\alpha_1)\cap\mathrm{Supp}(\alpha_2)|=w_1\\ |\mathrm{Supp}(\beta_1)\cap\mathrm{Supp}(\beta2)|=w_2}} \mathbb{E}_{V\leftarrow\mathcal{D}}[1_{\alpha_2|\alpha_1}\cdot 1_{\beta_2|\beta_1}\cdot|A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|]\end{aligned} \tag{5}$$

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

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

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

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 equals $\binom{r}{w_1} \cdot \binom{k}{w_2} \cdot n^{2(D - w_1 - w_2)}$. 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 such that $\gamma \cdot \beta_1$ and $\gamma \cdot \beta_2$ are both multilinear and $\gamma \cdot \beta_1 = \gamma \cdot \beta_2$. This is equal to $\binom{N - 2k + w_2}{m - (k - w_2)}$. 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 7.3, we now upper bound the expression

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

Claim 7.4. *Let α_1, β_1 be monomials such that $\alpha_1\in\mathcal{M}^{[r]}$ and $\beta_1\in M(\alpha_1)$. Then*

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

Proof. From Claim 7.3, it follows that

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

is at most

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

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

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

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 3.6

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

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 $p=n^{-\epsilon}$ and $k=\theta(n)$, $g'(w_2)\leq k^{w_2}\cdot n^{\epsilon w_2-2w_2}\cdot\theta(1)^{w_2}$. In particular, $g'(w_2)$ is upper bounded by a decreasing function of w_2 and takes the maximum value 1 at $w=0$. Hence,

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

By a similar reasoning,

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

So

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

is upper bounded by

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

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

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

□

Now, plugging this bound back into Equation 5, we get

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

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,

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

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

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

So, we obtain Lemma 6.4.

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.

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 and $(k'+2)r' = n$. 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. 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 the 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.

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

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

Thus

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

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

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

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

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.

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 $(\overline{IMM}|_V)^{(i)}_X$ 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.

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 $\alpha, \beta \in \mathcal{T}(IMM|_V)$,

$$|\text{Supp}(\alpha) \setminus \text{Supp}(\beta)| = |\text{Supp}(\beta) \setminus \text{Supp}(\alpha)| \geq r' - r$$

The following lemma shows that such a set exists with a probability 1 over $V \leftarrow \mathcal{D}$.

Lemma 8.1. *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 $\alpha, \alpha' \in \mathcal{T}(IMM|_V)$,*

$$|\mathit{Supp}(\alpha) \setminus \mathit{Supp}(\alpha')| = |\mathit{Supp}(\alpha') \setminus \mathit{Supp}(\alpha)| \geq r' - r$$

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 $q = \tilde{n}^{3/4}$[5] for each of the Y matrices. 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 $IMM|_V$.

[5] 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.

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 8.1, we define

- $T_1(IMM|_V) = \sum_{\substack{\alpha\in\mathcal{T}(IMM|_V)\\ \beta\in\mathrm{Supp}(\partial_\alpha(IMM_{\tilde{n},n}))}} 1_{\alpha,\beta}\cdot|S_m(\alpha,\beta)|$
- $T_2(IMM|_V) = \sum_{\substack{\alpha\in\mathcal{T}(IMM|_V)\\ \beta,\gamma\in\mathrm{Supp}(\partial_\alpha(IMM_{\tilde{n},n}))\\ \beta\neq\gamma}} 1_{\alpha,\beta,\gamma}\cdot|S_m(\alpha,\gamma)\cap S_m(\alpha,\beta)|$
- $T_3(IMM|_V) = \sum_{\substack{\alpha_1,\alpha_2\in\mathcal{T}(IMM|_V)\\ \beta_1\in\mathrm{Supp}(\partial_{\alpha_1}(IMM_{\tilde{n},n}))\\ \beta_2\in\mathrm{Supp}(\partial_{\alpha_2}(IMM_{\tilde{n},n}))\\ (\alpha_1,\beta_1)\neq(\alpha_2,\beta_2)}} 1_{\alpha_1,\alpha_2,\beta_1,\beta_2}\cdot|A_m(\alpha_1,\beta_1)\cap A_m(\alpha_2,\beta_2)|$

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.
-
$$\begin{aligned}T_2|_V &= \sum_{\alpha\in\mathcal{T}(IMM|_V)}\left(\sum_{\beta,\gamma\in\overline{IMM}|_V(\alpha)} |S_m(\alpha,\gamma)\cap S_m(\alpha,\beta)|\right)\\ &= \sum_{\alpha\in\mathcal{T}(IMM|_V)}\left(\sum_{\beta,\gamma\in\overline{IMM}|_V(\alpha)} \binom{N-k-\Delta(\beta,\gamma)}{m}\right)\end{aligned}$$

  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.

- $$\begin{aligned} T_3|_V &= \sum_{\substack{\alpha_1,\alpha_2 \in \mathcal{T}(IMM|_V) \\ \beta_1 \in \overline{IMM}|_V(\alpha_1) \\ \beta_2 \in \overline{IMM}|_V(\alpha_2) \\ (\alpha_1,\beta_1)\neq(\alpha_2,\beta_2)}} |A_m(\alpha_1,\beta_1) \cap A_m(\alpha_2,\beta_2)| \\ &\leq \sum_{\substack{\alpha_1,\alpha_2 \in \mathcal{T}(IMM|_V) \\ \beta_1 \in \overline{IMM}|_V(\alpha_1) \\ \beta_2 \in \overline{IMM}|_V(\alpha_2) \\ (\alpha_1,\beta_1)\neq(\alpha_2,\beta_2)}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \end{aligned}$$

  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 γ of degree m such that $\gamma \cdot \beta_1$ and $\gamma \cdot \beta_2$ are both multilinear, and $\gamma \cdot \beta_1 = \gamma \cdot \beta_2$.

For every pair of monomials $\alpha, \alpha' \in \mathcal{T}(IMM|_V)$, we define

- $T_1|_V(\alpha) = \sum_{\beta \in \overline{IMM}|_V(\alpha)} |S_m(\alpha,\beta)|$
- $T_2|_V(\alpha) = \sum_{\beta,\gamma \in \overline{IMM}|_V(\alpha)} \binom{N-k-\Delta(\beta,\gamma)}{m}$
- 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 3.8 and Lemma 5.3 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 3.8 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 5.1, $\Phi_{\mathcal{G}_V,m}(IMM|_V)$ is large.

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 8.2. *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[6] 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 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 equal to $\frac{\binom{\tilde{n}-t}{\tilde{n}^{3/4}-t}}{\binom{\tilde{n}}{\tilde{n}^{3/4}}}$. 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 3.6. 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. □

[6] We say that a product gate survives the random restriction if none of the variables feeding in to it are set to zero.

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 8.3. *For all $\alpha \in \mathcal{T}(IMM|_V)$, for all $V \leftarrow \mathcal{D}$*

$$T_1|_V(\alpha) = D^k \cdot \binom{N-k}{m}.$$

Lemma 8.4.

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V] \le n^r \cdot D^k \cdot \binom{N-k}{m} \cdot n^{o(r)}$$

Lemma 8.5.

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_3|_V] \le n^r \cdot D^k \cdot O(n^{(4/\Lambda)r}) \cdot \binom{N-k}{m}$$

We will also need the following lemma, which implies Lemma 8.4 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 8.6. $\forall \alpha \in (\overline{IMM}|_V)_Y$,

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V(\alpha)] \le D^k \cdot \binom{N-k}{m} \cdot n^{o(r)}$$

We will prove these lemmas in Section 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 8.7.

$$Pr_{V \leftarrow \mathcal{D}}\left[T_2|_V < 20 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}] \wedge T_3|_V < 20 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_3|_{V'}]\right] \ge 0.9$$

Proof. The proof follows from the Markov's inequality and the union bound. □

Lemma 8.7 implies the following lemma, which we will use to prove a lower bound on the complexity of a random restriction of the $IMM_{\tilde{n},n}$.

Lemma 8.8. *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:*

$$|\mathcal{G}_V| \ge \frac{4}{5} \cdot |\mathcal{T}(IMM|_V)|$$

$$\forall \alpha \in \mathcal{G}_V, T_2|_V(\alpha) \le 100 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]/(n^r)$$

Proof. Let $V \subseteq \mathcal{V}$ be such that the bounds in Lemma 8.7 hold. Let $\mathcal{G}_V$ be the set of $\alpha \in \mathcal{T}(IMM|_V)$ such that $T_2|_V(\alpha) \le 100 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]/(n^r)$. We will now argue that $|\mathcal{G}_V| \ge \frac{4}{5} \cdot |\mathcal{T}(IMM|_V)|$. Let us assume this is not true, then $\sum_{\alpha \in \mathcal{T}(IMM|_V)} T_2|_V(\alpha) \ge \sum_{\alpha \in \mathcal{T}(IMM|_V) \setminus \mathcal{G}_V} T_2|_V(\alpha) > \frac{1}{5} \cdot 100 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]/(n^r) \cdot |\mathcal{T}(IMM|_V)| = 20 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]$ 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 8.9. *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*

$$\Phi_{\mathcal{G}_V,m}(IMM|_V) \ge \frac{n^r}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot D^k \cdot \binom{N-k}{m}$$

Proof. Lemma 8.8 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

$$|\mathcal{G}_V| \geq \frac{4}{5} \cdot |\mathcal{T}(IMM|_V)|$$
$$\forall \alpha \in \mathcal{G}_V, T_2|_V(\alpha) \leq 100 \cdot \mathbb{E}_{V' \leftarrow \mathcal{D}}[T_2|_{V'}]/(n^r).$$

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 3.8, 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 5.3, it follows that for all $\alpha \in \mathcal{G}_V$

$$\sum_{\beta \in \overline{IMM}|_V(\alpha)} |A_m(\alpha, \beta)| \geq \left| \bigcup_{\beta \in \overline{IMM}|_V(\alpha)} S_m(\alpha, \beta) \right| \geq \frac{1}{O(n^{o(r)})} \cdot D^k \cdot \binom{N-k}{m}$$

Consequently,

$$\sum_{\alpha \in \mathcal{G}_V} \sum_{\beta \in \overline{IMM}|_V(\alpha)} |A_m(\alpha, \beta)| \geq \frac{1}{O(n^{o(r)})} \cdot D^k \cdot \binom{N-k}{m} \cdot |\mathcal{G}_V| \geq \frac{n^r}{O(n^{o(r)})} \cdot D^k \cdot \binom{N-k}{m}$$

Also,

$$\sum_{\alpha_1, \alpha_2 \in \mathcal{G}_V} T_3|_V(\alpha_1, \alpha_2) \leq \sum_{\alpha_1, \alpha_2 \in \mathcal{T}(IMM|_V)} T_3|_V(\alpha_1, \alpha_2) = T_3|_V < 20\mathbb{E}_{V' \leftarrow \mathcal{D}}[T_3|_{V'}] \leq O(n^{(4/\Lambda)r}) \cdot n^r D^k \cdot \binom{N-k}{m},$$

and hence

$$\sum_{\substack{\alpha_1, \alpha_2 \in \mathcal{G}_V \\ \beta_1 \in \overline{IMM}|_V(\alpha_1) \\ \beta_2 \in \overline{IMM}|_V(\alpha_2) \\ (\alpha_1, \beta_1) \neq (\alpha_2, \beta_2)}} |A_m(\alpha_1, \beta_1) \cap A_m(\alpha_2, \beta_2)| = \sum_{\alpha_1, \alpha_2 \in \mathcal{G}_V} T_3|_V(\alpha_1, \alpha_2) \leq O(n^{(4/\Lambda)r}) \cdot n^r D^k \cdot \binom{N-k}{m}$$

So, we have

$$\sum_{\substack{\alpha_1, \alpha_2 \in \mathcal{G}_V \\ \beta_1 \in \overline{IMM}|_V(\alpha_1) \\ \beta_2 \in \overline{IMM}|_V(\alpha_2) \\ (\alpha_1, \beta_1) \neq (\alpha_2, \beta_2)}} |A_m(\alpha_1, \beta_1) \cap A_m(\alpha_2, \beta_2)| \leq O(n^{(4/\Lambda)r}) \cdot n^{o(r)} \cdot \sum_{\alpha \in \mathcal{G}_V} \sum_{\beta \in \overline{IMM}|_V(\alpha)} |A_m(\alpha, \beta)|$$

Therefore, by Lemma 3.8, we have

$$\left|\bigcup_{\substack{\alpha\in\mathcal{G}_V\\ \beta\in\overline{IMM}|_V(\alpha)}} A_m(\alpha,\beta)\right| \geq \frac{1}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}}\cdot \sum_{\substack{\alpha\in\mathcal{G}_V\\ \beta\in\overline{IMM}|_V(\alpha)}} |A_m(\alpha,\beta)| \geq \frac{n^r}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}}\cdot D^k\cdot \binom{N-k}{m}$$

Now by Lemma 5.1,

$$\Phi_{\mathcal{G}_V,m}(IMM|_V) \geq \frac{n^r}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}}\cdot D^k\cdot \binom{N-k}{m}$$

□

8.8 Wrapping up the proof

We will now complete the proof of the main theorem.

Theorem 8.10. *Any homogeneous* $\Sigma\Pi\Sigma\Pi$ *circuit computing the polynomial* $IMM_{\tilde{n},n}$ *has size at least* $2^{\Omega(\sqrt{n}\log n)}$.

Proof. Let C be a homogeneous $\Sigma\Pi\Sigma\Pi$ 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 8.2 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,

$$\Phi_{\mathcal{G}_V,m}(C|_V) \leq \text{Size}(C|_V)\cdot \binom{\lceil\frac{2n}{s}\rceil + r'}{r'}\cdot\binom{N}{m+r's} \tag{6}$$

From Lemma 8.9, 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

$$\Phi_{\mathcal{G}_V,m}(IMM|_V) \geq \frac{n^r}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}}\cdot D^k\cdot \binom{N-k}{m} \tag{7}$$

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

$$\text{Size}(C) \geq \frac{\frac{n^r}{O(n^{(4/\Lambda)r})\cdot n^{o(r)}}\cdot D^k\cdot\binom{N-k}{m}}{\binom{\lceil\frac{2n}{s}\rceil+r'}{r'}\cdot\binom{N}{m+r's}} \tag{8}$$

$$\tag{9}$$

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 8, we get

$$\begin{aligned}
\text{Size}(C) &\geq \frac{\frac{n^r}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot D^k \cdot \binom{N-k}{m}}{\binom{\lceil \frac{2n}{s} \rceil + r'}{r'} \cdot \binom{N}{m+r's}} \\
&\geq \frac{1}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot \left(\frac{N}{m}\right)^k \cdot 2^{-O(\sqrt{n})} \cdot \frac{m^{r's}}{N^k} \cdot \frac{(N-m)^k}{(N-m)^{r's}} \\
&= \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot \left(\frac{N-m}{m}\right)^{k-r's} \\
&= \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot \left(\frac{1+\frac{\ln n}{\Gamma\sqrt{n}}}{1-\frac{\ln n}{\Gamma\sqrt{n}}}\right)^{k-r's} && \text{by substituting } m = \frac{N}{2}\left(1 - \frac{\ln n}{\Gamma\sqrt{n}}\right) \\
&\geq \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot \left(1+\frac{\ln n}{\Gamma\sqrt{n}}\right)^{k-r's} \\
&\geq \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot e^{(n-2r'-r's)\frac{\ln n}{\Gamma\sqrt{n}}} && \text{since } k = n - 2r' \\
&\geq \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot n^{\frac{\sqrt{n}}{\Gamma} - \frac{r'(2+s)}{\Gamma\sqrt{n}}} \\
&\geq \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)r}) \cdot n^{o(r)}} \cdot n^{\frac{\sqrt{n}}{\Gamma} - \frac{\Lambda r(2+s)}{\Gamma\sqrt{n}}} \\
&\geq \frac{2^{-O(\sqrt{n})}}{O(n^{(4/\Lambda)\sqrt{n}}) \cdot n^{o(r)}} \cdot n^{\frac{\sqrt{n}-\Lambda s}{\Gamma}} && \text{by substituting } r = \sqrt{n}
\end{aligned}$$

Now, by substituting $\Lambda = 32$, $\Gamma = 2 + o(1)$ and $s = \frac{\sqrt{n}}{64}$, we obtain

$$\text{Size(C)} \geq 2^{-O(\sqrt{n})} \cdot n^{\Omega(\sqrt{n})}.$$

□

9 Calculations for $IMM_{\tilde{n},n}$

In this section, we provide the calculations which establish the bounds in Lemma 8.3, Lemma 8.4, Lemma 8.5. In the next section, we will first prove technical results that will be the building blocks of the lemmas.

9.1 Preliminary lemmas

Proposition 9.1. *For all $\beta \in \overline{IMM}_X$,*

$$\mathbb{E}_{V \leftarrow \mathcal{D}}\left[\sum_{\gamma \in (\overline{IMM}|_V)_X} D^{-\Delta(\beta,\gamma)}\right] \leq n^{o(r)}.$$

The proof follows from Lemma 9.2 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 9.2. *For all $\beta^{(i)} \in \overline{IMM}_X^{(i)}$,*

$$\mathbb{E}_{V \leftarrow \mathcal{D}}\left[\sum_{\gamma^{(i)} \in (\overline{IMM}|_V)_X^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}\right] \leq O(1).$$

Proof. The proof follows immediately from Lemmas 9.3 and 9.4 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 $\gamma^{(i)} \in (\overline{IMM}|_V)_X^{(i)}$ 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 $\gamma^{(i)} \in (\overline{IMM}|_V)_X^{(i)}$ 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 $\mathcal{A}_V^{(i)}(\beta^{(i)}) \cup \mathcal{B}_V^{(i)}(\beta^{(i)}) = (\overline{IMM}|_V)_X^{(i)}$.

Thus we have partitioned the set of $\gamma^{(i)} \in (\overline{IMM}|_V)_X^{(i)}$ into two sets $\mathcal{A}_V^{(i)}(\beta^{(i)})$ and $\mathcal{B}_V^{(i)}(\beta^{(i)})$, and we estimate the expression in Lemma 9.2 separately as $\gamma^{(i)}$ varies in these sets. This calculation is carried out in Lemmas 9.3 and 9.4 below.

Lemma 9.3. *For all $\beta^{(i)} \in \overline{IMM}_X^{(i)}$,*

$$\mathbb{E}_{V \leftarrow \mathcal{D}}\left[\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}\right] \leq O(1).$$

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 $\gamma^{(i)} \in (\overline{IMM}|_V)_X^{(i)}$ 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 Section 8.2. Thus we get that

$$\begin{aligned}\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i,j)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} &\leq \prod_{j'=j+1}^{k'} \mathsf{Deg}(X^{(i,j')}) \cdot D^{-(k'-j)} \\ &= D^j \prod_{j'=1}^{j} \mathsf{Deg}(X^{(i,j')})^{-1}\end{aligned}$$

Now for $j = 0$, the expression above equals 1. For $j > k' - 2\log n$, since $\mathsf{Deg}(X^{(i,j)}) = 1$, thus

$$\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i,j)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \le D^{-(k'-j)}.$$

For $j \le k' - 2\log n$, using the fact that $D < 2$, $\mathsf{Deg}(X^{(i,1)}) = n^\eta$ and $\mathsf{Deg}(X^{(i,j')}) = 2$ for $2 \le j' \le k' - 2\log n$, we get that

$$\begin{aligned}
\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i,j)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} &\le D^j \prod_{j'=1}^{j} \mathsf{Deg}(X^{(i,j')})^{-1} \\
&= \frac{D}{\mathsf{Deg}(X^{(i,1)})} \cdot \prod_{j'=2}^{j} \frac{D}{\mathsf{Deg}(X^{(i,j')})} \\
&\le \frac{2}{n^\eta}
\end{aligned}$$

Putting together these values for all values of j, and using the fact that $k' < n/2$, we get that

$$\begin{aligned}
\sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} &= \sum_{j=0}^{k'} \sum_{\gamma^{(i)} \in \mathcal{B}_V^{(i,j)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \\
&\le 1 + (k' - 2\log n) \cdot \frac{2}{n^\eta} + \sum_{j=k'-2\log n+1}^{k'} D^{-(k'-j)} \\
&\le 2 + \sum_{j=0}^{2\log n} D^{-j} \\
&\le 2 + \frac{1}{1 - D^{-1}} \\
&\le 5
\end{aligned}$$

□

Lemma 9.4. *For all* $\beta^{(i)} \in \overline{IMM}_X^{(i)}$,

$$\mathbb{E}_{V \leftarrow \mathcal{D}} \left[\sum_{\gamma^{(i)} \in \mathcal{A}_V^{(i)}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right] \le O(1/n).$$

Proof. For $\gamma^{(i)} \in \mathcal{A}_V^{(i)}(\beta^{(i)})$, we call a coordinate j such that $2 \le j \le k'$ 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.

$$\mathbb{E}_{V \leftarrow \mathcal{D}}\left[|\mathcal{A}^{(i)}_{V,t,T}(\beta^{(i)})|\right] \leq \prod_{j \in T} \mathsf{Deg}(X^{(i,j)}) \cdot \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}}.$$

Before the final computation, we need the following simple lemma:

Lemma 9.5. *$\forall i \in [r']$, $\forall T \subseteq [k']$, $\left(\prod_{j \in T} \mathsf{Deg}(X^{(i,j)})\right) \cdot D^{-|T|} \leq n^2$.*

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

$$\begin{aligned}
\mathbb{E}_{V \leftarrow \mathcal{D}}[\sum_{\gamma^{(i)} \in \mathcal{A}^{(i)}_{V,t,T}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}] &\leq \prod_{j \in T} \mathsf{Deg}(X^{(i,j)}) \cdot \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}} \cdot D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \\
&= \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}} \cdot \left(\prod_{j \in T} \mathsf{Deg}(X^{(i,j)})\right) \cdot D^{-|T|} \\
&\leq \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}} \cdot n^2. \qquad \text{(by Lemma 9.5)}
\end{aligned}$$

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

$$\mathbb{E}_{V \leftarrow \mathcal{D}}\left[\sum_{\gamma^{(i)} \in \mathcal{A}^{(i)}_{V,t}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}\right] \leq \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}} \cdot n^2 \cdot 2 \cdot \binom{k'}{t}.$$

Summing over the various choices of t, we get that

$$\mathbb{E}_{V \leftarrow \mathcal{D}}\left[\sum_{\gamma^{(i)} \in \mathcal{A}^{(i)}_{V}(\beta^{(i)})} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}\right] \leq \sum_{t=1}^{k'} \frac{1}{\tilde{n}^{\max\{1,(t-1)/2\}}} \cdot n^2 \cdot 2 \cdot \binom{k'}{t}.$$

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 9.1.

Proof of Proposition 9.1. For all $\beta \in \overline{IMM}_X$, observe that

$$\sum_{\gamma \in (\overline{IMM}|_V)_X} D^{-\Delta(\beta,\gamma)} = \prod_{i \in [r']} \sum_{\gamma^{(i)} \in (\overline{IMM}|_V)^{(i)}_X} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}.$$

Moreover, since the choice of $V \leftarrow \mathcal{D}$ chooses variables in distinct matrices independently, thus

$$\mathbb{E}_{V \leftarrow \mathcal{D}}\left[\sum_{\gamma \in (\overline{IMM}|_V)_X} D^{-\Delta(\beta,\gamma)}\right] = \prod_{i \in [r']} \mathbb{E}_{V \leftarrow \mathcal{D}}\left[\sum_{\gamma^{(i)} \in (\overline{IMM}|_V)^{(i)}_X} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}\right] \leq (O(1))^{r'} \leq n^{o(r)},$$

Where the second to last inequality follows from Lemma 9.2, and the last inequality follows form the fact that $r' = O(r)$.

□

9.2 Expected value of $T_1(IMM|_V)$

We now prove Lemma 8.3.

Proof of Lemma 8.3. For all $\alpha \in \mathcal{T}(IMM|_V)$,

$$\begin{aligned} T_1|_V(\alpha) &= \sum_{\beta \in \overline{IMM}|_V(\alpha)} |S_m(\alpha,\beta)| \\ &= \sum_{\beta \in \overline{IMM}|_V(\alpha)} \binom{N-k}{m} \\ &= D^k \cdot \binom{N-k}{m} \end{aligned}$$

□

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 8.4 and Lemma 8.6. 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 8.4 and Lemma 8.6.

Lemma 9.6. *For $\alpha \in \overline{IMM}_Y$ and $\beta \in \overline{IMM}(\alpha)$,*

$$\mathbb{E}_{V\leftarrow\mathcal{D}}[T_2|_V(\alpha,\beta)] \leq \binom{N-k}{m} \cdot n^{o(r)}.$$

Proof.

$$\begin{aligned}
T_2|_V(\alpha,\beta) &= \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} \binom{N-k-\Delta(\beta,\gamma)}{m} \\
&= \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} \binom{N-k-\Delta(\beta,\gamma)}{m} \cdot \frac{\binom{N-k}{m}}{\binom{N-k}{m}} \\
&= \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} \frac{\binom{N-k-\Delta(\beta,\gamma)}{m}}{\binom{N-k}{m}} \\
&\approx \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} \left(\frac{N-m}{N}\right)^{\Delta(\beta,\gamma)} && \text{by Lemma 3.6} \\
&\leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} D^{-\Delta(\beta,\gamma)}
\end{aligned}$$

Thus,

$$\mathbb{E}_{V\leftarrow\mathcal{D}}\left[T_2|_V(\alpha,\beta)\right] \leq \binom{N-k}{m} \cdot \mathbb{E}_{V\leftarrow\mathcal{D}}\left[\sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha) \\ \gamma \neq \beta}} D^{-\Delta(\beta,\gamma)}\right] \leq \binom{N-k}{m} \cdot n^{o(r)},$$

where the second inequality follows from Proposition 9.1. □

Proof of Lemma 8.6. $\forall \alpha \in (\overline{IMM}|_V)_Y$,

$$\begin{aligned}\mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V(\alpha)] &\leq \sum_{\beta \in \overline{IMM}|_V(\alpha)} \mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V(\alpha, \beta)] \\ &= D^k \cdot \binom{N-k}{m} \cdot n^{o(r)}.\end{aligned}$$

□

Proof of Lemma 8.4.

$$\begin{aligned}\mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V] &= \sum_{\alpha \in \mathcal{T}(IMM|_V)} \mathbb{E}_{V \leftarrow \mathcal{D}}[T_2|_V(\alpha)] \\ &\leq \sum_{\alpha \in \mathcal{T}(IMM|_V)} D^k \cdot \binom{N-k}{m} \cdot n^{o(r)} \\ &= n^r \cdot D^k \cdot \binom{N-k}{m} \cdot n^{o(r)}\end{aligned}$$

□

9.4 Expected value of $T_3(IMM|_V)$

We now prove Lemma 8.5.

Proof of Lemma 8.5. 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

$$T_3^{=}|_V(\alpha) = \sum_{\substack{\beta, \gamma \in \overline{IMM}|_V(\alpha) \\ \beta \neq \gamma}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \tag{10}$$

For $\alpha, \alpha' \in \overline{IMM}_Y$ such that $\alpha \neq \alpha'$, let

$$T_3^{\neq}|_V(\alpha, \alpha') = \sum_{\substack{\beta \in \overline{IMM}|_V(\alpha) \\ \gamma \in \overline{IMM}|_V(\alpha')}} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \tag{11}$$

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

$$\begin{aligned}\binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} &= \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \cdot \frac{\binom{N-k}{m}}{\binom{N-k}{m}} \\ &= \binom{N-k}{m} \cdot \frac{\binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)}}{\binom{N-k}{m}}\end{aligned}$$

Now by an application of Lemma 3.6, we obtain

$$\binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)} \approx \binom{N-k}{m} \cdot \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)} \tag{12}$$

Since by our choice of parameters $D < N/m$, plugging back Equation 12 into Equation 10, 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 9.1, we obtain

$$\mathbb{E}_{V\leftarrow\mathcal{D}}\left[T_3^{=}|_V(\alpha)\right] \leq \binom{N-k}{m} \cdot D^k \cdot n^{o(r)}.$$

and hence

$$\mathbb{E}_{V\leftarrow\mathcal{D}}\left[T_3^{=}|_V\right] \leq n^r \cdot \binom{N-k}{m} \cdot D^k \cdot n^{o(r)}. \tag{13}$$

Thus, 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

$$T_3^{\neq}|_V(\alpha,\alpha',\beta) = \sum_{\gamma\in\overline{IMM}|_V(\alpha')} \binom{N-k-\Delta(\beta,\gamma)}{m-\Delta(\beta,\gamma)}$$

Now by an application of Equation 12, 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 $\epsilon' = 2/\Lambda$ be a constant. We now partition the sum over γ into two parts, depending on whether $\Delta(\beta,\gamma) \geq (1-\epsilon')k$ or whether $\Delta(\beta,\gamma) < (1-\epsilon')k$. For $\alpha,\alpha' \in \mathcal{T}(IMM|_V)$ such that $\alpha \neq \alpha'$, and for $\beta \in \overline{IMM}|_V(\alpha)$, let

$$T_{3_{\text{large}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta) = \binom{N-k}{m} \cdot \left(\sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) \geq (1-\epsilon')k}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)} \right)$$

and

$$T_{3_{\text{small}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta) = \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\epsilon')k}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)}$$

Thus

$$\begin{aligned}
T_{3_{\text{large}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta) &\leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) \geq (1-\epsilon')k}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)} \\
&= \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) \geq (1-\epsilon')k}} \left(\frac{N-m}{N}\right)^{\Delta(\beta,\gamma)} \cdot \left(\frac{m}{N-m}\right)^{\Delta(\beta,\gamma)} \\
&\leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) \geq (1-\epsilon')k}} \left(\frac{N-m}{N}\right)^{\Delta(\beta,\gamma)} \cdot \left(\frac{m}{N-m}\right)^{(1-\epsilon')k} \qquad \left(\text{since } \frac{m}{N-m} < 1\right)
\end{aligned}$$

Now, by our choice of parameters, $\left(\frac{m}{N-m}\right)^k = n^{-r}$ and $D = \frac{N}{N-m}$, we get

$$T_{3_{\text{large}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta) \leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) \geq (1-\epsilon')k}} D^{-\Delta(\beta,\gamma)} \cdot n^{-(1-\epsilon')r}$$

From here, by applying Proposition 9.1, we obtain

$$\mathbb{E}_{V \leftarrow \mathcal{D}}\left[T_{3_{\text{large}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta)\right] \leq \binom{N-k}{m} \cdot n^{o(r)} \cdot n^{-(1-\epsilon')r} \leq \binom{N-k}{m} \cdot O(n^{(2\epsilon'-1)r}), \qquad (14)$$

We will now bound

$$T_{3_{\text{small}\Delta}}^{\neq}|_V(\alpha,\alpha'\beta) = \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\epsilon')k}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)}$$

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 $\beta \in \overline{IMM}|_V(\alpha)$,

$$\begin{aligned}
T^{\neq}_{3_{\text{small}\Delta}}|_V(\alpha, \alpha'\beta) &\leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\epsilon')k}} \left(\frac{m}{N}\right)^{\Delta(\beta,\gamma)} \\
&= \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\epsilon')k}} \left(\frac{N-m}{N}\right)^{\Delta(\beta,\gamma)} \cdot \left(\frac{m}{N-m}\right)^{\Delta(\beta,\gamma)} \\
&\leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\epsilon')k}} \left(\frac{N-m}{N}\right)^{\Delta(\beta,\gamma)} \qquad \left(\text{since } \frac{m}{N-m} < 1\right) \\
&= \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\epsilon')k}} D^{-\Delta(\beta,\gamma)}
\end{aligned}$$

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 $\Delta(\gamma, \beta) < (1-\epsilon')k = (1-\epsilon')k'r'$, thus $|A(\beta, \gamma)| \geq \epsilon' r'$. Also, let $B(\alpha, \alpha') \subseteq [r']$ be the set of all $i \in [r']$ such that $\alpha^{(i)} = \alpha'^{(i)}$. Then by Lemma 8.1, for $\alpha \neq \alpha'$, $|B(\alpha, \alpha')| \leq r$.

Claim 9.7. *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 $\Delta(\beta, \gamma) < (1-\epsilon')k$. 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 $|A(\beta, \gamma) \setminus B(\alpha, \alpha')| \geq \epsilon' r' - r$.*

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 $x^{(i,1)}_{s',t'}$f or 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 $|C| = \epsilon' r' - r$, 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 $\Delta(\beta, \gamma) < (1-\epsilon')k$ 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 restriction. 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

$$T^{\neq}_{3_{\text{small}\Delta}}|_V(\alpha, \alpha'\beta) \leq \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\epsilon')k}} D^{-\Delta(\beta,\gamma)}$$

$$= \binom{N-k}{m} \cdot \sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha') \\ \Delta(\gamma,\beta) < (1-\epsilon')k}} \prod_{i \in [r']} D^{-\Delta(\beta^{(i)},\gamma^{(i)})}$$

$$\leq \binom{N-k}{m} \cdot \sum_{\substack{C \subset [r'], \\ |C| = \epsilon' r' - r}} \left(\sum_{\gamma \in M_C(\beta,\alpha')} \left(\prod_{i \in C} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \cdot \prod_{i \in [r'] \setminus C} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right) \right)$$

$$\leq \binom{N-k}{m} \cdot \sum_{\substack{C \subset [r'], \\ |C| = \epsilon' r' - r}} \left(\prod_{i \in C} \left(\sum_{\gamma^{(i)} \in M_C(\alpha')^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right) \cdot \prod_{i \in [r'] \setminus C} \left(\sum_{\gamma^{(i)} \in M_C(\alpha')^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right) \right)$$

Now, observe that $i \in C$, $M_C(\alpha')^{(i)} \subseteq \mathcal{A}_V^{(i)}(\beta^{(i)})$. Thus, by Lemma 9.4 and Lemma 9.2, we get that

$$\mathbb{E}_{V \leftarrow \mathcal{D}} \left[T^{\neq}_{3_{\text{small}\Delta}}|_V(\alpha, \alpha'\beta) \right] \leq \binom{N-k}{m} \cdot \mathbb{E}_{V \leftarrow \mathcal{D}} \left[\sum_{\substack{\gamma \in \overline{IMM}|_V(\alpha'), \\ \Delta(\gamma,\beta) < (1-\epsilon')k}} D^{-\Delta(\beta,\gamma)} \right]$$

$$\leq \binom{N-k}{m} \cdot \sum_{\substack{C \subset [r'], \\ |C| = \epsilon' r' - r}} \left(\prod_{i \in C} \mathbb{E}_{V \leftarrow \mathcal{D}} \left[\left(\sum_{\gamma^{(i)} \in M_C(\alpha')^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right) \right] \prod_{i \in [r'] \setminus C} \mathbb{E}_{V \leftarrow \mathcal{D}} \left[\left(\sum_{\gamma^{(i)} \in M_C(\alpha')^{(i)}} D^{-\Delta(\beta^{(i)},\gamma^{(i)})} \right) \right] \right)$$

$$\leq \binom{N-k}{m} \cdot \binom{r'}{\epsilon' r' - r} \cdot \left(O\left(\frac{1}{n}\right) \right)^{\epsilon' r' - r} 2^{O(r')}$$

$$= \binom{N-k}{m} \cdot \left(O\left(\frac{1}{n}\right) \right)^{\epsilon' r' - r} \cdot n^{o(r)}$$

Thus since $\epsilon' r' - r > r$,

$$\mathbb{E}[T^{\neq}_{3_{\text{small}\Delta}}(\alpha, \alpha'\beta)] \leq \binom{N-k}{m} \cdot \left(\frac{1}{n}\right)^{\epsilon' r' - r} \cdot n^{o(r)} \leq \binom{N-k}{m} \cdot n^{-r+o(r)}.$$

Putting this together with earlier computation showing that

$$\mathbb{E}[T^{\neq}_{3_{\text{large}\Delta}}(\alpha, \alpha'\beta)] \leq \binom{N-k}{m} \cdot O(n^{(2\epsilon'-1)r}),$$

we conclude that

$$\mathbb{E}[T^{\neq}_{3}(\alpha, \alpha'\beta)] \leq \binom{N-k}{m} \cdot O(n^{(2\epsilon'-1)r}).$$

Summing over $\beta \in \overline{IMM}|_V(\alpha)$, we get that

$$\mathbb{E}_{V \leftarrow \mathcal{D}} \left[T^{\neq}_{3}|_V(\alpha, \alpha') \right] \leq \binom{N-k}{m} \cdot D^k \cdot n^{(2\epsilon'-1)r}.$$

Summing over $\alpha, \alpha' \in \mathcal{T}(IMM|_V)$ such that $\alpha \neq \alpha'$, we get that

$$\mathbb{E}_{V \leftarrow \mathcal{D}} \left[T_3^{\neq}|_V \right] \leq n^{2r} \cdot \binom{N-k}{m} \cdot D^k \cdot n^{(2\epsilon'-1)r} = n^r \cdot \binom{N-k}{m} \cdot D^k \cdot n^{(2\epsilon')r}.$$

Putting this together with the bound in Equation 13, we conclude that

$$\mathbb{E}_{V \leftarrow \mathcal{D}}[T_3|_V] \leq n^r \cdot \binom{N-k}{m} \cdot D^k \cdot n^{\frac{4}{\Lambda}r}.$$

□

10 Open problems

Our results (and those by [KLSS14]) give $n^{\Omega(\sqrt{n})}$ lower bounds for polynomials computed by homogeneous $\Sigma\Pi\Sigma\Pi$ circuits. This suggests a very natural strategy of trying to prove lower bounds for any class of circuits $\mathcal{C}$. If one can show that some polynomial $P \in \mathcal{C}$ can be computed by a $n^{o(\sqrt{n})}$ sized homogeneous $\Sigma\Pi\Sigma\Pi$ circuit, then our results would immediately imply a lower bound for $\mathcal{C}$.

Recall that the depth reduction of Tavenas [Tav13] shows that ever polynomial inVP can be expressed as a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit of size $n^{O(\sqrt{n})}$. Unfortunately since our lower bounds hold for a polynomial in VP, thus the bound on the size of the depth 4 circuit obtained in the depth reduction cannot be improved. Although they cannot be improved for general circuits in VP, they might be possible to improve for other rich and interesting classes of circuits such as formulas or even homogeneous formulas. (The results of [KS] had shown that a more efficient depth reduction for homogeneous formulas is not possible when one wants to reduce to homogeneous $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits, but for general homogeneous $\Sigma\Pi\Sigma\Pi$ circuits this might still be possible.)

Another more general question that seems even more important now is to truly understand the potential of the shifted partial derivative method (and its variants as used in this work and earlier works) for proving lower bounds for general arithmetic circtuits. These techniques do seem to be giving significantly stronger lower bounds than we were able to show some years ago. Do they have the potential of separating VP from VNP? Or is there some inherent underlying reason that suggests we might need different techniques?

References

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

[BS83] Walter Baur and Volker Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22:317–330, 1983.

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

[CM14] Suryajith Chillara and Partha Mukhopadhyay. On the limits of depth reduction at depth 3 over small finite fields. *CoRR*, abs/1401.0189, 2014.

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

[GK98] Dima Grigoriev and Marek Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing (STOC)*, pages 577–582, 1998.

[GKKS13a] A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi. Approaching the chasm at depth four. *In Proceedings of CCC*, 2013.

[GKKS13b] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth three. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:26, 2013.

[GR98] D. Grigoriev and A. Razborov. Exponential complexity lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, FOCS '98, pages 269–278, 1998.

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

[KLSS] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. Superpolynomial lower bounds for depth-4 homogeneous arithmetic formulas. *To appear in STOC' 2014.*

[KLSS14] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An exponential lower bound for homogeneous depth four arithmetic formulas. *Electronic Colloquium on Computational Complexity (ECCC)*, 2014.

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

[KS] Mrinal Kumar and Shubhangi Saraf. The limits of depth reduction for arithmetic formulas: It's all about the top fan-in. *To appear in STOC' 2014.*

[KS13] Mrinal Kumar and Shubhangi Saraf. Superpolynomial lower bounds for general homogeneous depth 4 arithmetic circuits. *CoRR*, abs/1312.5978, 2013.

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

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

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

[Str73] V. Strassen. Die berechnungskomplexiät von elementarsymmetrischen funktionen und von interpolationskoeffizienten. *Numer. Math*, 20:238–251, 1973.

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

[Val79] L. G. Valiant. Completeness classes in algebra. In *Proceedings of the 11th Annual STOC*, STOC '79, pages 249–261, New York, NY, USA, 1979. ACM.

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

A Proof of Lemma 3.7

Proof. We will prove the lemma via contradiction. We will in fact, show that

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

Since, for all x, $f(x) \leq g(x)$, this would imply that

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

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. □

B Proof of Lemma 3.8

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. □

ECCC ISSN 1433-8092

http://eccc.hpi-web.de