An Exponential Lower Bound for Homogeneous Depth Four Arithmetic Formulas

Neeraj Kayal
Microsoft Research India
neeraka@microsoft.com

Nutan Limaye
Indian Institute of Technology Bombay
nutan@cse.iitb.ac.in

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

Srikanth Srinivasan
Indian Institute of Technology Bombay
srikanth@math.iitb.ac.in

April 1, 2014

Abstract

We show here a $2^{\Omega(\sqrt{d}\cdot\log N)}$ size lower bound for homogeneous depth four arithmetic formulas. That is, we give an explicit family of polynomials of degree d on N variables (with $N = d^3$ in our case) with $0,1$-coefficients such that for any representation of a polynomial f in this family of the form

$$f = \sum_i \prod_j Q_{ij},$$

where the Q_{ij}'s are homogeneous polynomials (recall that a polynomial is said to be homogeneous if all its monomials have the same degree), it must hold that

$$\sum_{i,j} (\text{Number of monomials of } Q_{ij}) \quad \geq \quad 2^{\Omega(\sqrt{d}\cdot\log N)}.$$

The abovementioned family which we refer to as the Nisan-Wigderson design-based family of polynomials, is in the complexity class VNP. For polynomial families in VP we show the following: Any homogeneous depth four arithmetic formula computing the Iterated Matrix Multiplication polynomial $\text{IMM}_{n,d}$ — the $(1,1)$-th entry of the product of d generic $n \times n$ matrices — has size $n^{\Omega(\log n)}$, if $d = \Omega(\log^2 n)$. Moreover, any homogeneous depth four formula computing the determinant polynomial Det_n — the determinant of a generic $n \times n$ matrix — has size $n^{\Omega(\log n)}$. Our work builds on the recent lower bound results [Kay12, GKKS13a, KSS13, FLMS13, KS13c] and yields an improved quantitative bound as compared to the recent indepedent work of [KS13b].

1 Introduction

The problem of proving super-polynomial lower bounds for arithmetic circuits occupies a central position in algebraic complexity theory, much like the problem of proving super-polynomial lower bounds for boolean circuits does in Boolean complexity. The model of arithmetic circuits is an algebraic analogue of the model of Boolean circuits: an arithmetic circuit contains addition ($+$) and multiplication ($\times$) gates and it naturally computes a polynomial in the input variables over some underlying field. We typically allow the input edges to a $+$ gate to be labelled with arbitrary constants from the underlying field $\mathbb{F}$ so that a $+$ gate can in fact compute an arbitrary $\mathbb{F}$-linear combination of its inputs. Proving super-polynomial arithmetic circuit lower bounds for an explicit family of polynomials, say the Permanent family, amounts to showing that $\mathsf{VP} \neq \mathsf{VNP}$ [Val79]. The complexity classes VP and VNP consist of families of polynomials and can be viewed as algebraic analogues of the classes P and NP respectively[1]. The hope is that it might be possible to use algebraic and geometric insights along with the structure of arithmetic circuits to make progress towards settling the VP vs VNP question. Till date, research on arithmetic circuits has produced several interesting results that have enriched our understanding of the lower bound problem and the related problems on polynomial identity testing & reconstruction (or learning) of arithmetic circuits. The survey [SY10] gives an account of some of the results and outstanding open questions in this area.

Constant Depth Circuits. While the available lower bounds for general circuits are very modest, progress in this direction has been made in the form of lower bounds for restricted (but still nontrivial and interesting) subclasses of arithmetic circuits. One such restriction is in the form of the depth of a circuit[2]. The study of constant depth circuits has gained momentum in the recent years after a striking connection was shown between lower bounds for general circuits and that for depth-4 & depth-3 formulas. First recall that a polynomial f is said to be *homogeneous* if all its monomials have the same degree. An arithmetic circuit is said to be homogeneous if it is syntactically homogeneous, i.e. at every intermediate $+$ gate, the inputs all have the same formal degree[3]. Building on the *depth reduction* results of [VSBR83, AJMV98], a string of works [AV08, Koi12, Tav13] arrived at the following result: A $2^{\omega(\sqrt{d}\log N)}$ size lower bound for depth-4 homogeneous formulas[4], computing a degree-d, N-variate polynomial (in a polynomial family), implies a super-polynomial lower bound for general circuits. Further, if the polynomial family belongs to VNP then such a lower bound would imply $\mathsf{VP} \neq \mathsf{VNP}$[5].

Previous work on super-polynomial lower bounds. Raz [Raz09] showed that any multilinear formula computing the determinant Det_n (or the permanent Perm_n) polynomial has $n^{\Omega(\log n)}$ size

[1] It is known that if VNP can be computed by arithmetic circuits of polynomial size size and degree and which have the additional property that the constants from the underlying field have polynomially bounded bitlengths then it must follow that $\mathsf{P} = \mathsf{NP}$ (cf. [SV85]).

[2] Recall that the depth of a circuit is the maximum length of any path from an input node to the output node.

[3] The formal degree of a node in a circuit is defined inductively in the natural manner - leaf nodes labelled with variables (respectively with field constants) have formal degree 1 (respectively zero) and every internal $+$ gate (resp. $\times$ gate) is said to have formal degree equal to the maximum of (resp. the sum of) the formal degrees of its children.

[4] with bottom fan-in bounded by $O(\sqrt{d})$

[5] A similar implication is true even for depth-3 formulas, although at the loss of the homogeneity condition - due to [GKKS13b].

with subsequent separations[6] and refinements[7] in [Raz06] and in [RY09]. The formal degree of a homogeneous formula is bounded by the degree of the computed polynomial - a feature that is quite effective in proving lower bounds using *partial derivatives* based methods. The approach of proving lower bounds by studying the space of partial derivatives of the computed polynomial was introduced by Nisan and Wigderson [NW97], who showed an exponential lower bound for homogeneous depth-3 formulas[8]. (For depth-3 formulas over fixed finite fields, an exponential lower bound was shown by [GK98, GR98].) Indeed, the super-polynomial lower bounds obtained by [Raz09, Raz06, RY09], as well as some others like [ASSS12], are based upon studying partial derivatives or associated matrices involving partial derivatives like the Jacobian or the Hessian[9]. The situation for depth-4 homogeneous formulas is substantially rectified by the recent work of [Kay12, GKKS13a], followed by the work of [KSS13] and [FLMS13]. These works have led to a $2^{\Omega(\sqrt{d}\log N)}$ lower bound for depth-4 homogeneous formulas with bottom fan-in $O(\sqrt{d})$ (where d is the degree of the N-variate 'target' polynomial on which the lower bound is shown). Further, [KSS13] and [FLMS13] together imply a super-polynomial separation between *algebraic branching programs* and *regular formulas* - two natural sub-models of arithmetic circuits[10]. A seemingly tempting problem left open in these works is, if the lower bound of $2^{\Omega(\sqrt{d}\log N)}$ in the above statement could be improved to $2^{\omega(\sqrt{d}\log N)}$, a super-polynomial lower bound for general circuits would ensue immediately. Another recent work [KS13a] has shown an exponential lower bound for depth-4 homogeneous formulas with constant top fan-in. At the heart of these results lies the study of the space of *shifted partial derivatives* of polynomials and an associated measure called the *dimension of the shifted partials* - a technique introduced in [Kay12, GKKS13a]. Loosely speaking, the dimension of the shifted partials of a polynomial g refers to the dimension of the $\mathbb{F}$-linear vector space generated by the set of polynomials obtained by multiplying (shifting) the partial derivatives of g with monomials of suitable degrees.

Our results. In an attempt to understand the strength of the shifted partials method better, a recurring open problem stated in [KSS13, FLMS13, KS13a, Tav13] is to show a super-polynomial lower bound for homogeneous depth-4 formulas. Whether the shifted partial measure can be used to prove such a result or not is not exactly clear to us. This very recent work by [KS13c] seems to suggest that the answer is likely in the negative. However, this does not rule out the possibility of using a different measure, perhaps closely related to the shifted partials, to achieve the same. It turns out that indeed it is possible to use a slightly modified (or augmented) version of the shifted partial measure to show an exponential lower bound for depth-4 homogeneous formulas. For the ease of reference in this paper, we will call this modified measure the *projected shifted partials*. Loosely speaking, the idea is to *shift the derivatives of a polynomial by a carefully chosen set of monomials and then view these after 'projecting' them to an appropriate set of monomials*. Our results are formally stated below.

Theorem 1. *Let $\mathbb{F}$ be any field of characteristic zero. There is an explicit family of polynomials*

[6] Building upon [Raz09], a super-polynomial gap between multilinear circuits and formulas was obtained in [Raz06].

[7] Also building upon [Raz09], a significantly better bound was later shown for bounded (i.e. constant) depth multilinear circuits [RY09]: A depth-d multilinear circuit computing Det_n or Perm_n has size $2^{n^{\Omega(1/d)}}$.

[8] Prior to this work, Nisan [Nis91] showed an exponential lower bound for noncommutative arithmetic formulas

[9] A recent survey by Chen, Kayal and Wigderson [CKW11] gives more applications of partial derivatives.

[10] In fact, a very recent work of [KS13c] shows a super-polynomial separation between *general formulas* and *regular formulas*.

of degree d in $N = d^3$ variables with zero-one coefficients such that any homogeneous $\Sigma\Pi\Sigma\Pi$ formula over $\mathbb{F}$ computing this family must have size at least $2^{\Omega(\sqrt{d}\cdot\log N)}$. In other words, for any representation of the degree d polynomial f in the family, of the form

$$f = \sum_i \prod_j Q_{ij},$$

where the Q_{ij}'s are homogeneous polynomials, it must hold that

$$\sum_{i,j} (\textit{Number of monomials of } Q_{ij}) \quad \geq \quad 2^{\Omega\left(\sqrt{d}\cdot\log N\right)}.$$

Remark 2. The above theorem continues to hold for any $N \geq d^{2+\varepsilon}$, for any constant $\varepsilon > 0$. For clarity of presentation, we work with $N = d^3$ in what follows.

The explicit polynomial f in the theorem above is a variant of the Nisan-Wigderson design-based polynomial introduced in [KSS13] and further studied in [KS13c, KS13b]. While this family of polynomials is explicit (in VNP), it is not known to be efficiently computable. Thus, as it stands, our main theorem has two limitations - it is valid only over fields of characteristic zero and the explicit family of polynomials that we give is not known to be efficiently computable. We partially address these limitations by proving super-polynomial lower bounds which are quantitatively worse but hold over any field for a family of polynomials in VP.

Theorem 3. *Over* any *field $\mathbb{F}$, any depth-4 homogeneous formula computing the Iterated Matrix Multiplication polynomial $IMM_{n,d}$ — the $(1,1)$-th entry of the product of d generic $n \times n$ matrices — has $n^{\Omega(\log n)}$ size, assuming $d = \Omega(\log^2 n)$. If $d \leq \varepsilon \log^2 n$ for a sufficiently small $\varepsilon > 0$ then any depth-4 homogeneous formula computing the $IMM_{n,d}$ polynomial has size $n^{\Omega(\sqrt{d})}$.*

We note that these bounds are incomparable with those proved in [FLMS13]. While [FLMS13] prove an exponential lower bound — $2^{\Omega(\sqrt{d})\log n}$ for $d = n^{\Omega(1)}$ — for depth-4 homogeneous *multilinear* formulas computing $\mathrm{IMM}_{n,d}$, we prove a weaker quasipolynomial lower bound for the more general model of depth-4 homogeneous formulas.

Theorem 4. *Over any field $\mathbb{F}$, any depth-4 homogeneous formula computing the determinant polynomial Det_n — the determinant of a generic $n \times n$ matrix — has size $n^{\Omega(\log n)}$.*

The rest of the paper is devoted to proving these results.

A recent independent result by [KS13b]. Very recently, Kumar and Saraf [KS13b] independently proved a superpolynomial ($n^{\Omega(\log \log n)}$) lower bound for homogeneous depth four circuits using another nice augmentation of the shifted partial measure that they call *bounded support shifted partials*. We do not know if this measure can be used to prove an exponential lower bound. Indeed, they explicitly state the problem of proving exponential lower bounds for homogeneous depth four circuits which we happen to achieve here.

2 Overview of our proof

We now give an outline of the proof of theorem 1. Let $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be a polynomial of degree d on N variables over a field $\mathbb{F}$. Consider a representation of f of the form

$$f = \sum_{i=1}^{s} \prod_{j} Q_{ij}, \tag{1}$$

where the Q_{ij}'s are homogeneous polynomials. Note that any polynomial can be written in this way - the challenge is to prove a lower bound on the total number of monomials appearing the Q_{ij}'s. For each $i \in [s]$, the i-th term in such a representation is defined to be $T_i = \prod_j Q_{ij}$. First observe that we can assume without loss of generality that the degree of each term T_i is at most d (as we can simply discard terms of degree larger than d without changing the output). So now assume that the total number of monomials in this representation is small, say $2^{o(\sqrt{d} \cdot \log N)}$ (else we have nothing to prove). In particular, our assumption means that every Q_{ij} has at most $2^{o(\sqrt{d} \cdot \log N)}$ monomials.

Using Random Restrictions to reduce the support size. In the first step, we consider the identity (1) and in that set each variable to zero independently at random with probability $(1-p)$ (a variable is left untouched with probability p.) Then any monomial m in any of the Q_{ij}'s which contains t distinct variables will now survive (i.e. remain nonzero under this substitution) with probability p^t. So if we choose $p = \frac{1}{d^{\Theta(1)}}$ then via an application of the union bound we deduce that all monomials of support at least $t = \Omega(\sqrt{d})$ will be 'killed' (i.e. set to zero) under this substitution[11]. For ease of subsequent exposition, let us introduce the following notation/terminology.

1. **Support.** Let $m = x_1^{e_1} \cdot x_2^{e_2} \cdot \ldots \cdot x_N^{e_N}$ in $\mathbb{F}[x_1, x_2, \ldots, x_N]$ be a monomial. The support of m, denoted $\mathrm{Supp}(m)$ is the subset of variables appearing in it, i.e.

   $$\mathrm{Supp}(m) \stackrel{\text{def}}{=} \{i \ : \ e_i \geq 1\} \subseteq [N].$$

   The support size of a polynomial f, denoted $|\mathrm{Supp}(f)|$ is the maximum support size of any monomial appearing in f.

2. **Substitution maps.** Let $R \subseteq [N]$ be a set. The substitution map $\sigma_R : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}[\mathbf{x}]$ is the map which sets all the variables in R to zero, i.e. $\sigma_R(f) \stackrel{\text{def}}{=} f|_{x_i=0 \ \forall i \in R}$. Formally, $\sigma_R : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}[\mathbf{x}]$ is a homomorphism such that for any monomial $m \in \mathbb{F}[\mathbf{x}]$, $\sigma_R(m) = m$ if the monomial m is supported outside R and is zero otherwise.

So the above discussion can now be summarized as follows. Let $t = \Theta(\sqrt{d})$ be a suitable integer. By choosing a set R at random in the above manner and applying σ_R to the identity (1), we obtain (with high probability) another identity

$$\sigma_R(f) = \sum_{i=1}^{s} \prod_{j} \sigma_R(Q_{ij}), \quad \text{where } \forall i,j: \ \sigma_R(Q_{ij}) \text{ is homogeneous and } |\mathrm{Supp}(\sigma_R(Q_{ij}))| \leq t. \tag{2}$$

[11] This reduction from homogeneous $\Sigma\Pi\Sigma\Pi$ formulas to low support $\Sigma\Pi\Sigma\Pi$ formulas was communicated to the first author by Avi Wigderson. It was recently exploited by Kumar and Saraf in [KS13b] and also independently discovered by some of the other authors of the present work.

In this manner our problem reduces to proving lower bounds for representations of the form (2) which we refer to as t-supported homogeneous $\Sigma\Pi\Sigma\Pi$ circuits.

Lower bounds for low support homogeneous $\Sigma\Pi\Sigma\Pi$ circuits. We first note that the degree of a polynomial is an upper bound on its support size. Prior work by [Kay12, GKKS13a, KSS13, FLMS13] had proved lower bounds for similar looking representations but in which the degree of every Q_{ij}, rather than its support was bounded by t. We build on this work to devise a complexity measure that we refer to as *dimension of projected shifted partials*. We define this measure as follows.

1. **The projection map.** Let $s, e \geq 1$ be integers. The linear map $\pi_{e,s} : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}[\mathbf{x}]$ maps a polynomial $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ to the component of degree e and support s of $f(\mathbf{x})$. Formally, it is defined as follows. We need to only specify it for monomials and it then extends by linearity to all of $\mathbb{F}[\mathbf{x}]$. For a monomial $m \in \mathbb{F}[\mathbf{x}]$, $\pi_{e,s}(m)$ equals m if m has degree exactly e and support size exactly s and zero otherwise.

2. **The Complexity Measure.** Let k, ℓ, r be integer parameters and $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be a multivariate polynomial. $\boldsymbol{\partial}^{=k} f$ shall denote the set of all k-th order partial derivatives of f. Let $\mathbf{x}^{(=\ell,=s)}$ denote the set of monomials of degree exactly ℓ and support exactly s over the variables in $\mathbf{x}$. Let $A, B \subseteq \mathbb{F}[\mathbf{x}]$ be any two sets of polynomials. $A \cdot B$ stands for the set

$$A \cdot B \stackrel{\text{def}}{=} \{f \cdot g \ : \ f \in A \text{ and } g \in B\}.$$

   For a linear map $\pi : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}[\mathbf{x}]$, $\pi(A)$ denotes the set

$$\pi(A) \stackrel{\text{def}}{=} \{\pi(f) \ : \ f \in A\}.$$

   The *dimension of projected shifted partial derivatives* of f (DPSP for short) is defined as

$$\mathsf{DPSP}_{k,\ell,e}(f) \stackrel{\text{def}}{=} \dim\left(\pi_{\ell+e,\ell+e}\left(\mathbf{x}^{(=\ell,=\ell)} \cdot \boldsymbol{\partial}^{=k} f\right)\right).$$

Intuitively, by shifting (i.e. multiplying) the partial derivatives by a carefully chosen set of monomials and then projecting them to another appropriate set of monomials, we are able to ignore factors (paying a relatively small cost) in which the support size is much larger than the degree so that effectively we are reduced to the case where all the Q_{ij}'s have small degree. Specifically, we show that this measure is relatively small for t-supported homogeneous $\Sigma\Pi\Sigma\Pi$ circuits (corollary 14 in section 4). We then find an explicit polynomial f whose projected shifted partials has large dimension and thereby obtain a $2^{\Omega\left(\frac{d}{t}\cdot\log N\right)}$ lower bound for t-supported homogeneous $\Sigma\Pi\Sigma\Pi$ circuits computing f. We further show that the dimension of projected shifted partials of f remains quite large even under random restrictions (with high probability) thereby obtaining a $2^{\Omega\left(\sqrt{d}\cdot\log N\right)}$ lower bound overall for general homogeneous $\Sigma\Pi\Sigma\Pi$ circuits.

Lower bounding the dimension of projected shifted partials. A crucial component of this proof is to show that the dimension of projected shifted partials of our explicit family of polynomials is large[12]. From the definition, it follows that this quantity is equal to the rank of a certain matrix

[12] In prior work one needed to estimate the dimension of shifted partials of a given f and it was shown that in many interesting cases this could be successfully accomplished simply by counting leading monomials. It seems that counting of leading monomials perhaps will not yield good enough estimates for the modified measure used here.

$M(f)$ whose entries are zero or equal to the coefficient of an appropriate monomial of f. In order to show that $\text{rank}(M(f))$ is large for our choice of f, we use a lemma by Noga Alon[13] [Alo09]. It goes as follows. Let $B(f) \stackrel{\text{def}}{=} M(f)^T \cdot M(f)$. Clearly, the rank of $B(f)$ is a lower bound on the rank of $M(f)$. By an application of Cauchy-Schwarz on the vector of nonzero eigenvalues of $B(f)$, one sees that

$$\text{rank}(B(f)) \geq \frac{\text{Tr}(B(f))^2}{\text{Tr}(B(f)^2)}.$$

We estimate $\text{Tr}(B(f))^2$ and $\text{Tr}(B(f)^2)$ and show that the ratio is large for our choice of f (even under random restrictions).

Organization. The rest of the paper is devoted to fleshing out this outline into a full proof. For the sake of clarity of exposition, we first focus our attention on t-supported homogeneous $\Sigma\Pi\Sigma\Pi$ circuits. We first give an upper bound (in section 4) on the dimension of projected shifted partials of any homogeneous t-supported $\Sigma\Pi\Sigma\Pi$ circuit $\mathcal{C}$. In section 5 we then give the construction of our polynomial f and show that choosing the parameters appropriately yields a lower bound of $2^{\Omega\left(\frac{d}{t}\cdot\log N\right)}$ on the top fanin of homogeneous t-supported $\Sigma\Pi\Sigma\Pi$ circuits computing f - assuming that f has large projected shifted partials dimension. In section 6 we show that our polynomial does indeed have a large projected shifted partials dimension. Finally, in section 7 we analyze the effect of random restrictions and show that the dimension of shifted partials of f remains large under random restrictions thereby yielding a $2^{\Omega\left(\sqrt{d}\cdot\log N\right)}$ lower bound overall. We wrap up by showing quantitatively weaker lower bounds for efficiently computable polynomials over any field in section 8.

3 Preliminaries

Vector Spaces of Polynomials and linear maps. Let $U, V \subseteq \mathbb{F}[\mathbf{x}]$ be two vector spaces of polynomials and let $\pi : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}[\mathbf{x}]$ be a linear map. Define

$$\pi(U) \stackrel{\text{def}}{=} \{\pi(f) \quad : \quad f \in U\} \subseteq \mathbb{F}[\mathbf{x}].$$

Note that $\pi(U)$ must be a subspace in $\mathbb{F}[\mathbf{x}]$. Also define

$$U + V \stackrel{\text{def}}{=} \mathbb{F}\text{-span}\left(\{f + g \quad : \quad f \in U, g \in V\}\right).$$

Let us record a basic fact from linear algebra as applicable to us.

Proposition 5. *Let $U, V \subseteq \mathbb{F}[\mathbf{x}]$ be two vector spaces of polynomials and let $\pi : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}[\mathbf{x}]$ be any linear map. Then*

$$\pi(U + V) = \pi(U) + \pi(V) \quad \textit{and} \quad \dim(\pi(U)) \leq \dim(U).$$

Numerical estimates.

Proposition 6 (Stirling's Formula, cf. [Rom])**.** $\ln(n!) \quad = \quad n \ln n - n + O(\ln n)$

[13] We learnt of the usefulness of this lemma from the beautiful recent work by Barak, Dvir, Wigderson and Yehudayoff [BDYW11] and a subsequent improvement by Dvir, Saraf and Wigderson [DSW13].

Stirling's formula can be used to obtain the following estimates (proofs of which are in appendix section A).

Lemma 7. *Let $a(n)$, $f(n)$, $g(n) : \mathbb{Z}_{>0} \to \mathbb{Z}$ be integer valued function such that $(|f| + |g|) = o(a)$. Then,*

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

Depth-4 arithmetic formulas. We recall some basic definitions regarding arithmetic circuits and formulas; for a more thorough introduction, see the survey [SY10]. Let Y be a finite set of variables. An arithmetic formula C over $\mathbb{F}$ is a rooted tree the leaves of which are labelled by variables in Y and elements of the field $\mathbb{F}$, and internal nodes (called *gates*) by $+$ and $\times$. This computes a polynomial $f \in \mathbb{F}[Y]$ in a natural way. By the *size* of a formula, we mean the number of vertices in the tree, and by the *depth* of a formula, we mean the longest root-to-leaf path in the tree. Our focus here is on *depth-4 formulas* [14], which are formulas that can be written as sums of products of sums of products, otherwise known as $\Sigma\Pi\Sigma\Pi$ formulas. We will prove lower bounds for *homogeneous* $\Sigma\Pi\Sigma\Pi$ formulas which are $\Sigma\Pi\Sigma\Pi$ formulas such that each node computes a homogeneous polynomial (i.e. a polynomial whose every monomial has the same degree). Given a $\Sigma\Pi\Sigma\Pi$ formula, the layer 0 nodes will refer to the leaf nodes, the layer 1 nodes to the Π-gates just above the leaf nodes, etc. The *top fan-in* refers to the fan-in of the root node on layer 4. We also consider variants of $\Sigma\Pi\Sigma\Pi$ formulas with bounds on the fan-ins of the Π gates. By $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}$ formulas, we mean $\Sigma\Pi\Sigma\Pi$ formulas where the fan-ins of the layer 1 and layer 3 Π gates are *at most* t and D respectively.

4 Upper bounding the measure for low support $\Sigma\Pi\Sigma\Pi$ circuits.

Consider a homogeneous $\Sigma\Pi\Sigma\Pi$ ciruit $\mathcal{C}$ of the form

$$\mathcal{C} = \sum_i \prod_j Q_{ij}, \quad \text{where } |\mathrm{Supp}(Q_{ij})| \leq t \text{ for every } Q_{ij}.$$

We will see how the measure defined in section 2 can be used to pinpoint a weakness of such a circuit. Let us first note two simple properties of our projection map π.

Proposition 8. *Let $Q(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be a homogeneous polynomial of degree d and $m(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be a monomial of degree a. Then*

$$\pi_{d+a,d+a}\left(m(\mathbf{x}) \cdot Q(\mathbf{x})\right) = \begin{cases} 0 & \text{if } |\mathrm{Supp}(m)| < a \\ m(\mathbf{x}) \cdot \sigma_A(\pi_{d,d}(Q)) = m(\mathbf{x}) \cdot \pi_{d,d}(\sigma_A(Q)) & \text{if } A \stackrel{\text{def}}{=} \mathrm{Supp}(m) \text{ has size } a. \end{cases}$$

Our measure, namely

$$\mathsf{DPSP}_{k,\ell,e}(f) \stackrel{\text{def}}{=} \pi_{\ell+e,\ell+e}(\mathbf{x}^{=(\ell,\ell)} \cdot \partial^{=k} f)$$

has the following properties.

[14] we will interchangeably use the terms 'depth-4 circuits', as depth-4 circuits can be converted to depth-4 formulas with only a polynomial blow-up in size

Proposition 9. *For any pair of polynomials $f, g \in \mathbb{F}[\mathbf{x}]$ and any 3-tuple of integers k, ℓ, e*

1. **[Subadditivity.]**
$$\mathsf{DPSP}_{k,\ell,e}(f+g) \leq \mathsf{DPSP}_{k,\ell,e}(f) + \mathsf{DPSP}_{k,\ell,e}(g).$$

2. **[Subprojectivity.]** *If $g = \sigma_A(f)$ for some subset A, i.e. g is obtained from f by setting some subset A of variables to zero, then*
$$\mathsf{DPSP}_{k,\ell,e}(g) \leq \mathsf{DPSP}_{k,\ell,e}(f).$$

3. **[Zeroness for low-support polynomials.]** *If all monomials of f have support strictly less than e then*
$$\mathsf{DPSP}_{k,\ell,e}(f) = 0.$$

The proof is an easy verification. We will now upper bound how large the measure can be for any term T of a low support homogeneous $\Sigma\Pi\Sigma\Pi$-circuit

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

and then via subadditivity derive an upperbound for the entire circuit $\mathcal{C}$ as well. So let us focus on a term T in our t-supported homogeneous $\Sigma\Pi\Sigma\Pi$-circuit $\mathcal{C}$ so that T is of the form

$$T = Q_1 \cdot Q_2 \cdot \ldots \cdot Q_m, \quad |\mathrm{Supp}(Q_i)| \leq t \quad \text{for each } i \in [m],$$

where the Q_i's are homogeneous polynomials and T is of degree d. We will now upper bound $\mathsf{DPSP}_{k,\ell,d-k}(T)$.

Preprocessing. First note that we can assume without loss of generality that every Q_i (except perhaps one) has degree at least $t/2$ for if not, then we can replace two such Q_i's by their product $(Q_i \cdot Q_j)$. The product $(Q_i \cdot Q_j)$ has degree at most t and therefore also support at most t. Continuing this process of combining factors of small degree, we end up in a situation where every factor (except perhaps one) has degree at least $t/2$. In such a situation, the number of factors m can at most be

$$m \leq 1 + \frac{d}{t/2} = 1 + \frac{2d}{t}.$$

Proposition 10. *If $\mathsf{DPSP}_{k,\ell,d-k}(T) > 0$ then for any subset of k factors of T, the sum of their degrees must be at most $(kt + k)$.*

Proof. Assume that

$$\mathsf{DPSP}_{k,\ell,d-k}(T) > 0.$$

Then by part (3) of Proposition 9 it follows that $\mathrm{Supp}(T) \geq (d-k)$. Now consider a subset of the

factors $A \subseteq [m]$ of size k. Since

$$\begin{aligned}
(d-k) &\leq |\mathrm{Supp}(T)| \\
\sum_{i\in[m]} \deg(Q_i) - k &\leq |\mathrm{Supp}(T)| \\
\sum_{i\in[m]} \deg(Q_i) - k &\leq \sum_{i\in[m]} |\mathrm{Supp}(Q_i)| \\
\sum_{i\in[m]} (\deg(Q_i) - |\mathrm{Supp}(Q_i)|) &\leq k \\
\sum_{i\in A} (\deg(Q_i) - |\mathrm{Supp}(Q_i)|) &\leq k \quad \text{(as each summand is positive)} \\
\sum_{i\in A} \deg(Q_i) &\leq \sum_{i\in A} |\mathrm{Supp}(Q_i)| + k \\
\sum_{i\in A} \deg(Q_i) &\leq kt + k.
\end{aligned}$$

□

Computing the derivatives. We now compute the derivatives of our term T and examine what the projected shifted partial derivatives of T look like. Let us introduce the relevant sets and subspaces of polynomials which occur here. For a subset of the factors $A \in \binom{[m]}{k}$ of size k, let

$$d_A \stackrel{\text{def}}{=} \sum_{i\in A} \deg(Q_i)$$

and let

$$V_A \stackrel{\text{def}}{=} \mathbb{F}\text{-span}\left(\mathbf{x}^{(=\ell+d_A-k,\leq \ell+kt)} \cdot \prod_{i\notin A} Q_i\right).$$

Then

Proposition 11.

$$\mathbf{x}^{(=\ell,=\ell)} \cdot \left(\boldsymbol{\partial}^{=k} T\right) \subseteq \sum_{A\in\binom{[m]}{k}} V_A.$$

Combining the above with proposition 5 we have

Corollary 12.

$$\pi_{\ell+d-k,\ell+d-k}\left(\mathbf{x}^{(=\ell,=\ell)} \cdot \left(\boldsymbol{\partial}^{=k} T\right)\right) \subseteq \sum_{A\in\binom{[m]}{k}} \pi_{\ell+d-k,\ell+d-k}\left(V_A\right).$$

In particular,

$$\mathsf{DPSP}_{k,\ell,d-k}(T) = \dim\left(\pi_{\ell+d-k,\ell+d-k}\left(\mathbf{x}^{(=\ell,=\ell)} \cdot \left(\boldsymbol{\partial}^{=k} T\right)\right)\right) \leq \sum_{A\in\binom{[m]}{k}} \dim\left(\pi_{\ell+d-k,\ell+d-k}\left(V_A\right)\right)$$

Now fix an $A \in \binom{[m]}{k}$ and consider the vector space V_A defined above. The generators of V_A consist of polynomials of the form

$$g(\mathbf{x}) = m(\mathbf{x}) \cdot \left(\prod_{i \notin A} Q_i \right),$$

where $m(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ is a monomial of degree $\ell + d_A - k$ and $\left(\prod_{i \notin A} Q_i\right)$ is of degree $(d - d_A)$. By proposition 8 we have that if $m(\mathbf{x})$ is not multilinear then

$$\pi_{\ell+d-k,\ell+d-k}(g) = 0.$$

So assume that $m(\mathbf{x})$ is a multilinear monomial,

$$m(\mathbf{x}) = \mathbf{x}_S, \quad S \in \binom{[N]}{\ell + d_A - k}.$$

By proposition 8

$$\pi_{\ell+d-k,\ell+d-k}(g) = \mathbf{x}_S \cdot \pi_{d-d_A,d-d_A}\left(\sigma_S(\prod_{i \notin A} Q_i)\right).$$

Thus

$$\pi_{\ell+d-k,\ell+d-k}\left(V_A\right) \subseteq \mathbb{F}\text{-span}\left(\left\{\mathbf{x}_S \cdot \pi_{d-d_A,d-d_A}\left(\sigma_S(\prod_{i \notin A} Q_i)\right) : S \in \binom{[N]}{\ell + d_A - k}\right\}\right)$$

In particular,

$$\begin{aligned} \dim\left(\pi_{\ell+d-k,\ell+d-k}\left(V_A\right)\right) &\leq \binom{N}{\ell + d_A - k} \\ &\leq \binom{N}{\ell + kt} \quad \left(\text{for } \ell + kt < \frac{N}{2}, \text{using Proposition 10}\right). \end{aligned}$$

Combining this with the above observations we have

Lemma 13. *Let T be a term of the form*

$$T = Q_1 \cdot Q_2 \cdot \ldots \cdot Q_m, \quad |\text{Supp}(Q_i)| \leq t \quad \text{for each } i \in [m],$$

where the Q_i's are homogeneous polynomials and T is of degree d. For any k and any $\ell < \frac{N}{2} - kt$ we have

$$\mathsf{DPSP}_{k,\ell,d-k}(T) \leq \binom{2d/t + 1}{k} \cdot \binom{N}{\ell + k \cdot t}.$$

Combining the above upper bound for a term with the subadditivity of our measure we immediately get:

Corollary 14. *Let $\mathcal{C}$ be a t-supported degree d homogeneous $\Sigma\Pi\Sigma\Pi$ circuit with top fanin s, i.e $\mathcal{C}$ is a degree d homogeneous circuit of the form*

$$\mathcal{C} = \sum_{i=1}^{s} Q_{i1} \cdot Q_{i2} \cdot \ldots \cdot Q_{im_i}, \quad |\text{Supp}(Q_{ij})| \leq t.$$

Then for every k and every $\ell < \frac{N}{2} - kt$ we have

$$\mathsf{DPSP}_{k,\ell,d-k}(\mathcal{C}) \leq s \cdot \binom{2d/t+1}{k} \cdot \binom{N}{\ell + k \cdot t}.$$

Consequently, for any N-variate homogeneous polynomial f of degree d, any homogeneous t-supported $\Sigma\Pi\Sigma\Pi$-circuit C computing f must have top fanin at least

$$s \geq \frac{\mathsf{DPSP}_{k,\ell,d-k}(f)}{\binom{2d/t+1}{k} \cdot \binom{N}{\ell+k\cdot t}}.$$

In the next section we construct an explicit polynomial f for which $\mathsf{DPSP}_{k,\ell,d-k}(f)$ is large and then use the above to deduce a lower bound on the top fanin of any t-supported $\Sigma\Pi\Sigma\Pi$-circuit computing f.

5 The lower bound for low support homogeneous $\Sigma\Pi\Sigma\Pi$ circuits.

We will now construct an explicit homogeneous, multilinear polynomial f of degree d on $N = d^3$ variables for which our measure, namely $\mathsf{DPSP}_{k,\ell,d-k}(f)$ is large. We will then see that this implies that any t-supported $\Sigma\Pi\Sigma\Pi$-circuit computing f must have large top fanin.

5.1 The Construction of an Explicit Polynomial

Our explicit polynomial is parametrized by an integer parameter r that we call NW_r and it is a variant of the Nisan-Wigderson design polynomial from [KSS13]. Let d be a prime power and $\mathbb{F}_d$ be the finite field of size d. Let $\mathbb{F}_{d^2} \supseteq \mathbb{F}_d$ be the quadratic extension field of $\mathbb{F}_d$. We refer to the elements of the finite field $\mathbb{F}_{d^2}$ simply as $\{1, 2, \cdots, d^2\}$ where the first d among these belong to the subfield $\mathbb{F}_d$. Fix an integer r. Our explicit polynomial is:

$$\mathsf{NW}_r(x_{1,1}, x_{1,2}, \ldots, x_{d,d^2}) \stackrel{\text{def}}{=} \sum_{h(z) \in \mathbb{F}_{d^2}[z], \deg(h) \leq r} \prod_{i \in [d]} x_{i,h(i)}.$$

From the definition above, it is clear that for all r, NW_r is an explicit homogeneous, multilinear polynomial of degree d on $N = d^3$ variables. our main technical lemma stated below is a lower bound on the dimension of projected shifted partials of the design polynomial NW_r.

Lemma 15. [Main Technical Lemma.] *Let NW_r be the Nisan-Wigderson design-based polynomial defined above. Over any field $\mathbb{F}$ of characteristic zero, for $r = \frac{d}{3}$ and $k = o(d)$ and $\ell = \frac{N}{2} \cdot \left(1 - \frac{k \ln d}{d}\right)$ we have*

$$\mathsf{DPSP}_{k,\ell,d-k}(\mathsf{NW}_r) \geq \frac{1}{d^{O(1)}} \cdot \min\left(\binom{N}{\ell + d - k}, \binom{d}{k}^2 \cdot d^k \cdot k! \cdot \binom{N}{\ell} \right).$$

We first see how to apply this lemma to deduce a lower bound on the top fanin of any t-supported homogeneous $\Sigma\Pi\Sigma\Pi$ circuit computing $\mathsf{NW}_{d/3}$ while postponing the proof of this lemma to section 6. So consider a t-supported $\Sigma\Pi\Sigma\Pi$ circuit $\mathcal{C}$ of top fanin s computing $\mathsf{NW}_{d/3}$. We fix our choice of parameters as follows:

$$k = \delta \cdot \frac{d}{t} \quad \text{(for a small enough constant } \delta > 0), \quad \ell = \frac{N}{2} \cdot \left(1 - \frac{k \ln d}{d}\right) \tag{3}$$

By corollary 14 we get

$$\begin{aligned}
s &\geq \frac{\mathsf{DPSP}_{k,\ell,d-k}(\mathsf{NW}_{d/3})}{\binom{2d/t+1}{k}\cdot\binom{N}{\ell+k\cdot t}} \\
&\geq \frac{1}{d^{O(1)}\cdot\binom{2d/t+1}{k}}\cdot\min\left(\frac{\binom{d}{k}^2\cdot d^k\cdot k!\cdot\binom{N}{\ell}}{\binom{N}{\ell+kt}},\frac{\binom{N}{\ell+d-k}}{\binom{N}{\ell+kt}}\right) \quad \text{(using lemma 15)}\,. \\
&= \frac{1}{2^{O(d/t)}}\cdot\min\left(\binom{d}{k}^2\cdot d^k\cdot k!\cdot\frac{(\ell+kt)!}{\ell!}\cdot\frac{(N-\ell-kt)!}{(N-\ell)!},\frac{(\ell+kt)!}{(\ell+d-k)!}\cdot\frac{(N-\ell-kt)!}{(N-\ell-d+k)!}\right) \\
&= \frac{1}{2^{O(d/t)}}\cdot\min\left(\binom{d}{k}^2\cdot d^k\cdot k!\cdot e^{(-kt)\cdot\ln\frac{N-\ell}{\ell}+o(1)},e^{(d-k-kt)\cdot\ln\frac{N-\ell}{\ell}+o(1)}\right) \quad \text{(Using lemma 7)} \\
&= \frac{1}{2^{O(d/t)}}\cdot\min\left(\binom{d}{k}^2\cdot d^k\cdot k!\cdot e^{(-kt)\cdot\ln\frac{1+(k/d)\ln d}{1-(k/d)\ln d}},e^{(d-k-kt)\cdot\ln\frac{1+(k/d)\ln d}{1-(k/d)\ln d}}\right) \\
&\geq 2^{\Omega\left(\frac{d}{t}\cdot\log N\right)} \quad \left(\text{for a small enough constant } \delta \text{ and } t=\Omega(\log^2 d)\right)
\end{aligned}$$

This gives the claimed lower bound on the top fanin s of any t-supported homogeneous $\Sigma\Pi\Sigma\Pi$-circuit computing $\mathsf{NW}_{d/3}$.

6 Proof of the main technical lemma

In this section we prove lemma 15, i.e. we show that the dimension of projected shifted partial derivatives of the Nisan-Wigderson design based polynomial is within a $poly(N)$ factor of the maximum possible. Let $e \stackrel{\text{def}}{=} (d-k)$ throughout the rest of this section.

Preliminaries. Note that in the construction in section 5 of NW_r, there is a 1-1 correspondence between the variable indices in $[N]$ and points in $\mathbb{F}_d \times \mathbb{F}_{d^2}$, which we will often identify simply with $[d]\times[d^2]$. Being homogeneous and multilinear of degree d, the monomials of NW_r are in 1-1 correspondence with sets in $\binom{[N]}{d} \equiv \binom{[d]\times[d^2]}{d}$. Indeed, from the construction it is clear that the coefficient of any monomial in NW_r is either 0 or 1 and that there is a 1-1 correspondence between monomials in the support of NW_r and univariate polynomials of degree at most r in $\mathbb{F}_{d^2}[z]$. Now since two distinct polynomials of degree r over a field have at most r common roots we get:

Proposition 16. [A basic property of our construction.] *For any two distinct sets $D_1, D_2 \in \binom{[d]\times[d^2]}{d}$ in the support of NW_r, we have*

$$\begin{aligned}
|D_1 \cap D_2| &\leq r \\
&< \frac{e}{2} \quad (\textit{for } r = d/3 \textit{ and } k = o(d).)
\end{aligned}$$

Our goal for the remainder of this section is to lower bound $\mathsf{DPSP}_{k,\ell,d-k}(NW_r)$ which is defined as the $\mathbb{F}$-linear dimension of the following set of polynomials.

$$\mathsf{DPSP}_{k,\ell,d-k}(\mathsf{NW}_r) = \dim\left(\pi_{\ell+d-k,\ell+d-k}\left(\mathbf{x}^{(=\ell,=\ell)}\cdot\boldsymbol{\partial}^{=k}\mathsf{NW}_r\right)\right).$$

Reformulating our goal in terms of the rank of an explicit matrix. Let f be any homogeneous multilinear polynomial of degree d on N variables. By multilinearity, the only derivatives of f that survive are those with respect to multilinear monomials. Thus we have

$$\boldsymbol{\partial}^{=k} f = \left\{ \partial^C f \ : \ C \in \binom{[N]}{k} \right\}.$$

Note that every k-th order derivative of f is homogeneous and multilinear of degree $(d-k)$. Combining this with proposition 8 we get that

$$\pi_{\ell+d-k,\ell+d-k}(\mathbf{x}^{(=\ell,=\ell)} \cdot \boldsymbol{\partial}^{=k} f) = \left\{ \mathbf{x}_A \cdot \sigma_A \left(\partial^C f\right) \ : \ A \in \binom{[N]}{\ell}, \quad C \in \binom{[N]}{k} \right\}.$$

Thus we have

Proposition 17. *For any homogeneous multilinear polynomial f of degree d on N variables and for all integers k and ℓ:*

$$\mathsf{DPSP}_{k,\ell,d-k}(f) = \dim\left(\left\{ \mathbf{x}_A \cdot \sigma_A \left(\partial^C f\right) \ : \ A \in \binom{[N]}{\ell}, \quad C \in \binom{[N]}{k} \right\} \right).$$

Now the $\mathbb{F}$-linear dimension of any set of polynomials is the same as the rank of the matrix corresponding to our set of polynomials in the natural way. Specifically,

Proposition 18. *Let f be a homogeneous multilinear polynomial of degree d on N variables. Let k, ℓ be integers. Define a matrix $M(f)$ as follows. The rows of $M(f)$ are labelled by pairs of subsets $(A, C) \in \binom{[N]}{\ell} \times \binom{[N]}{k}$ and columns are indexed by subsets $S \in \binom{[N]}{\ell+e}$. Each row (A, C) corresponds to the polynomial*

$$f_{A,C} \stackrel{\text{def}}{=} \mathbf{x}_A \cdot \sigma_A \left(\partial^C f\right)$$

in the following way. The S-th entry of the row (A, C) is the coefficient of $\mathbf{x}_S$ in the polynomial $f_{A,C}$. Then,

$$\mathsf{DPSP}_{k,\ell,d-k}(f) = \mathit{rank}(M(f)).$$

So our problem is equivalent to lower bounding the rank of the matrix $M(f)$ for our constructed polynomial f. Now note that the entries of $M(f)$ are coefficients of appropriate monomials of f and it will be helpful to us in what follows to keep track of this information. We will do it by assigning a label to each cell of $M(f)$ as follows. We will think of every location in the matrix $M(f)$ being labelled with either a set $D \in \binom{[N]}{d}$ or the label `InvalidSet` depending on whether that entry contains the coefficient of the monomial $\mathbf{x}_D$ of f or it would have been zero regardless of the actual coefficients of f. Specifically, let us introduce the following notation. For sets A, B define:

1. $$A \setminus\setminus B = \begin{cases} A \setminus B & \text{if} B \subseteq A \\ \texttt{InvalidSet} & \text{otherwise} \end{cases}$$

2. $$A \uplus B = \begin{cases} A \cup B & \text{if} B \cap A = \emptyset \\ \texttt{InvalidSet} & \text{otherwise} \end{cases}$$

Then the label of the $((A,C),S)$-th cell of $M(f)$ is defined to be the set $(S \setminus\setminus A) \uplus C$. Equivalently, if the label of a cell of the (A,C)-th row of M is a set D then the column must be the one corresponding to $S = (D \setminus\setminus C) \uplus A$ (if C is not a subset of D or if $(D \setminus\setminus C)$ and A are not disjoint then D cannot occur in the row indexed by (A,C)). For the rest of this section, we will refer to $M(\mathsf{NW}_r)$ simply as the matrix M. Our goal then is to show that the rank of this matrix M is reasonably close (within a $\mathsf{poly}(d)$-factor) of the trivial upper bound, viz. the minimum of the number of rows and the number of columns of M. It turns out that our matrix M is a relatively sparse matrix and we will exploit this fact by using a relevant lemma from real matrix analysis to obtain a lower bound on its rank.

The Surrogate Rank. Consider the matrix $B \stackrel{\text{def}}{=} M^T \cdot M$. Then B is a real symmetric, positive semidefinite matrix. From the definition of B it is easy to show that:

Proposition 19. *Over any field $\mathbb{F}$ we have*

$$\mathit{rank}(B) \leq \mathit{rank}(M).$$

Over the field $\mathbb{R}$ of real numbers we have

$$\mathit{rank}(B) = \mathit{rank}(M).$$

So it suffices to lower bound the rank of B. By an application of Cauchy-Schwarz on the vector of nonzero eigenvalues of B, one obtains:

Lemma 20. *[Alo09] Over the field of real numbers $\mathbb{R}$ we have:*

$$\mathit{rank}(B) \geq \frac{\mathrm{Tr}(B)^2}{\mathrm{Tr}(B^2)}.$$

Let us call the quantity $\frac{\mathrm{Tr}(B)^2}{\mathrm{Tr}(B^2)}$ as the surrogate rank of M, denoted $\mathrm{SurRank}(M)$. It then suffices to show that this quantity is within a $\mathsf{poly}(d)$ factor of $U = \min(\binom{d^3}{\ell+e}, \binom{d^3}{\ell} \cdot \binom{d^3}{k})$. In the rest of this section, we will first derive an exact expression for $\mathrm{SurRank}(B)$ and then show that it is close to U.

6.1 Deriving an exact expression for $\mathrm{SurRank}(B)$.

We will now calculate an exact expression for $\mathrm{SurRank}(B)$, or equivalently an exact expression for $\mathrm{Tr}(B)$ and $\mathrm{Tr}(B^2)$.

Calculating $\mathrm{Tr}(B)$. Calculating $\mathrm{Tr}(B)$ is fairly straightforward. From the definition of the matrix B we have:

Proposition 21. *For any $0, \pm 1$ matrix M (i.e. a matrix all of whose entries are either 0, or $+1$ or -1) we have*

$$\mathrm{Tr}(B) = \mathrm{Tr}(M^T \cdot M) = \textit{number of nonzero entries in } M.$$

Now we can calculate the number of nonzero entries in M by going over all sets $D \in \binom{[N]}{d} \cap \mathrm{Supp}(\mathsf{NW}_r)$, calculating the number of cells of M labelled with D and adding these up. This yields:

Proposition 22.

$$\mathrm{Tr}(B) = d^{2r+2} \cdot \binom{d}{k} \cdot \binom{N-e}{\ell}.$$

Calculating $\mathrm{Tr}(B^2)$**.** From the definition of $B = M^T \cdot M$ and expanding out the relevant summations we get:

Proposition 23.

$$\mathrm{Tr}(B^2) = \sum_{(A_1,C_1),(A_2,C_2) \in \left(\binom{[N]}{\ell} \times \binom{[N]}{k}\right)^2} \quad \sum_{S_1,S_2 \in \binom{[N]}{\ell+e}^2} M_{(A_1,C_1),S_1} \cdot M_{(A_1,C_1),S_2} \cdot M_{(A_2,C_2),S_1} \cdot M_{(A_2,C_2),S_2}.$$

We will use the following notation in doing this calculation. For a pair of row indices $((A_1, C_1), (A_2, C_2)) \in \left(\binom{[N]}{\ell} \times \binom{[N]}{k}\right)^2$ and a pair of column indices $S_1, S_2 \in \left(\binom{[N]}{\ell+e}\right)^2$, the box $\mathbf{b}$ defined by them, denoted $\mathbf{b} = 2 - \mathrm{box}((A_1, C_1), (A_2, C_2), S_1, S_2)$ is the four-tuple of cells

$$(((A_1, C_1), S_1), ((A_1, C_1), S_2), ((A_2, C_2), S_1), ((A_2, C_2), S_2)).$$

Since all the entries of our matrix M are either 0 or 1 we have:

Proposition 24.

$$\mathrm{Tr}(B^2) = \textit{Number of boxes } \mathbf{b} \textit{ with all four entries nonzero.}$$

For a box $\mathbf{b} = 2-\mathrm{box}((A_1, C_1), (A_2, C_2), S_1, S_2)$, its tuple of labels, denoted $\mathrm{labels}(\mathbf{b})$ is the tuple of labels of the cells $((A_1, C_1), S_1), ((A_1, C_1), S_2), ((A_2, C_2), S_1), ((A_2, C_2), S_2))$ in that order. In other words,

$$\mathrm{labels}(\mathbf{b}) = ((S_1 \setminus\setminus A_1) \uplus C_1, (S_2 \setminus\setminus A_1) \uplus C_1, (S_1 \setminus\setminus A_2) \uplus C_2, (S_2 \setminus\setminus A_2) \uplus C_2).$$

We then have

Proposition 25. $\mathrm{Tr}(B^2)$ *equals the number of boxes*

$$\mathbf{b} = 2 - \textit{box}((A_1, C_1), (A_2, C_2), S_1, S_2)$$

such that all the four labels in $\mathrm{labels}(\mathbf{b})$ *are valid sets in the support of our design polynomial* NW_r*.*

So our problem boils down to counting the number of boxes in which all the four labels are valid sets in the support of our polynomial NW_r. Our key observation is that the sets labelling such boxes must satisfy certain constraints on pairwise intersection sizes and this will help rule out boxes with more than two distinct labels.

Proposition 26. *Suppose that all the labels of a box*

$$\mathbf{b} = 2 - \textit{box}((A_1, C_1), (A_2, C_2), S_1, S_2)$$

are valid sets:

$$\mathrm{labels}(\mathbf{b}) = (D_{11}, D_{12}, D_{21}, D_{22}) \in \binom{[N]}{d}^4.$$

Then we must have that either

1. $|D_{11} \cap D_{12}| \geq \frac{e}{2}$ **and** $|D_{21} \cap D_{22}| \geq \frac{e}{2}$ *or,*

2. $|D_{11} \cap D_{21}| \geq \frac{e}{2}$ **and** $|D_{12} \cap D_{22}| \geq \frac{e}{2}$.

Proof. First observe that

$$D_{11} \cap D_{12} \supseteq (A_2 \setminus A_1) \quad \text{and} \quad D_{21} \cap D_{22} \supseteq (A_1 \setminus A_2). \tag{4}$$

Next observe that

$$D_{11} \cap D_{21} \supseteq S_1 \setminus (A_1 \cup A_2) \quad \text{and} \quad D_{12} \cap D_{22} \supseteq S_2 \setminus (A_1 \cup A_2). \tag{5}$$

Now let $|A_1 \cap A_2| = v$.

Case 1. $v \leq (\ell - \frac{e}{2})$. Then the containment (4) implies that

$$|D_{11} \cap D_{12}| \geq (\ell - v) \geq \frac{e}{2} \quad \text{and} \quad |D_{21} \cap D_{22}| \geq (\ell - v) \geq \frac{e}{2}.$$

Case 2. $v \geq (\ell - \frac{e}{2})$. Then the containment (5) implies that

$$|D_{11} \cap D_{21}| \geq (\ell + e) - (\ell + \ell - v) \geq \frac{e}{2} \quad \text{and} \quad |D_{12} \cap D_{22}| \geq (\ell + e) - (\ell + \ell - v) \geq \frac{e}{2}.$$

□

Indeed, the above observation is why we choose our polynomial f to be a design polynomial with $r < \frac{e}{2}$ since the design polynomial property ensures that any two distinct sets D_1 and D_2 in the support of NW_r have intersection size at most $r < \frac{e}{2}$. This means that any box $\mathbf{b}$ that contributes to $\mathrm{Tr}(B^2)$ must have the property that its label set $\mathrm{labels}(\mathbf{b})$ contains at most two distinct sets in the support of NW_r.

Corollary 27. *For any two distinct sets $D_1, D_2 \in \binom{[N]}{d}$ define*

$$\begin{aligned} \mu_0(D_1) &\stackrel{\text{def}}{=} \{\textit{box } \mathbf{b} : \mathrm{labels}(\mathbf{b}) = (D_1, D_1, D_1, D_1)\} \\ \mu_1(D_1, D_2) &\stackrel{\text{def}}{=} \{\textit{box } \mathbf{b} : \mathrm{labels}(\mathbf{b}) = (D_1, D_2, D_1, D_2)\} \\ \mu_2(D_1, D_2) &\stackrel{\text{def}}{=} \{\textit{box } \mathbf{b} : \mathrm{labels}(\mathbf{b}) = (D_1, D_1, D_2, D_2)\} \end{aligned}$$

Let the support of NW_r, denoted $\mathrm{Supp}(\mathsf{NW}_r) \subset \binom{[N]}{d}$, be the set of all sets $D \in \binom{[N]}{d}$ such that the coefficient of the monomial $\mathbf{x}_D$ in NW_r is nonzero. Then

$$\mathrm{Tr}(B^2) = \sum_{D_1 \in \mathrm{Supp}(\mathsf{NW}_r)} |\mu_0(D_1)| \quad + \sum_{D_1 \neq D_2 \in \mathrm{Supp}(\mathsf{NW}_r)} |\mu_1(D_1, D_2)| + \sum_{D_1 \neq D_2 \in \mathrm{Supp}(\mathsf{NW}_r)} |\mu_2(D_1, D_2)| \,.$$

By the way, proposition 26 rules out the existence of any box $\mathbf{b}$ having $\mathrm{labels}(\mathbf{b}) = (D_1, D_2, D_2, D_1)$ with $D_1, D_2 \in \mathrm{Supp}(\mathsf{NW}_r)$ and that is why there is no term in $\mathrm{Tr}(B^2)$ corresponding to such boxes. In what follows we will compute $\mathrm{Tr}(B^2)$ by deriving expressions for $|\mu_0(D_1)|, |\mu_1(D_1, D_2)|$ and $|\mu_2(D_1, D_2)|$ and then summing these up over $D_1, D_2 \in \mathrm{Supp}(\mathsf{NW}_r)$. We first observe:

Proposition 28. *For any set $D_1 \in \binom{[N]}{d}$ and any row (A, C) of M, there can be at most one cell in that row labelled with the set D_1.*

This means that any box $\mathbf{b} = 2 - \mathrm{box}((A_1, C_1), (A_2, C_2), S_1, S_2)$ contributing to either $\mu_0(D_1)$ or $\mu_2(D_1, D_2)$, the columns S_1 and S_2 must be the same.

6.2 Calculating $\mu_0(D_1)$.

Every box $\mathbf{b} \in \mu_0(D_1)$ is of the form $\mathbf{b} = 2-\mathrm{box}((A_1, C_1), (A_2, C_2), S_1, S_1)$. Let $u = |C_1 \cap C_2|$. Due to the type of the box we know that $(D_1 \setminus\setminus C_1) \uplus A_1 = (D_1 \setminus\setminus C_2) \uplus A_2$. This implies the following two things:

1. $C_1 \setminus (C_1 \cap C_2) \subseteq A_1$ and $C_2 \setminus (C_1 \cap C_2) \subseteq A_2$.
2. $A_1 \setminus (C_1 \setminus (C_1 \cap C_2)) = A_2 \setminus (C_2 \setminus (C_1 \cap C_2))$.

Due to the fact that $|C_1 \setminus (C_1 \cap C_2)| = |C_2 \setminus (C_1 \cap C_2)| = k - u$ and by 1 above, $k - u$ elements in A_1 and A_2 are fixed. Due to 2 above, A_1 and A_2 must agree on the rest of the elements which can be chosen from $([N] \setminus D_1) \cup (C_1 \cap C_2)$. Analyzing this situation gives

Proposition 29.

$$|\mu_0(D_1)| = \sum_{0 \le u \le k} \binom{N-d+u}{\ell-k+u} \cdot \binom{d}{u, k-u, k-u, d-2k+u}$$

6.3 Calculating $\mu_1(D_1, D_2)$.

Let $D_1, D_2 \in \binom{[N]}{d}$ be two distinct subsets in the support of NW_r. We consider a box $\mathbf{b} = 2-\mathrm{box}((A_1, C_1), (A_2, C_2), S_1, S_2)$ in $\mu_1(D_1, D_2)$ which is equivalent to saying that

$$(D_1 \setminus\setminus C_1) \uplus A_1 = (D_1 \setminus\setminus C_2) \uplus A_2 = S_1$$

and

$$(D_2 \setminus\setminus C_1) \uplus A_1 = (D_2 \setminus\setminus C_2) \uplus A_2 = S_2.$$

Note that here $(C_1 \cup C_2) \subseteq D_1 \cap D_2$. Analyzing this situation as in Section 6.2 gives

Proposition 30. *If $|D_1 \cap D_2| = w$ then*

$$|\mu_1(D_1, D_2)| = \sum_{0 \le u \le k} \binom{N-2d+w+u}{\ell-k+u} \cdot \binom{w}{u, k-u, k-u, w-2k+u}$$

6.4 Calculating $\mu_2(D_1, D_2)$.

Let $D_1, D_2 \in \binom{[N]}{d}$ be two distinct subsets in the support of NW_r. We consider a box $\mathbf{b} = 2-\mathrm{box}((A_1, C_1), (A_2, C_2), S_1, S_2)$ in $\mu_1(D_1, D_2)$. As we observed before this can happen only if $S_1 = S_2 = S$ (say). By definition we have

$$(D_1 \setminus\setminus C_1) \uplus A_1 = (D_2 \setminus\setminus C_2) \uplus A_2 = S$$

Here, let $C_1' = C_1 \cap (D_1 \cap D_2)$, let $C_2' = C_2 \cap (D_1 \cap D_2)$, and let $C = C_1' \cap C_2'$. Also let $u_1 = |C_1' \setminus C|$, $u_2 = |C_2' \setminus C|$, and $u = |C|$. Analyzing this situation as in Section 6.2 gives

Proposition 31. *If $|D_1 \cap D_2| = w$ then*

$$\begin{aligned} |\mu_2(D_1, D_2)| \;=\; & \sum_{0 \le u, u_1, u_2 \le k} \binom{N-2d+w+2k-u-u_1-u_2}{\ell-d+k+w-u-u_1-u_2} \\ & \cdot \binom{d-w}{k-u-u_1} \cdot \binom{d-w}{k-u-u_2} \cdot \binom{w}{u_1, u_2, u, w-u-u_1-u_2}. \end{aligned}$$

6.5 An exact expression for $\mathrm{SurRank}(B)$.

We are now ready to give an expression for $\mathrm{Tr}(B^2)$ and thereby for $\mathrm{SurRank}(B)$ as well. Let $R_d(w,r)$ denote the number of univariate polynomials in $\mathbb{F}_{d^2}[z]$ of degree at most r having exactly w distinct roots in the subfield $\mathbb{F}_d$. Then using the expression for $|\mu_0(D_1)|, |\mu_1(D_1,D_2)|$ and $|\mu_2(D_1,D_2)|$ calculated above we get

$$\mathrm{Tr}(B^2) = T_0 + T_1 + T_2,$$

where

$$\begin{aligned}
T_0 &= d^{2r+2} \cdot \sum_{0\le u\le k} \binom{N-d+u}{\ell-k+u} \cdot \binom{d}{u, k-u, k-u, d-2k+u} \\
T_1 &= d^{2r+2} \cdot \sum_{k\le w\le r} \sum_{0\le u\le k} R_d(w,r) \cdot \binom{N-2d+w+u}{\ell-k+u} \cdot \binom{w}{u, k-u, k-u, w-2k+u} \\
T_2 &= d^{2r+2} \cdot \sum_{0\le w\le r} \sum_{0\le u,u_1,u_2\le k} R_d(w,r) \cdot \binom{N-2d+w+2k-u-u_1-u_2}{\ell-d+k+w-u-u_1-u_2} \\
&\quad \cdot \binom{d-w}{k-u-u_1} \cdot \binom{d-w}{k-u-u_2} \cdot \binom{w}{u_1, u_2, u, w-u-u_1-u_2}.
\end{aligned}$$

6.6 Estimating the above expression.

First note that any polynomial $h(z) \in \mathbb{F}_{d^2}[z]$ of degree at most r that has w roots in $\mathbb{F}_d[z]$ must be of the form

$$h(z) = (z-\alpha_1)\cdot(z-\alpha_2)\cdot\ldots\cdot(z-\alpha_w)\cdot\hat{h}(z),$$

where each α_i is in $\mathbb{F}_d$ and $\hat{h}(z) \in \mathbb{F}_{d^2}[z]$ is of degree at most $(r-w)$. Thus we have

$$\begin{aligned}
R_d(w,r) &\le d^{2r-2w+2} \cdot \binom{d}{w} \\
&\le \frac{d^{2r+2}}{d^w \cdot w!}
\end{aligned}$$

Estimating T_0. We have

$$\begin{aligned}
\binom{d}{u, k-u, k-u, d-2k+u} &= \frac{(d)\cdot(d-1)\cdot\ldots(d-2k+u+1)}{u!\cdot(k-u)!\cdot(k-u)!} \\
&\le (d)\cdot(d-1)\cdot\ldots(d-2k+u+1) \\
&\le d^{2k-u}
\end{aligned}$$

and since $(\ell-k) \le (N-d)$ we have that for all $0 \le u \le k$:

$$\binom{N-d+u}{\ell-k+u} \le \binom{N-d+k}{\ell}.$$

So for $d \ge 2$

$$T_0 \le 2\cdot d^{2r+2+2k} \cdot \binom{N-d+k}{\ell}$$

We will now upper bound each of the sums T_1 and T_2.

Estimating T_1. Let

$$S(u,w) \stackrel{\text{def}}{=} \frac{1}{d^w \cdot w!} \cdot \binom{N-2d+w+u}{\ell-k+u} \cdot \binom{w}{u, k-u, k-u, w-2k+u}.$$

It turns out that $S(u,w)$ is maximized at $w=u=k$ (see section B.1 in appendix) and consequently we have:

Claim 32. *For any $d>4$ and $k<\frac{d}{4}$ and $\ell<\frac{N}{2}$:*

$$T_1 \leq (rk) \cdot \frac{d^{4r+4}}{d^k \cdot k!} \cdot \binom{N-2d+2k}{\ell}.$$

Estimating T_2. Let

$$\begin{aligned} S_1(w,u,u_1,u_2) \quad = \quad & \frac{1}{d^w \cdot w!} \cdot \binom{N-2d+w+2k-u-u_1-u_2}{\ell-d+k+w-u-u_1-u_2} \cdot \binom{d-w}{k-u-u_1} \cdot \\ & \cdot \binom{d-w}{k-u-u_2} \cdot \binom{w}{u_1,u_2,u,w-u-u_1-u_2}. \end{aligned}$$

It turns out that $S_1(w,u,u_1,u_2)$ is maximized at $w=u=u_1=u_2=0$ (see section B.2 in appendix). Consequently we have:

Claim 33. *For $\ell > \frac{N}{d}+2d$ and $k<\frac{d}{3}$ we have*

$$T_2 \leq (rk^3) \cdot d^{4r+4} \cdot \binom{N-2d+2k}{\ell-d+k} \cdot \binom{d}{k}^2.$$

Combining the above bounds, for the choice of parameters

$$r=\frac{d}{3}, \quad \text{and} \quad k=o(d) \quad \textit{and} \quad \ell=\frac{N}{2}\cdot\left(1-\frac{k\ln d}{d}\right), \tag{6}$$

we have:

$$\mathrm{Tr}(B^2) \quad \leq \quad 2\cdot d^{2r+2+2k}\cdot\binom{N-e}{\ell} + (rk)\cdot\frac{d^{4r+4}}{d^k\cdot k!}\cdot\binom{N-2e}{\ell} + (rk^3)\cdot d^{4r+4}\cdot\binom{N-2e}{\ell-e}\cdot\binom{d}{k}^2$$

Now observe that for the above choice of parameters r,k and ℓ we have

$$2\cdot d^{2r+2+2k}\cdot\binom{N-e}{\ell} \quad \leq \quad (rk)\cdot\frac{d^{4r+4}}{d^k\cdot k!}\cdot\binom{N-2e}{\ell}$$

so that overall

$$\mathrm{Tr}(B^2) \quad \leq \quad (2k^3d)\cdot(d^{4r+4})\cdot\max\left(\frac{1}{d^k\cdot k!}\cdot\binom{N-2e}{\ell}, \binom{N-2e}{\ell-e}\cdot\binom{d}{k}^2\right)$$

This means that

$$\begin{aligned}\mathrm{SurRank}(B) &= \frac{\mathrm{Tr}(B)^2}{\mathrm{Tr}(B^2)} \\ &\geq \frac{1}{2k^3 d} \cdot \min\left(\frac{\binom{d}{k}^2 \cdot d^k \cdot k! \cdot \binom{N-e}{\ell}^2}{\binom{N-2e}{\ell}}, \frac{\binom{N-e}{\ell}^2}{\binom{N-2e}{\ell-e}} \right) \\ &= \frac{1}{d^{O(1)}} \cdot \min\left(\binom{d}{k}^2 \cdot d^k \cdot k! \cdot \binom{N}{\ell}, \binom{N}{\ell+e} \right) \quad \text{(using (6))}\,.\end{aligned}$$

This proves our main technical lemma, namely lemma 15.

7 The lower bound for general homogeneous ΣΠΣΠ circuits.

As hinted in the introduction, the problem of lower bounding the size of general homogeneous ΣΠΣΠ circuits reduces to proving lower bounds for low support homogeneous ΣΠΣΠ circuits. We now give the details of this reduction.

Definition 1. *For a real number $p \in (0,1]$, define the distribution $\mathcal{D}_p$ on subsets of $[N]$ obtained by choosing every element in $[N]$ independently at random with probability $(1-p)$. Thus, $\mathcal{D}_p : 2^{[N]} \mapsto (0,1]$ and for any $R \subseteq [N]$ we have*

$$\mathcal{D}_p(R) = (1-p)^{|R|} \cdot p^{N-|R|}.$$

Let NW_r be the Nisan-Wigderson design polynomial as constructed in section 5. Let us consider a homogeneous ΣΠΣΠ-circuit $\mathcal{C}$ computing it, i.e. consider any representation of NW_r of the form

$$\mathsf{NW}_r = \sum_i \prod_j Q_{ij}, \tag{7}$$

where the Q_{ij}'s are also homogeneous polynomials. Suppose that the total number of monomials in the polynomials Q_{ij}'s is bounded by $\mathfrak{s}$. Then the following holds true:

Lemma 34. *For any homomorphism $\sigma_R : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}[\mathbf{x}]$ we have*

$$\sigma_R(\mathsf{NW}_r) = \sum_i \prod_j \sigma_R(Q_{ij}).$$

For a set R chosen randomly according to $\mathcal{D}_p$, we have:

$$\Pr_{R \sim \mathcal{D}_p}[\exists i,j : \quad \sigma_R(Q_{ij}) \text{ contains a monomial of support more than } t] \leq \mathfrak{s} \cdot p^t.$$

Proof. The distribution $\mathcal{D}_p$ "kills" a variable x with probability $(1-p)$, i.e. $\sigma_R(x) = 0$ with probability $(1-p)$. Now a monomial m of support size more than t contains at least $t+1$ distinct variables. Each of these variables "survives" with probability p so that overall the monomial m survives with probability at most p^{t+1}, i.e.

$$\Pr_{R \sim \mathcal{D}_p}[\sigma_R(m) \neq 0] \leq p^{t+1} < p^t.$$

By the union bound, the probability that some $\sigma_R(Q_{ij})$ contains a monomial of support t is at most $\mathfrak{s} \cdot p^t$. □

Choosing the parameters t, p and $\mathfrak{s}$: Set $t = \sqrt{d}$, $p = d^{-\varepsilon}$ (for an sufficiently small $\varepsilon > 0$ to be fixed later), and suppose $\mathfrak{s} < 2^{\frac{\varepsilon}{2}\sqrt{d}\log d}$. Then,

$$\Pr_{R\sim\mathcal{D}_p}[\exists i,j: \quad \sigma_R(Q_{ij}) \text{ contains a monomial of support more than } t] < 2^{-\frac{\varepsilon}{2}\sqrt{d}\log d} \ll 1.$$

This means, there are "plenty of" subsets R such that the circuit $\mathcal{C}$ restricted to the variables in R (i.e. $\sigma_R(\mathcal{C})$) is a t-supported homogeneous depth-4 circuit. If we can now show that there exists such an R that also keeps $\mathsf{DPSP}_{k,\ell,e}(\sigma_R(\mathsf{NW}_r))$ sufficiently close to $\min\left(\binom{N}{k}\cdot\binom{N}{\ell}, \binom{N}{\ell+e}\right)$ then we are done as before (by our discussion in Section 5.1). The following lemma together with Lemma 34 show this.

Lemma 35.

$$\Pr_{R\sim\mathcal{D}_p}\left[\mathsf{DPSP}_{k,\ell,e}(\sigma_R(\mathsf{NW}_r)) < \frac{p^k}{d^{\Theta(1)}}\cdot\min\left(\binom{N}{k}\cdot\binom{N}{\ell}, \binom{N}{\ell+e}\right)\right] < \frac{1}{d^{\Theta(1)}}.$$

Proof. To prove this lemma, we need to examine how the setting of variables in $R \sim \mathcal{D}_p$ to zero effects the dimension of projected shifted partials of NW_r. Clearly

$$\sigma_R(\mathsf{NW}_r) = \sum_{D\in\mathrm{Supp}(\mathsf{NW}_r)} e_D \cdot \mathbf{x}_D,$$

where e_D is an indicator variable such that $e_D = 1$ if $\sigma_R(\mathbf{x}_D) \neq 0$, and $e_D = 0$ otherwise. By definition,

$$\mathsf{DPSP}_{k,\ell,e}(\sigma_R(\mathsf{NW}_r)) = \dim\left(\pi_{\ell+e,\ell+e}(\mathbf{x}^{(=\ell,=\ell)}\cdot\boldsymbol{\partial}^{=k}\sigma_R(\mathsf{NW}_r))\right)$$

Like before (by proposition 18), the above measure corresponds to the rank of a matrix $M_R := M(\sigma_R(\mathsf{NW}_r))$, which in turn equals the rank of $B_R = M_R^T\cdot M_R$ over the field of reals. $\mathrm{rank}(B_R)$ is lower bounded by $\frac{\mathrm{Tr}(B_R)^2}{\mathrm{Tr}(B_R^2)}$ so that it suffices to show that this ratio, namely $\frac{\mathrm{Tr}(B_R)^2}{\mathrm{Tr}(B_R^2)}$ is sufficiently large with high probability when $R\sim\mathcal{D}_p$. Hereafter, we will refer to $\sigma_R(\mathsf{NW}_r)$ as g at some places, and the number of monomials in $\sigma_R(\mathsf{NW}_r)$ as $\mu(g)$. Let $\mathcal{E}_{R\sim\mathcal{D}_p}[Y]$ denote the expected value of a random variable Y when R is chosen according to the distribution $\mathcal{D}_p$. We will at times simply write $\mathcal{E}[Y]$ or $\Pr[\cdot]$ forgoing the subscript $R\sim\mathcal{D}_p$. Note that

$$\begin{aligned}\mu(g) &= \sum_{D\in\mathrm{Supp}(\mathsf{NW}_r)} e_D\\ \Rightarrow \mathcal{E}_{R\sim\mathcal{D}_p}[\mu(g)] &= p^d\cdot d^{2r+2} = \gamma \text{ (say)}\end{aligned}$$

Claim 36. $\Pr_{R\sim\mathcal{D}_p}\left[\mathrm{Tr}(B_R) \le \frac{1}{2}\cdot p^d\cdot\mathrm{Tr}(B)\right] \le \frac{5}{pd}$, *if* $0<\varepsilon<\frac{2}{3}$ *and* $d > \max\left(2^{\frac{1}{1-\varepsilon}}, \frac{1-\varepsilon}{2/3-\varepsilon}\right)$.

Proof. As in proposition 21, $\mathrm{Tr}(B_R) = \mathrm{Tr}(M_R^T\cdot M_R) =$ number of nonzero entries in M_R. Arguing along the same line as in proposition 22,

$$\begin{aligned}\mathrm{Tr}(B_R) &= \mu(g)\cdot\binom{d}{k}\cdot\binom{N-e}{\ell}\\ \Rightarrow \mathcal{E}[\mathrm{Tr}(B_R)] &= \gamma\cdot\binom{d}{k}\cdot\binom{N-e}{\ell} = p^d\cdot\mathrm{Tr}(B)\end{aligned}$$

Hence,

$$\Pr\left[\mathrm{Tr}(B_R) \leq \frac{1}{2}\cdot p^d \cdot \mathrm{Tr}(B)\right] = \Pr\left[\mu(g) \leq \frac{1}{2}\cdot\gamma\right].$$

It turns out that the variance of $\mu(g)$, denoted by $\mathsf{Var}(\mu(g))$, can be upper bounded as follows (see section C in the appendix).

$$\begin{aligned}\mathsf{Var}(\mu(g)) &\leq \gamma\cdot(1-p^d)+\gamma^2\cdot\frac{2}{pd} \quad (\text{if } d > 2^{\frac{1}{1-\varepsilon}})\\ \Rightarrow \Pr\left[\mu(g)\leq\frac{1}{2}\cdot\gamma\right] &\leq \frac{5}{pd} \quad (\text{by Chebyshev's inequality, if } d > \frac{1-\varepsilon}{2/3-\varepsilon})\end{aligned}$$

□

Claim 37. $\Pr\left[\mathrm{Tr}(B_R^2) \geq (2k^3d^2)\cdot\gamma^2\cdot\max\left(\frac{1}{(pd)^k\cdot k!}\cdot\binom{N-2e}{\ell}, \binom{N-2e}{\ell-e}\cdot\binom{d}{k}^2\right)\right] \leq \frac{1}{d}$.

Proof. We argue along the same line as in section 6.1. By propositions 24 and 25, $\mathrm{Tr}(B_R^2)$ equals the number of boxes in M_R such that all the four labels are valid sets in the support of $\sigma_R(\mathsf{NW}_r)$. Observe that proposition 26 is applicable in this setting as well because $\mathrm{Supp}(\sigma_R(\mathsf{NW}_r)) \subseteq \mathrm{Supp}(\mathsf{NW}_r)$. Which means, following the same definitions of $\mu_0(D_1)$, $\mu_1(D_1,D_2)$ and $\mu_2(D_1,D_2)$ (as in corollary 27), we arrive at the following equations:

$$\begin{aligned}\mathrm{Tr}(B_R^2) &= T_0' + T_1' + T_2', \text{ where}\\ T_0' &= \sum_{D_1\in\mathrm{Supp}(\mathsf{NW}_r)} e_{D_1}\cdot|\mu_0(D_1)|\\ T_1' &= \sum_{D_1\neq D_2\in\mathrm{Supp}(\mathsf{NW}_r)} e_{D_1}\cdot e_{D_2}\cdot|\mu_1(D_1,D_2)|\\ T_2' &= \sum_{D_1\neq D_2\in\mathrm{Supp}(\mathsf{NW}_r)} e_{D_1}\cdot e_{D_2}\cdot|\mu_2(D_1,D_2)|,\end{aligned}$$

where e_D's are the indicator variables as defined above. Now, using the facts that $\mathcal{E}[e_D] = p^d$ and $\mathcal{E}[e_{D_1}\cdot e_{D_2}] = p^{d-w}$ if $|D_1\cap D_2| = w$, and mimicking the calculations of sections 6.5 and 6.6, we get the following upper bound.

$$\begin{aligned}\mathcal{E}\left[\mathrm{Tr}(B_R^2)\right] &= \mathcal{E}[T_0'] + \mathcal{E}[T_1'] + \mathcal{E}[T_2']\\ &\leq (2k^3d)\cdot\gamma^2\cdot\max\left(\frac{1}{(pd)^k\cdot k!}\cdot\binom{N-2e}{\ell}, \binom{N-2e}{\ell-e}\cdot\binom{d}{k}^2\right)\end{aligned}$$

By Markov's inequality, $\Pr\left[\mathrm{Tr}(B_R^2) \geq d\cdot\mathcal{E}\left[\mathrm{Tr}(B_R^2)\right]\right] \leq \frac{1}{d}$. This proves Claim 37. □

Using Claims 36 and 37, with probability at least $1 - \frac{6}{pd}$,

$$\begin{aligned}
\mathrm{Tr}(B_R) &> \frac{1}{2} \cdot p^d \cdot \mathrm{Tr}(B) \text{ and} \\
\mathrm{Tr}(B_R^2) &< (2k^3d^2) \cdot \gamma^2 \cdot \max\left(\frac{1}{(pd)^k \cdot k!} \cdot \binom{N-2e}{\ell}, \binom{N-2e}{\ell-e} \cdot \binom{d}{k}^2 \right), \\
\Rightarrow \mathrm{rank}(B_R) &\geq \frac{\mathrm{Tr}(B_R)^2}{\mathrm{Tr}(B_R^2)} \\
&> \frac{p^k}{d^{\Theta(1)}} \cdot \min\left(\binom{N}{k} \cdot \binom{N}{\ell}, \binom{N}{\ell+e} \right) \text{ (by mimicking the previous calculations)}
\end{aligned}$$

The proves Lemma 35 as $\mathrm{rank}(B_R) = \mathsf{DPSP}_{k,\ell,e}(\sigma_R(\mathsf{NW}_r))$. □

By Lemma 34 and 35, and applying union bound, there exists a subset R such that $\sigma_R(\mathcal{C})$ is a t-supported homogeneous depth-4 circuit and

$$\mathsf{DPSP}_{k,\ell,e}(\sigma_R(\mathsf{NW}_r)) \geq \frac{p^k}{d^{\Theta(1)}} \cdot \min\left(\binom{N}{k} \cdot \binom{N}{\ell}, \binom{N}{\ell+e} \right).$$

If we choose a sufficiently small constant ε then $p^k = d^{-\varepsilon k}$ is sufficiently large and the top fanin of $\sigma_R(\mathcal{C})$ (also the top fanin of $\mathcal{C}$) is $2^{\Omega(\sqrt{d}\cdot\log N)}$. Recall that we arrived at this conclusion assuming that the *total sparsity* of $\mathcal{C}$, which was denoted by $\mathfrak{s}$, is less than $2^{\varepsilon/2\cdot\sqrt{d}\cdot\log d}$. Therefore, overall we get a lower bound of $2^{\Omega(\sqrt{d}\cdot\log N)}$ on the size of the homogeneous depth-4 circuit $\mathcal{C}$ computing NW_r.

8 Lower bounds for polynomial families in VP

The Iterated Matrix Multiplication polynomial. Fix any $n, d \in \mathbb{N}$ such that $n, d \geq 2$. Define sets of variables $X_1, \ldots, X_d$ as follows. If $p \in \{1, d\}$, $X_p = \left\{ x_j^{(p)} \,\middle|\, j \in [n] \right\}$ is a set of n variables; otherwise $X_p = \left\{ x_{j,k}^{(p)} \,\middle|\, j,k \in [n] \right\}$ is a set of n^2 variables. Let $X = \bigcup_{p\in[d]} X_p$ and $N := |X| = (d-2)n^2 + 2n$. We think of X_1 and X_d as row and column vectors of variables respectively and of X_p ($p \in [d] \setminus \{1, d\}$) as $n \times n$ matrices of variables. Now, we define the $\mathrm{IMM}_{n,d}(X)$ polynomial as the (unique entry of) the product of the matrices $X_1 \cdots X_d$. Formally,

$$\mathrm{IMM}_{n,d}(X) = \sum_{j_1,\ldots,j_{d-1}} x_{j_1}^{(1)} x_{j_1,j_2}^{(2)} \cdots x_{j_{d-2},j_{d-1}}^{(d-1)} x_{j_{d-1}}^{(d)}$$

An alternate, combinatorial and quite useful way of looking the above polynomial is through the lens of *Algebraic Branching Programs* (ABPs) (see, e.g., [SY10]). Consider a homogeneous ABP $\mathcal{A}$ defined over vertex sets $V_0, \ldots, V_d$ where $V_0 = \{v^{(0)}\}$, $V_d = \{v^{(d)}\}$, and $V_p = \{v_i^{(p)} \mid i \in [n]\}$ for $p \in [d-1]$. The ABP contains all possible edges between V_p and V_{p+1} for $p \in \{0, \ldots, d-1\}$. Each edge e is labelled with a *distinct* variable from X: the edge $e = (v^{(0)}, v_j^{(1)})$ is labelled with $x_j^{(1)}$; $e = (v_i^{(p)}, v_j^{(p+1)})$ is labelled with $x_{i,j}^{(p+1)}$; finally, $e = (v_i^{(d-1)}, v^{(d)})$ is labelled with $x_j^{(d)}$. The ABP computes a polynomial by summing over all paths ρ from $v^{(0)}$ to $v^{(d)}$ the monomial which is obtained by multiplying the variables labelling the edges along the path. It is easily verified that

the polynomial computed this way is $\mathrm{IMM}_{n,d}$. Throughout, we omit mention of the set of variables X if the values of n and d are fixed. Recall that a monomial over the variables in X is said to be *multilinear* if it is not divisible by x^2 for any $x \in X$. Given a monomial $\mathbf{x^i}$, we define the *matrix support of* $\mathbf{x^i}$ — denoted $\mathrm{MSupp}(\mathbf{x^i})$ — to be the set of all $p \in [d]$ such that m is divisible by some $x \in X_p$. We call a monomial $\mathbf{x^i}$ *set-multilinear* if it is multilinear and furthermore, it is divisible by exactly one variable in X_p for each $p \in \mathrm{MSupp}(\mathbf{x^i})$.

Throughout this section, we fix some $n, d \in \mathbb{N}$ and work with $X = \bigcup_{p\in[d]} X_p$, the set of variables over which $\mathrm{IMM}_{n,d}$ is defined.

Remark 38. In what follows we choose to work with a measure that is a slight variant of the dimension of the projected shifted partial measure (defined in section 2). This is only for ease of exposition. We call this variant the dimensions of the *shifted projected partials*.

The measure. Let f be a polynomial in $\mathbb{F}[x_1, \ldots, x_N]$ of degree d. Let S_1 and S_2 be certain fixed subsets of monomials in the N variables. For a polynomial $g = \sum_{\mathbf{i}} c_{\mathbf{i}} \mathbf{x^i}$, where $c_{\mathbf{i}} \in \mathbb{F}$, define $\pi_{S_1}(g) := \sum_{\mathbf{x^i} \in S_1} c_{\mathbf{i}} \mathbf{x^i}$ i.e. $\pi_{S_1}(g)$ is the projection of g onto the monomials in S_1. Consider the following vector space, we call the space of the shifted projected partials of f:

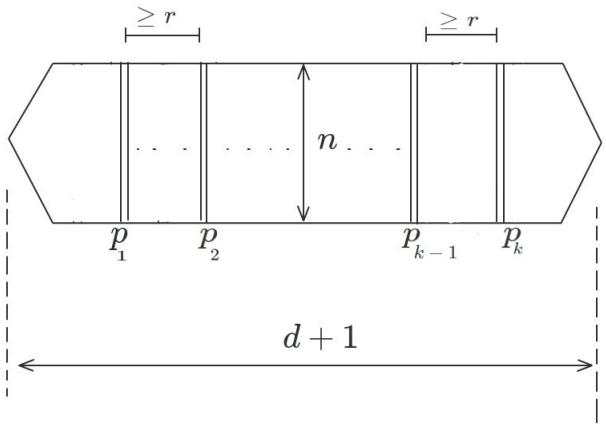
$$\mathcal{V}_{k,\ell}(f) := \mathrm{span}_{\mathbb{F}}\{\mathbf{x^i} \cdot \pi_{S_1}\left(\frac{\partial^k f}{\partial x_{j_1} \ldots \partial x_{j_k}}\right) : |\mathbf{i}| \leq \ell \text{ and } \prod_{q\in[k]} x_{j_q} \in S_2\}. \tag{8}$$

The measure is the dimension of this space, denoted by $\mu_{k,\ell}(f) := \dim(\mathcal{V}_{k,\ell}(f))$. The choices of S_1 and S_2 used for $\mathrm{IMM}_{n,d}$ will be made precise in sections 8.1 and 8.3. The parameters k and ℓ will also be fixed in the analysis later. Since S_1 and S_2 are fixed, it is easy to verify that the measure obeys the subadditivity property.

Lemma 39 (Subadditivity Lemma)**.** *For any $f, g \in \mathbb{F}[X]$, we have $\mu_{k,\ell}(f+g) \leq \mu_{k,\ell}(f) + \mu_{k,\ell}(g)$.*

8.1 The derivatives of $\mathrm{IMM}_{n,d}$

We define the derivative operators as in [FLMS13]. Let $X_1, X_2, \ldots, X_d$ be the matrices that define $\mathrm{IMM}_{n,d}$. Let k be a parameter which will be fixed later and $r = \lfloor \frac{d}{k+1} \rfloor - 1$. We choose evenly spaced k indices $p_1, p_2, \ldots, p_k$, i.e. $p_1, p_2, \ldots, p_k$ are chosen so that for all $1 \leq q \leq k+1$, $p_q - (p_{q-1}+1)) \geq r$, where $p_0 = 0$ and $p_{k+1} = d+1$. Now we choose one variable each from the matrices $X_{p_1}, X_{p_2}, \ldots, X_{p_k}$, say $x^{(p_1)}_{i_1,j_1}, x^{(p_2)}_{i_2,j_2}, \ldots, x^{(p_k)}_{i_k,j_k}$, respectively and take derivatives with respect to them - this defines the set S_2 in Equation (8). More precisely, for any $I = (i_1, j_1, \ldots, i_k, j_k) \in [n]^{2k}$, let m_I denote the monomial $x^{(p_1)}_{i_1,j_1} x^{(p_2)}_{i_2,j_2} \ldots x^{(p_k)}_{i_k,j_k}$ and for a polynomial $F \in \mathbb{F}[X]$, let $\partial_I F$ denote $\left(\frac{\partial^k F}{\partial x^{(p_1)}_{i_1,j_1} \ldots \partial x^{(p_k)}_{i_k,j_k}}\right)$. Then S_2 is the set $\{m_I \mid I \in [n]^{2k}\}$.



8.2 Restriction applied to $\mathrm{IMM}_{n,d}(X)$

We will define a restriction as in Section 6 of [FLMS13]. Fix $p'_1, \ldots, p'_{k+1} \in [d]$ such that for each $q \in [k+1]$, we have $\min\{p'_q - (p_{q-1}+1), p_q - (p'_q+1)\} \geq \lfloor \frac{r-1}{2} \rfloor$, where $p_0, \ldots, p_{k+1}, r$ are as defined in section 8.1. Let $P' = \{p_q \mid q \in [k]\} \cup \{p'_q \mid q \in [k+1]\}$. For $j_1, j_d \in [n]$ and tuple of bijections $B = (\phi_p \in S_n : p \in [d] \setminus (P' \cup \{1, d\}))$, we define the restriction $\tau = \tau_{j_1, j_d, B}$ as follows: For $x \in X$

$$\tau(x) = \begin{cases} 0 & \text{if } x = x_j^{(1)} \text{ for } j \neq j_1, \\ 0 & \text{if } x = x_j^{(d)} \text{ for } j \neq j_d, \\ 0 & \text{if } x = x_{i,j}^{(p)} \text{ for } p \in [d] \setminus (P' \cup \{1, d\}) \text{ and } \phi_p(i) \neq j, \\ x & \text{otherwise.} \end{cases}$$

We denote by $\mathcal{R}$ the set of all such restrictions. Given a restriction $\sigma \in \mathcal{R}$ and a polynomial $f \in \mathbb{F}[X]$, we denote by $f|_\sigma$ the polynomial $f(\sigma(x) : x \in X)$. Let $\tau_0 = \tau_{1,1,B_0}$ where B_0 is a tuple of identity permutations and let $F = \mathrm{IMM}_{n,d}|_{\tau_0}$.

8.3 Measure $\mu_{k,\ell}$ applied to a restriction of $\mathrm{IMM}_{n,d}(X)$

Just as in [FLMS13], we work with the special restriction $F = \mathrm{IMM}_{n,d}|_{\tau_0}$ for the ease of presentation. The lower bound on the measure given by Lemma 40 (below) holds for every restriction τ applied to $\mathrm{IMM}_{n,d}$ i.e. for every $\mathrm{IMM}_{n,d}|_\tau$. In [FLMS13] it was proved that the dimension of the shifted partials space of F is *large*. It turns out that the measure $\mu_{k,\ell}(F)$ is exactly equal to the the dimension of the shifted partials space of F, if the set S_1 in Equation (8) is defined as follows.

The projection π_{S_1}: The map π_{S_1} becomes well defined once we specify the set S_1. Let $p_1, p_2, \ldots, p_k$ be as defined in Section 8.1. The set S_1 is defined as the set of all set-multilinear monomials which are supported on variables in $X \setminus \left(\cup_q X_{p_q}\right)$. We can now prove this lemma formally.

Lemma 40. *Let $k, \ell \in \mathbb{N}$ be arbitrary parameters such that $20k < d < \ell$ and $k \geq 2$. Then,*

$$\mu_{k,\ell}(F) \geq M \cdot \binom{N+\ell}{\ell} - M^2 \cdot \binom{N+\ell-d/40}{\ell-d/40},$$

where $M = \lfloor n^{1.5k} \rfloor$.

The proof of this lemma follows that of [FLMS13, Lemma 11] closely. For completeness, the entire proof is presented in section D of the appendix.

8.4 Lower bounds for certain $\Sigma\Pi\Sigma\Pi$ formulas

In this section, we prove a lower bound for certain variants of $\Sigma\Pi\Sigma\Pi$ formulas that we define below. Fix n and d and let X be the set of input variables to $\mathrm{IMM}_{n,d}$. Let Z denote the set $\bigcup_{p \in P'} X_p$ — where P' is as defined in section 8.2 — and $Y = X \setminus Z$. Let $\mathcal{J}$ denote the ideal generated by all the *non*-set-multilinear monomials over X.

Given $X' \subseteq X$ and $f \in \mathbb{F}[X]$, we denote by $\deg_{X'}(f)$ the degree of f seen as a polynomial over the variables in X' with coefficients from $\mathbb{F}[X \setminus X']$.

Definition 2 ($\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}_Y$ formulas)**.** *An $\Sigma\Pi\Sigma\Pi$ formula C is said to be an $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}_Y$ formula if the fan-ins of its layer 3 multiplication gates are bounded by D, and the layer 1 Π gates in C compute monomials $\mathbf{x^i}$ s.t. $\deg_Y(\mathbf{x^i}) \leq t$.*

The main result of this section is the following:

Lemma 41. *For large enough $n, d \in \mathbb{N}$, any $D \in \mathbb{N}$ and $t, k \in \mathbb{N}$ such that $t \geq 4$ and $kt \leq d/1000$, the following holds. Let C be a $\Sigma\Pi^{[D]}\Sigma\Pi^{[\lceil t/2 \rceil]}_Y$ formula such that $C = \mathrm{IMM}_{n,d}|_\sigma \pmod{\mathcal{J}}$ for some $\sigma \in \mathcal{R}$. Then, the top fan-in of C is at least $\frac{1}{4 \cdot 2^d} \left(\frac{n^{1.25} k}{eD} \right)^k$.*

The proof of the above combines Lemma 40 along with an upper bound on the dimension of the shifted projected partial derivative space of C. To be precise, we prove the following:

Lemma 42. *For any $n, d, D, k, \ell \geq 2$, we have the following. Let C be a $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}_Y$ formula over the variables in X of top fan-in s and let f be any polynomial from $\mathcal{J}$. Then, we have*

$$\mu_{k,\ell}(C + f) \leq s \cdot 2^d \cdot \binom{D}{k} \cdot \binom{N + \ell + (t+1)k}{\ell + (t+1)k}$$

Assuming the above lemma, let us finish the proof of Lemma 41. We will need the following technical facts (see [FLMS13, Section 3] for the proof of Fact 2).

Fact 1. *For any integers N, ℓ, r such that $r < \ell$, we have*

$$\left(\frac{N+\ell}{\ell} \right)^r \leq \frac{\binom{N+\ell}{\ell}}{\binom{N+\ell-r}{\ell-r}} \leq \left(\frac{N+\ell-r}{\ell-r} \right)^r .$$

Fact 2. *For any integers $n, d \geq 2$, $N = (d-2)n^2 + 2dn$ and $t \geq 1$, there exists an integer $\ell > d$ such that $n^{1/16} \leq \left(\frac{N+\ell}{\ell} \right)^t \leq n^{1/4}$.*

Proof of Lemma 41. [FLMS13, Claim 14] observe that all the polynomials $\mathrm{IMM}_{n,d}|_\sigma$ are equivalent in the sense that they can be transformed to one another by permuting the variables in each X_p ($p \in [d]$) suitably, which also preserves the ideal $\mathcal{J}$. Thus, it suffices to prove the lemma for $F = \mathrm{IMM}_{n,d}|_{\tau_0}$ only.

By Fact 2, we can fix $\ell \in \mathbb{N}$ such that $n^{1/16} \leq \left(\frac{N+\ell}{\ell}\right)^t \leq n^{1/4}$. For this choice of ℓ, we first lower bound $\mu_{k,\ell}(F)$ using Lemma 40, which tells us that

$$\mu_{k,\ell}(F) \geq M \cdot \binom{N+\ell}{\ell} - M^2 \binom{N+\ell-d/40}{\ell-d/40} \tag{9}$$

where $M = \lfloor n^{1.5k} \rfloor$.
Note that for our setting of parameters, we have

$$\begin{aligned} \frac{M\binom{N+\ell}{\ell}}{M^2\binom{N+\ell-d/40}{\ell-d/40}} &\geq \frac{1}{n^{1.5k}} \cdot \left(\frac{N+\ell}{\ell}\right)^{d/40} && \text{(by Fact 1)} \\ &\geq \frac{(n^{1/16t})^{d/40}}{n^{1.5k}} \geq n^{\Omega(k)} \geq 2 \end{aligned}$$

for large enough n. Thus, using the above and (9), we obtain that

$$\mu_{k,\ell}(F) \geq \frac{M}{2} \cdot \binom{N+\ell}{\ell} \tag{10}$$

Now, since $C = F \pmod{\mathcal{J}}$, we have $F = C + f$ for some polynomial $f \in \mathcal{J}$. Then, Lemma 42 and Inequality (10) above together imply that

$$\begin{aligned} s &\geq \frac{M}{2 \cdot 2^d \cdot \binom{D}{k}} \cdot \frac{\binom{N+\ell}{\ell}}{\binom{N+\ell+(\lceil t/2 \rceil+1)k}{\ell+(\lceil t/2 \rceil+1)k}} \geq \frac{1}{2 \cdot 2^d} \frac{n^{1.5k}/2}{(\frac{eD}{k})^k} \cdot \frac{\binom{N+\ell}{\ell}}{\binom{N+\ell+tk}{\ell+tk}} \\ &\geq \frac{1}{4 \cdot 2^d} \frac{n^{1.5k}}{(\frac{eD}{k})^k} \cdot \frac{1}{\left(\frac{N+\ell}{\ell}\right)^{tk}} && \text{(by Fact 1)} \\ &\geq \frac{1}{4 \cdot 2^d} \left(\frac{n^{1.5}k}{eD \cdot \left(\frac{N+\ell}{\ell}\right)^t}\right)^k \geq \frac{1}{4 \cdot 2^d} \left(\frac{n^{1.5}k}{eD \cdot n^{1/4}}\right)^k && \text{(by choice of } \ell) \\ &\geq \frac{1}{4 \cdot 2^d} \left(\frac{n^{1.25}k}{eD}\right)^k, \end{aligned}$$

which implies the lemma. □

All that remains is to prove Lemma 42, which is done in section E of the appendix.

8.5 Lower bounds for $\Sigma\Pi\Sigma\Pi$ homogeneous formulas

In this section, we prove Theorems 3 and 4. The idea of the proof is to show that if $\mathrm{IMM}_{n,d}$ or Det_n has a small $\Sigma\Pi\Sigma\Pi$ homogeneous formula, then there is a restriction $\sigma \in \mathcal{R}$ such that $\mathrm{IMM}_{n,d}|_\sigma$ has a small $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}_Y$ formula $\pmod{\mathcal{J}}$ (for suitably chosen t and k). We then appeal to Lemma 41 to get the result. We first prove a restriction lemma for $\mathrm{IMM}_{n,d}$.

Lemma 43. *For all large enough $n, d \in \mathbb{N}$, any $D, t, k \geq 1$, we have the following. If $\mathrm{IMM}_{n,d}$ has a $\Sigma\Pi^{[D]}\Sigma\Pi$ formula of size $\mathfrak{s} < n^{t/10}$, then there is a restriction $\sigma \in \mathcal{R}$ and a $\Sigma\Pi^{[D]}\Sigma\Pi^{[\lceil t/2 \rceil]}_Y$ formula C' of size at most $\mathfrak{s}$ such that $C' = \mathrm{IMM}_{n,d}|_\sigma \pmod{\mathcal{J}}$. Moreover, if C is also homogeneous, then we can find a homogeneous C' satisfying the above.*

Proof. We show that a random $\sigma \in \mathcal{R}$ will meet our requirements with good probability. Formally, choose $j_1, j_d \in [d]$ and $B = (\phi_p \in S_n : p \in [d] \setminus (P' \cup \{1, d\}))$ each independently and uniformly at random and set $\sigma = \tau_{j_1,j_d,B}$ as defined in section 8.2. Note that for $p \in [d] \setminus P'$, each variable $x \in X_p$ is set to 0 with probability $1 - 1/n$; moreover, the restrictions in $X_p, X_{p'}$ for $p \neq p'$ are independent.

Let C_1 be the formula obtained by setting all variables $x \in X$ to $\sigma(x)$ and removing Π-gates at layer 1 which have an input set to 0; clearly, C_1 is a $\Sigma\Pi^{[D]}\Sigma\Pi$ formula that computes $\mathrm{IMM}_{n,d}|_\sigma$. We call a $\sigma \in \mathcal{R}$ *good* if every gate g at layer 1 in C computing a *set-multilinear* monomial such that $\deg_Y(g) > \lceil t/2 \rceil$ has as input some variable that is set to 0 by σ and hence removed from C_1. We claim that σ is good with probability at least $1/2$.

To see this, consider any gate g at layer 1 in C computing a set-multilinear monomial $\mathbf{x^i}$ such that $\deg_Y(g) = |\mathrm{MSupp}(\mathbf{x^i}) \cap ([d] \setminus P')| > \lceil t/2 \rceil$. We can factor $\mathbf{x^i}$ as $(\prod_{p \in \mathrm{MSupp}(\mathbf{x^i}) \cap ([d] \setminus P')} x_p) \cdot \mathbf{x^j}$ for some monomial $\mathbf{x^j}$. Then, g survives in C_1 iff no variable x_p ($p \in \mathrm{MSupp}(\mathbf{x^i}) \cap ([d] \setminus P')$) is set to 0 by σ. Since the probability that each such x_p is not set to 0 is at most $1/n$ and this event is independent for distinct p, the probability that g survives in C_1 is at most $\frac{1}{n^{t/2}}$. Taking a union bound over all such g — of which there are at most $\mathfrak{s}$ many — we see that the probability that any such g survives in C_1 is at most $\mathfrak{s} \cdot \frac{1}{n^{t/2}} \leq 1/2$ for large n since $\mathfrak{s} < n^{t/10}$.

Now, fix any good σ and $C_1 = C|_\sigma$ which computes $\mathrm{IMM}_{n,d}|_\sigma$. Let C' denote the formula obtained from C_1 by removing all gates g at layer 1 such that $\deg_Y(g) > \lceil t/2 \rceil$. By our choice of σ, all such gates compute non-set-multilinear monomials in $\mathcal{J}$. Thus, $C' = \mathrm{IMM}_{n,d}|_\sigma \pmod{\mathcal{J}}$ as claimed in the lemma statement. Moreover, it is clear that C' has size at most the size of C which is $\mathfrak{s}$.
Finally, note that C' was obtained from C by removing some of the monomials computed at layer 1 in C. If C is homogeneous, then we can assume w.l.o.g. that all the monomials feeding into a Σ-gate at layer 2 have the same degree. It thus follows that if C was a homogeneous formula, then so is C'. □

We now prove the lower bound for $\mathrm{IMM}_{n,d}$.

Proof of Theorem 3. We first fix the parameters that we will be using. Choose t, k such that $t = \min\{\lfloor \sqrt{d} \rfloor, \lfloor \log n/5000 \rfloor\}$ and $d/4000 \leq kt \leq d/2000$. Let C be a homogeneous formula of size $\mathfrak{s}$ computing $\mathrm{IMM}_{n,d}$. Note that since C is homogeneous, it is in particular a $\Sigma\Pi^{[d]}\Sigma\Pi$ formula. If $\mathfrak{s} \geq n^{t/10}$, then we have the claimed lower bound and thus we are done.
Otherwise, we can use Lemma 43 and obtain a restriction $\sigma \in \mathcal{R}$ and a *homogeneous* $\Sigma\Pi^{[d]}\Sigma\Pi_Y^{[\lceil t/2 \rceil]}$ formula C' of size at most $\mathfrak{s}$ such that $C' = \mathrm{IMM}_{n,d}|_\sigma \pmod{\mathcal{J}}$. Note that, in particular, the *top fan-in* s of C' is at most $\mathfrak{s}$.

Since C' is $\Sigma\Pi^{[d]}\Sigma\Pi_Y^{[\lceil t/2 \rceil]}$, each input polynomial f to a Π-gate g at layer 3 in C' satisfies $\deg_Y(f) \leq \lceil t/2 \rceil$. We now apply the following transformation to C': if any Π-gate g at layer 3 in C' has two inputs f_1, f_2 such that $\deg_Y(f_1), \deg_Y(f_2) < t/4$, then we replace them by a brute force $\Sigma\Pi_Y^{[\lceil t/2 \rceil]}$ formula computing their homogeneous product $f_1 f_2$. This process clearly ensures that the formula remains $\Sigma\Pi^{[d]}\Sigma\Pi_Y^{[\lceil t/2 \rceil]}$ and moreover, does not increase the top fan-in of C'. We repeatedly apply this transformation to C' until we have an equivalent homogeneous formula C'' of top fan-in at most s that moreover has the property that any Π-gate at layer 3 has at most one input f such

that $\deg_Y(f) < t/4$. In particular, this last property along with the homogeneity of C'' ensures that any layer 3 Π-gate in C'' has fan-in at most $4d/t+1 \leq 5d/t$. Hence, C'' is a $\Sigma\Pi^{[\lfloor 5d/t \rfloor]}\Sigma\Pi_Y^{[\lceil t/2 \rceil]}$ formula of top fan-in at most s such that $C'' = \mathrm{IMM}_{n,d}|_\sigma \pmod{\mathcal{J}}$.

Lemma 41 tells us that by our choice of k and t and for large enough n, we have

$$\mathfrak{s} \geq s \geq \frac{1}{4 \cdot 2^d} \cdot \left(\frac{nkt}{5ed}\right)^k \geq \frac{1}{4 \cdot 2^d} \cdot \left(\frac{n}{60000}\right)^k \geq \frac{n^{d/4500t}}{4 \cdot 2^d} \geq \max\left\{\frac{n^{\Omega(\sqrt{d})}}{4 \cdot 2^d}, 2^{\Omega(d)}\right\}.$$

Note that the above lower bound is $n^{\Omega(\sqrt{d})}$ when $d < \varepsilon \log^2 n$ for a small enough $\varepsilon > 0$; for $d = \Omega(\log^2 n)$, the above is $n^{\Omega(\log n)}$. Thus, we have the theorem. □

We now turn to the lower bound for Det_n. We first need a lemma due to Toda [Tod92]. Given parameters $n_1, d_1 \geq 2$, we let $X(n_1, d_1)$ denote the set of variables over which the polynomial IMM_{n_1,d_1} is defined.

Lemma 44. *For any $n_1, d_1 \in \mathbb{N}^+$ and any $n \geq n_1 d_1$, there is an $n \times n$ matrix M whose entries are either 0, 1, or variables from $X(n_1, d_1)$ such that $\mathsf{Det}_n(M) = \mathrm{IMM}_{n_1,d_1}$.*

Proof of Theorem 4. For large enough n, we can fix n_1 and $d_1 = \Theta(\log^2 n_1)$ such that $n/2 \leq n_1 d_1 \leq n$. Set $t = \lfloor \log n_1/25000 \rfloor$ and $k \geq 10$ such that $d_1/4000 \leq kt \leq d_1/2000$.

Assume that C is a homogeneous $\Sigma\Pi\Sigma\Pi$ formula of size $\mathfrak{s}$ for the polynomial $\mathsf{Det}_n(y_{i,j} : i, j \in [n])$. In particular, note that C is a $\Sigma\Pi^{[n]}\Sigma\Pi$ formula. If $\mathfrak{s} \geq n_1^{t/10} = n^{\Omega(\log n)}$, then we have the claimed lower bound and we are done. Otherwise, we can use Lemma 44 to transform C into an $\Sigma\Pi^{[n]}\Sigma\Pi$ formula C_1 of size at most $\mathfrak{s}$ for IMM_{n_1,d_1} by substituting each $y_{i,j}$ by $M_{i,j}$ throughout C. Now, we can apply Lemma 43 to C_1 and obtain a restriction $\sigma \in \mathcal{R}$ and a $\Sigma\Pi^{[n]}\Sigma\Pi_Y^{[\lceil t/2 \rceil]}$ formula C' of size at most $\mathfrak{s}$ such that $C' = \mathrm{IMM}_{n_1,d_1}|_\sigma \pmod{\mathcal{J}}$. Note in particular that the top fan-in s of C' is at most $\mathfrak{s}$.

Lemma 41 tells us that for large enough n we have

$$s \geq \frac{1}{4 \cdot 2^{d_1}} \cdot \left(\frac{n_1^{1.25}}{en}\right)^k \geq \frac{1}{4 \cdot 2^{d_1}} \cdot \left(\frac{n_1^{1.25}}{2en_1d_1}\right)^k \geq \frac{n^{k/5}}{4 \cdot 2^{d_1}} \geq \frac{n_1^{d_1/20000t}}{4 \cdot 2^{d_1}} = 2^{\Omega(d_1)} = n_1^{\Omega(\log n_1)} = n^{\Omega(\log n)},$$

and since $s \leq \mathfrak{s}$, we have the theorem. □

9 Conclusion

As mentioned in the introduction, proving good enough lower bounds (specifically $2^{\omega(\sqrt{d}\cdot\log N)}$) for homogeneous depth four formulas yields superpolynomial lower bounds for general arithmetic circuits. Our lower bound of $2^{\Omega(\sqrt{d}\cdot\log N)}$ comes temptingly close to this threshold. So a very natural question would be to improve the exponent. A more modest aim might be to further understand the power and limitations of our techniques/complexity measure. With this intent we formulate a concrete conjecture that might serve as the goal of such an undertaking.

Conjecture 45. *There exist a (family of) homogeneous polynomial(s) f of degree d in $N = d^{O(1)}$ variables which can be computed by $\mathsf{poly}(d)$-sized homogeneous circuits of depth six but for which any homogeneous circuit of depth four must have superpolynomial (in d) size.*

Acknowledgements

NK would like to thank Avi Wigderson for many helpful discussions including pointing out the use of random restrictions to reduce a general homogeneous $\Sigma\Pi\Sigma\Pi$ circuit into one with low support. NL and SS would like to thank Hervé Fournier and Guillaume Malod for useful discussions. CS and SS would like to thank Arnab Bhattacharya and Ramprasad Saptharishi for their feedback and encouragement.

References

[AJMV98] Eric Allender, Jia Jiao, Meena Mahajan, and V. Vinay. Non-Commutative Arithmetic Circuits: Depth Reduction and Size Lower Bounds. *Theor. Comput. Sci.*, 209(1-2):47–86, 1998.

[Alo09] Noga Alon. Perturbed Identity Matrices Have High Rank: Proof and Applications. *Combinatorics, Probability & Computing*, 18(1-2):3–15, 2009.

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

[AV08] Manindra Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *FOCS*, pages 67–75. IEEE Computer Society, 2008.

[BDYW11] Boaz Barak, Zeev Dvir, Amir Yehudayoff, and Avi Wigderson. Rank bounds for design matrices with applications toc ombinatorial geometry and locally correctable codes. In *STOC*, pages 519–528, 2011.

[CKW11] Xi Chen, Neeraj Kayal, and Avi Wigderson. Partial Derivatives in Arithmetic Complexity and Beyond. *Foundations and Trends in Theoretical Computer Science*, 6(1-2):1–138, 2011.

[DSW13] Zeev Dvir, Shubhangi Saraf, and Avi Wigderson. Breaking the quadratic barrier for 3-LCCs over the Reals. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:160, 2013.

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

[GK98] Dima Grigoriev and Marek Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *STOC*, pages 577–582, 1998.

[GKKS13a] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the chasm at depth four. In *Proceedings of the Conference on Computational Complexity (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] Dima Grigoriev and Alexander A. Razborov. Exponential complexity lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. In *FOCS*, pages 269–278, 1998.

[Gur10] Venkatesan Guruswami. Introduction to coding theory, Lecture 2: Gilbert-Varshamov bound. http://www.cs.cmu.edu/~venkatg/teaching/codingtheory/, 2010.

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

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

[KS13a] Mrinal Kumar and Shubhangi Saraf. Lower Bounds for Depth 4 Homogenous Circuits with Bounded Top Fanin. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:68, 2013.

[KS13b] Mrinal Kumar and Shubhangi Saraf. Superpolynomial lower bounds for general homogeneous depth 4 arithmetic circuits. Technical Report 181, Electronic Colloquium on Computational Complexity (ECCC), 2013.

[KS13c] Mrinal Kumar and Shubhangi Saraf. The Limits of Depth Reduction for Arithmetic Formulas: Its all about the top fan-in. Technical Report 153, Electronic Colloquium on Computational Complexity (ECCC), 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.

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

[NW97] Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997.

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

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

[Rom] Dan Romik. Stirlings approximation for n!: The ultimate short proof? *The American Mathematical Monthly*, 107(6):556557.

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

[SV85] Sven Skyum and Leslie G. Valiant. A Complexity Theory Based on Boolean Algebra. *J. ACM*, 32(2):484–502, 1985.

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

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

[Tod92] S. Toda. Classes of Arithmetic Circuits Capturing the Complexity of Computing the Determinant. *IEICE Transactions on Information and Systems*, E75-D:116–124, 1992.

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

[VSBR83] Leslie G. Valiant, Sven Skyum, S. Berkowitz, and Charles Rackoff. Fast Parallel Computation of Polynomials Using Few Processors. *SIAM J. Comput.*, 12(4):641–644, 1983.

A Proof of preliminaries

Lemma 46. *Let $a(n)$, $f(n)$, $g(n) : \mathbb{Z}_{>0} \to \mathbb{Z}$ be integer valued function such that $(f+g) = o(a)$. Then,*

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

Proof.

$$\frac{(a+f)!}{(a-g)!} = (a+f)(a+f-1)\dots(a-g+1)$$

$$\implies \quad a^{f+g}\left(1-\frac{g}{a}\right)^{f+g} \quad \leq \quad \frac{(a+f)!}{(a-g)!} \quad \leq \quad a^{f+g}\left(1+\frac{f}{a}\right)^{f+g}$$

$$\implies \quad (f+g)\ln\left(1-\frac{g}{a}\right) \quad \leq \quad \ln\frac{(a+f)!}{(a-g)!} - (f+g)\ln a \quad \leq \quad (f+g)\ln\left(1+\frac{f}{a}\right)$$

Using the fact that $\frac{x}{1+x} \leq \ln(1+x) \leq x$ for $x > -1$, it is easy to see that both the LHS and RHS are bounded by $O\left(\frac{f^2+g^2}{a}\right)$. □

B Proof details for section 6.6

B.1 Estimating T_1

The following calculations show that $S(u,w)$ is maximized at $w = u = k$.

$$\begin{aligned}
\frac{S(u,w+1)}{S(u,w)} &= \frac{1}{d}\cdot\frac{(N-2d+u+w+1)}{(N-\ell-2d+k+w+1)\cdot(w-2k+u+1)} \\
&\leq \frac{1}{d}\cdot\frac{(N)}{(N-\ell-2d)} \quad (\text{since } w-2k+u\geq 0, u+w+1\leq k+w+1\leq 2d) \\
&\leq \frac{1}{4}\cdot\frac{(N)}{(\frac{N}{2}-2d)} \quad \left(\text{for } \ell\leq\frac{N}{2}, d\geq 4\right) \\
&< 1 \quad \left(\text{since } N=d^3\gg d\right)
\end{aligned}$$

Thus for a fixed u, $S(u,w)$ is maximized at $w = 2k-u$. Now for any $u \leq (k-1)$ we have

$$\begin{aligned}
\frac{S(u+1,2k-u-1)}{S(u,2k-u)} &= d\cdot\frac{N-\ell-2d+3k-u}{\ell-k+u+1}\cdot\frac{(k-u)^2}{(u+1)} \\
&> d\cdot\frac{N-\ell-2d}{\ell}\cdot\frac{1}{k} \\
&> 4\cdot\frac{N-\ell-2d}{\ell} \quad (\text{for } k\leq 4d) \\
&> \frac{8}{N}\cdot(\frac{N}{2}-2d) \quad \left(\text{for } \ell<\frac{N}{2}\right) \\
&> 1 \quad \left(\text{since } N=d^3>\frac{16d}{3}\right)
\end{aligned}$$

Thus $S(u,w)$ is maximized at $w = u = k$.

B.2 Estimating T_2

The following calculations show that $S_1(w,u,u_1,u_2)$ is maximized at $w = u = u_1 = u_2 = 0$.

$$\begin{aligned}
\frac{S_1(w+1,u,u_1,u_2)}{S_1(w,u,u_1,u_2)} &= \frac{1}{d}\cdot\frac{(N-2d+2k-u-u_1-u_2+w+1)}{\ell-d+k-u-u_1-u_2+w+1}\cdot\frac{d-k+u+u_1-w}{d-w}\cdot \\
&\quad \cdot\frac{d-k+u+u_2-w}{d-w}\cdot\frac{1}{w-u-u_1-u_2+1} \\
&\leq \frac{1}{d}\cdot\frac{N-2d+2k+w+1}{\ell-d+k-2k+w+1}\cdot\frac{1}{w-u-u_1-u_2+1} \quad (\text{since } u+u_1, u+u_2\leq k) \\
&\leq \frac{1}{d}\cdot\frac{N-2d+2k+d}{\ell-d-k+0} \quad (\text{since } 0\leq u+u_1+u_2<(w+1)<d) \\
&\leq \frac{1}{d}\cdot\frac{N}{\ell-2d} \quad \left(\text{for } k<\frac{d}{2}\right) \\
&< 1 \quad \left(\text{for } \ell>\frac{N}{d}+2d\right)
\end{aligned}$$

So for a fixed u, u_1, u_2, $S(w, u, u_1, u_2)$ is maximized at $w = u + u_1 + u_2$. Let $S_2(u, u_1, u_2) = S_1(u + u_1 + u_2, u, u_1, u_2)$. Then for any $u \leq (k-1)$ we have

$$\begin{aligned}
\frac{S_2(u+1, u_1, u_2)}{S_2(u, u_1, u_2)} &= \frac{1}{d} \cdot \frac{k - u_1 - u}{d - u_1 - u_2 - u} \cdot \frac{k - u_2 - u}{d - u_1 - u_2 - u} \cdot \frac{1}{u+1} \\
&\leq \frac{1}{d} \cdot \frac{k}{d - 2k} \cdot \frac{k}{d - 2k} (\text{since } u + u_1, u + u_2 \leq k) \\
&< \frac{1}{d} \quad \left(\text{for } k < \frac{d}{3}\right) \\
&< 1
\end{aligned}$$

Thus for fixed u_1 and u_2, $S_2(u, u_1, u_2)$ is maximized at $u = 0$. Proceeding in a manner similar to above we see that $S_2(u, u_1, u_2)$ is maximized at

$$u = u_1 = u_2 = 0.$$

C Variance of $\mu(g)$

Recall that

$$\begin{aligned}
\mu(g) &= \sum_{D \in \text{Supp}(\mathsf{NW}_r)} e_D \\
\Rightarrow \mathcal{E}[\mu(g)] &= p^d \cdot d^{2r+2} =: \gamma
\end{aligned}$$

Now, let us bound the variance of $\mu(g)$. In the summations below, D, D_1, D_2 run over all elements in $\text{Supp}(\mathsf{NW}_r)$.

$$\begin{aligned}
\mathsf{Var}(\mu(g)) &= \mathcal{E}[\mu(g)^2] - \mathcal{E}[\mu(g)]^2 \\
&= \mathcal{E}\left[\left(\sum_D e_D\right)^2\right] - \mathcal{E}\left[\sum_D e_D\right]^2 \\
&= \mathcal{E}\left[\sum_D e_D^2 + \sum_{\substack{D_1, D_2 \\ D_1 \neq D_2}} e_{D_1} \cdot e_{D_2}\right] - \left[\sum_D \mathcal{E}[e_D]\right]^2 \text{ (by linearity of expectation)} \\
&= \mathcal{E}\left[\sum_D e_D + \sum_{\substack{D_1, D_2 \\ D_1 \neq D_2}} e_{D_1} \cdot e_{D_2}\right] - \left[\sum_D \mathcal{E}[e_D]^2 + \sum_{\substack{D_1, D_2 \\ D_1 \neq D_2}} \mathcal{E}[e_{D_1}] \cdot \mathcal{E}[e_{D_2}]\right] \text{ (as } e_D^2 = e_D) \\
&= \mathcal{E}\left[\sum_D e_D\right] - \sum_D \mathcal{E}[e_D]^2 + \mathcal{E}\left[\sum_{\substack{D_1, D_2 \\ D_1 \neq D_2}} e_{D_1} \cdot e_{D_2}\right] - \sum_{\substack{D_1, D_2 \\ D_1 \neq D_2}} \mathcal{E}[e_{D_1}] \cdot \mathcal{E}[e_{D_2}]
\end{aligned}$$

$$
\begin{aligned}
\mathsf{Var}(\mu(g)) &= p^d \cdot d^{2r+2} - p^{2d} \cdot d^{2r+2} + \sum_{w=0}^{r} \left[\mathcal{E} \left[\sum_{\substack{D_1, D_2 \\ D_1 \neq D_2, |D_1 \cap D_2| = w}} e_{D_1} \cdot e_{D_2} \right] - \sum_{\substack{D_1, D_2 \\ D_1 \neq D_2, |D_1 \cap D_2| = w}} \mathcal{E}[e_{D_1}] \cdot \mathcal{E}[e_{D_2}] \right] \\
&= \gamma \cdot (1 - p^d) + \sum_{w=0}^{r} \left[\sum_{\substack{D_1, D_2 \\ D_1 \neq D_2, |D_1 \cap D_2| = w}} (\mathcal{E}[e_{D_1} \cdot e_{D_2}] - \mathcal{E}[e_{D_1}] \cdot \mathcal{E}[e_{D_2}]) \right] \\
&\qquad \text{(by linearity of expectation)} \\
&= \gamma \cdot (1 - p^d) + \sum_{w=0}^{r} \left[\sum_{\substack{D_1, D_2 \\ D_1 \neq D_2, |D_1 \cap D_2| = w}} \left(p^d \cdot p^{d-w} - p^d \cdot p^d \right) \right] \\
&\qquad \text{(as } \mathcal{E}[e_{D_2} | e_{D_2} = 1] = p^{d-w} \text{ if } |D_1 \cap D_2| = w) \\
&= \gamma \cdot (1 - p^d) + \sum_{w=1}^{r} \left[\sum_{D_1} \sum_{\substack{D_2 \\ D_1 \neq D_2, |D_1 \cap D_2| = w}} \left(p^{2d-w} - p^{2d} \right) \right] \\
&= \gamma \cdot (1 - p^d) + \sum_{w=1}^{r} \left[\sum_{D_1} R_d(w, r) \cdot p^{2d} \left(p^{-w} - 1 \right) \right] \text{ (recall } R_d(w,r) \text{ from section 6.5)} \\
&\leq \gamma \cdot (1 - p^d) + p^{2d} \cdot \sum_{w=1}^{r} \left[d^{2r+2} \cdot R_d(w, r) \cdot p^{-w} \right] \\
&\leq \gamma \cdot (1 - p^d) + \gamma^2 \cdot \sum_{w=1}^{r} \frac{1}{(pd)^w} \quad \left(\text{ since } R_d(w,r) \leq \frac{d^{2r+2}}{d^w \cdot w!} \right) \\
&\leq \gamma \cdot (1 - p^d) + \gamma^2 \cdot \frac{2}{pd} \quad (\text{as } pd = d^{1-\varepsilon} > 2 \text{ if } d > 2^{\frac{1}{1-\varepsilon}})
\end{aligned}
$$

D Proof of Lemma 40

By the definition of $\mu_{k,\ell}$, we have $\mu_{k,\ell}(F) = \dim(\mathcal{V}_{k,\ell}(f))$ where

$$
\begin{aligned}
\mathcal{V}_{k,\ell}(F) &:= \mathrm{span}_{\mathbb{F}} \{ \mathbf{x}^{\mathbf{i}} \cdot \pi_{S_1} \left(\frac{\partial^k F}{\partial x_{i_1, j_1}^{(p_1)} \dots \partial x_{i_k, j_k}^{(p_k)}} \right) \mid |\mathbf{i}| \leq \ell \text{ and } \prod_{q \in [k]} x_{i_q, j_q}^{(p_q)} \in S_2 \} \\
&= \mathrm{span}_{\mathbb{F}} \{ \mathbf{x}^{\mathbf{i}} \cdot \pi_{S_1} \left(\partial_I F \right) \mid |\mathbf{i}| \leq \ell \text{ and } I \in [n]^{2k} \}
\end{aligned}
$$

First observe that any $\partial_I F$ is a monomial given by $\rho_1 \rho_2 \dots \rho_{k+1}$, where

$$\rho_1 = \underbrace{\left(x_1^{(1)} \cdot \prod_{1<p<p_1'} x_{1,1}^{(p)}\right)}_{g_1^I} \cdot x_{1,i_1}^{(p_1')} \cdot \underbrace{\left(\prod_{p_1'<p<p_1} x_{i_1,i_1}^{(p)}\right)}_{h_1^I}$$

$$\rho_q = \underbrace{\left(\prod_{p_{q-1}<p<p_q'} x_{j_{q-1},j_{q-1}}^{(p)}\right)}_{g_q^I} \cdot x_{j_{q-1},i_q}^{(p_q')} \cdot \underbrace{\left(\prod_{p_q'<p<p_q} x_{i_q,i_q}^{(p)}\right)}_{h_q^I} \quad \text{(for } 1 < q < k+1\text{)}$$

$$\rho_{k+1} = \underbrace{\left(\prod_{p_k<p<p_{k+1}'} x_{j_k,j_k}^{(p)}\right)}_{g_{k+1}^I} \cdot x_{j_k,1}^{(p_{k+1}')} \cdot \underbrace{\left(\left(\prod_{p_{k+1}'<p<d} x_{1,1}^{(p)}\right) \cdot x_1^{(d)}\right)}_{h_{k+1}^I}$$

Due to the above structure of $\partial_I F$ we have the following claim.

Claim 47. $\forall I \in [n]^{2k}$, $\partial_I F \in S_1$.

Claim 47 implies that for all $I \in [n]^{2k}$, $\pi_{S_1}(\partial_I F) = \partial_I F$. Therefore, we get

$$\mathcal{V}_{k,\ell}(F) = \text{span}_{\mathbb{F}}\{\mathbf{x}^{\mathbf{i}} \cdot \partial_I F : |\mathbf{i}| \leq \ell \text{ and } I \in [n]^{2k}\}$$

The analysis of the dimension of $\mathcal{V}_{k,\ell}(F)$ is now very similar to the analysis of the dimension of the shifted partial derivative space of F as done in [FLMS13].
Let $\mathcal{M} = \{\mathbf{x}^{\mathbf{i}} \cdot \partial_I F : |\mathbf{i}| \leq \ell \text{ and } I \in [n]^{2k}\}$. Since $\mathcal{M}$ is a set of *monomials*, the dimension of the span of $\mathcal{M}$ is exactly $|\mathcal{M}|$.
Another way of looking at $\mathcal{M}$ is $\mathcal{M} = \bigcup_{I\in[n]^{2k}} \mathcal{M}_I$, where $\mathcal{M}_I := \{\mathbf{x}^{\mathbf{i}} \mid |\mathbf{i}| \leq \ell + d - k \text{ and } \partial_I F \text{ divides } \mathbf{x}^{\mathbf{i}}\}$. Therefore, we have the following claim.

Claim 48. *For F and $\mathcal{M}_I$ ($I \in [n]^{2k}$) as defined above, we have $\dim(\mathcal{V}_{k,\ell}(F)) = |\mathcal{M}|$, where $\mathcal{M} = \bigcup_{I\in[n]^{2k}} \mathcal{M}_I$.*

In what follows, we do not distinguish between multilinear monomials over the variable set X and subsets of X.

Claim 49. *For any $I, I' \in [n]^{2k}$, we have*

$$|\partial_{I'} F \setminus \partial_I F| \geq \Delta(I, I') \cdot \left\lfloor \frac{r-1}{2} \right\rfloor$$

where $\Delta(I, I')$ denotes the Hamming distance between I and I'.

Proof. Consider any $I, I' \in [n]^{2k}$. Say $I = (i_1, j_1, \ldots, i_k, j_k)$ and $I' = (i_1', j_1', \ldots, i_k', j_k')$. Then, using the notation from the definition of $\partial_I F$, we have

$$\begin{aligned}\partial_{I'} F \setminus \partial_I F &\supseteq \dot{\bigcup_{q\in[k]}} (g_{q+1}^{I'} \setminus g_{q+1}^I) \dot{\cup} \dot{\bigcup_{q\in[k]}} (h_q^{I'} \setminus h_q^I) \\ &\supseteq \dot{\bigcup_{q\in[k]:j_q\neq j_q'}} (g_{q+1}^{I'} \setminus g_{q+1}^I) \dot{\cup} \dot{\bigcup_{q\in[k]:i_q\neq i_q'}} (h_q^{I'} \setminus h_q^I).\end{aligned}$$

where $A \dot\cup B$ denotes the union of disjoint sets A and B.
Now, when $j_q \neq j'_q$, then the monomials g^I_{q+1} and $g^{I'}_{q+1}$ are disjoint and hence $|g^{I'}_{q+1} \setminus g^I_{q+1}| = |g^{I'}_{q+1}| \geq \left\lfloor \frac{r-1}{2} \right\rfloor$. Similarly, when $i_q \neq i'_q$, we have $|h^{I'}_q \setminus h^I_q| \geq \left\lfloor \frac{r-1}{2} \right\rfloor$.

$$\begin{aligned} |\partial_{I'} F \setminus \partial_I F| &\geq \sum_{q \in [k] : j_q \neq j'_q} |g^{I'}_{q+1} \setminus g^I_{q+1}| + \sum_{q \in [k] : i_q \neq i'_q} |h^{I'}_q \setminus h^I_q| \\ &\geq \Delta(I, I') \cdot \left\lfloor \frac{r-1}{2} \right\rfloor, \end{aligned}$$

which completes the proof of the claim. □

Claim 50. *For any $I \in [n]^{2k}$, we have $|\mathcal{M}_I| = \binom{N+\ell}{\ell}$.*

Proof. A monomial $\mathbf{x^i} \in \mathcal{M}_I$ iff there is a monomial $\mathbf{x^j}$ such that $\mathbf{j} \leq \ell$ and $\mathbf{x^i} = \mathbf{x^j} \cdot \partial_I F$. Thus, $|\mathcal{M}_I|$ is equal to the number of monomials of degree at most ℓ, which is $\binom{N+\ell}{\ell}$. □

Claim 51. *For any $I, I' \in [n]^{2k}$, we have*

$$|\mathcal{M}_I \cap \mathcal{M}_{I'}| = \binom{N + \ell - |(\partial_{I'} F \setminus \partial_I F)|}{\ell - |(\partial_{I'} F \setminus \partial_I F)|}.$$

Proof. Fix any I, I' as above. Any monomial $\mathbf{x^i} \in \mathcal{M}_I \cap \mathcal{M}_{I'}$ may be factored as $\mathbf{x^i} = \mathbf{x^j} \cdot \partial_I F \cdot (\partial_{I'} F \setminus \partial_I F)$. where $j \leq \ell + d - k - (d-k) - |(\partial_{I'} F \setminus \partial_I F)| = \ell - |(\partial_{I'} F \setminus \partial_I F)|$.
Thus, $|\mathcal{M}_I \cap \mathcal{M}_{I'}|$ is equal to the number of monomials of degree at most $\ell - |(\partial_{I'} F \setminus \partial_I F)|$, from which the claim follows. □

Claim 52. *For any $k \in \mathbb{N}$ and large enough $n \in \mathbb{N}$, there exists an $\mathcal{S} \subseteq [n]^{2k}$ such that*

- $|\mathcal{S}| = \lfloor n^{1.5k} \rfloor$,
- *For all distinct $I, I' \in \mathcal{S}$, we have $\Delta(I, I') \geq k/4$.*

Proof. We construct the set $\mathcal{S}$ by first greedily choosing vectors which have pairwise Hamming distance at least $k/4$ and then prove that the set thus formed has size $\lfloor n^{1.5k} \rfloor$. A standard volume argument [Gur10] gives that the set picked greedily as above has size at least $\frac{n^{2k}}{\mathrm{Vol}_n(2k, k/4)}$, where $\mathrm{Vol}_n(2k, k/4)$ stands for the volume of the Hamming ball of radius k for strings of length $2k$ over an alphabet of size n. It is easy to see that $\mathrm{Vol}_n(2k, k/4) = \sum_{i=0}^{k/4} \binom{2k}{i} (n-1)^i$, which is upper bounded by $2\binom{2k}{k/4}(n-1)^{k/4}$. This in turn is at most $n^{k/3}$ for large enough n. Therefore, $|\mathcal{S}|$ is at least $\frac{n^{2k}}{n^{k/3}}$, i.e. $|\mathcal{S}| \geq n^{5k/3}$. By choosing a subcollection of the vectors thus chosen, we can ensure that $|\mathcal{S}|$ is exactly $\lfloor n^{1.5k} \rfloor$. □

Recall that a very similar claim (Claim 10) proved in [FLMS13] gave $\mathcal{S} = \left\lfloor \left(\frac{n}{4}\right)^k \right\rfloor$. This size of $\mathcal{S}$ was sufficient for the proof of Lemma 11 in [FLMS13]. We will now see that a slightly larger sized $\mathcal{S}$ will be useful for us to prove Lemma 40.

Proof of Lemma 40. Fix $\mathcal{S}$ as guaranteed by Claim 52. By Claim 48, it suffices to lower bound $|\mathcal{M}|$. For this, we use inclusion-exclusion. Since $\mathcal{M} = \bigcup_I \mathcal{M}_I$, we have

$$\begin{aligned} |\mathcal{M}| &\geq |\bigcup_{I \in \mathcal{S}} \mathcal{M}_I| \\ &\geq \sum_{I \in \mathcal{S}} |\mathcal{M}_I| - \sum_{I \neq I' \in \mathcal{S}} |\mathcal{M}_I \cap \mathcal{M}_{I'}|. \end{aligned} \tag{11}$$

By Claim 50, we know that $|\mathcal{M}_I| = \binom{N+\ell}{\ell}$. By Claims 51 and 49 and our choice of $\mathcal{S}$, we see that for any distinct $I, I' \in \mathcal{S}$, we have

$$|\mathcal{M}_I \cap \mathcal{M}_{I'}| \leq \binom{N + \ell - k/4 \cdot \lfloor (r-1)/2 \rfloor}{\ell - k/4 \cdot \lfloor (r-1)/2 \rfloor} \leq \binom{N + \ell - d/40}{\ell - d/40}$$

where the last inequality follows since $\lfloor (r-1)/2 \rfloor \geq d/10k$ for $k \leq d/20$ (recall that r denotes $\left\lfloor \frac{d}{k+1} \right\rfloor - 1$).
Plugging the above into (11), we obtain

$$|\mathcal{M}| \geq |\mathcal{S}| \cdot \binom{N+\ell}{\ell} - |\mathcal{S}|^2 \cdot \binom{N + \ell - d/40}{\ell - d/40}.$$

Since $|\mathcal{S}| = \lfloor n^{1.5k} \rfloor$, the lemma follows.

□

E Proof of Lemma 42

Fix C and f as in the statement of the lemma. By Lemma 39, we know that $\mu_{k,\ell}(C + f) \leq \mu_{k,\ell}(C) + \mu_{k,\ell}(f)$. The latter term is handled first.

Claim 53. *For every $g \in \mathcal{J}$ and every $I \in [n]^{2k}$, we have $\pi_{S_1}(\partial_I g) = 0$. In particular, $\mu_{k,\ell}(g) = 0$.*

Proof. By linearity, it suffices to prove the above for every monomial $\mathbf{x^i} \in \mathcal{J}$. Since $\mathbf{x^i}$ is non-set multilinear, there exists some $p \in [d]$ and $x, y \in X_p$ (possibly equal) such that $xy | \mathbf{x^i}$. There are two cases to consider:

- $p \notin \{p_1, \ldots, p_k\}$: In this case, it is easy to see that $xy | \partial_I \mathbf{x^i}$ as well and hence $\pi_{S_1}(\partial_I \mathbf{x^i}) = 0$.
- $p \in \{p_1, \ldots, p_k\}$: Either x and y are distinct or $x = y$. In the former case, we note that since we derive w.r.t. at most one of x or y, it must be the case that either $x | \partial_I \mathbf{x^i}$ or $y | \partial_I \mathbf{x^i}$. In the latter case, since we derive at most once w.r.t. x, we have $x | \partial_I \mathbf{x^i}$. In either case, $\pi_{S_1}(\partial_I \mathbf{x^i}) = 0$.

□

Thus, we only need to bound $\mu_{k,\ell}(C)$. Assume that $C = \sum_{i=1}^{s} C_i$, where each C_i is a $\Pi^{[D]}\Sigma\Pi^{[t]}_Y$ formula. By Lemma 39, it suffices to show that for each $i \in [s]$, we have

$$\mu_{k,\ell}(C_i) \leq 2^d \cdot \binom{D}{k} \cdot \binom{N + \ell + (t+1)k}{\ell + (t+1)k} \tag{12}$$

Let $i \in [s]$ be fixed for the rest of the proof. We may assume that the top fan-in of C_i is exactly D and hence $C_i = \prod_{p\in[D]} Q_p$ where $\deg_Y(Q_p) \le t$ for each $p \in [d]$. Consider any $I \in [n]^{2k}$. By the product rule for differentiation, we can see that $\partial_I C_i$ can be written as

$$\partial_I(C_i) = \sum_{A\subseteq[D]:|A|=D-k} \left(\prod_{p\in A} Q_p\right) \cdot Q'_{I,A}$$

where for each A, $Q'_{I,A}$ satisfies $\deg_Y(Q'_{I,A}) \le tk$. Let Q_A denote $\prod_{p\in A} Q_p$. Hence we have

$$\left\{\partial_I(C_i) \;\middle|\; I \in [n]^{2k}\right\} \subseteq \mathrm{span}_{\mathbb{F}} \left\{Q_A \cdot \mathbf{x^j} \;\middle|\; A \subseteq [D], |A| = D-k, \deg_Y(\mathbf{x^j}) \le tk\right\}$$

Thus, we have by linearity,

$$\left\{\pi_{S_1}(\partial_I(C_i)) \;\middle|\; I \in [n]^{2k}\right\} \subseteq \mathrm{span}_{\mathbb{F}} \left\{\pi_{S_1}(Q_A \cdot \mathbf{x^j}) \;\middle|\; |A| = D-k, \deg_Y(\mathbf{x^j}) \le tk\right\}$$

Now, by the definition of π_{S_1}, $\pi_{S_1}(Q_A \cdot \mathbf{x^j}) = 0$ if either $\mathbf{x^j}$ is non-set-multilinear or it is divisible by a variable in $\bigcup_{q\in[k]} X_{p_q}$. Thus, in the expression above, we may range only over $\mathbf{x^j}$ that are set-multilinear and not divisible by any $x \in \bigcup_{q\in[k]} X_{p_q}$. In particular, this implies that $\deg_Z(\mathbf{x^j}) \le k$ (recall that $Z = \bigcup_{p\in P'} X_p$) and hence $|\mathbf{j}| = \deg_Y(\mathbf{x^j}) + \deg_Z(\mathbf{x^j}) \le tk + k = (t+1)k$. Thus, we get

$$\left\{\pi_{S_1}(\partial_I(C_i)) \;\middle|\; I \in [n]^{2k}\right\} \subseteq \mathrm{span}_{\mathbb{F}} \left\{\pi_{S_1}(Q_A \cdot \mathbf{x^j}) \;\middle|\; |A| = D-k, \mathbf{x^j} \in \mathcal{M}_X^{sm}, |\mathbf{j}| \le (t+1)k\right\} \tag{13}$$

where we use $\mathcal{M}_X^{sm}$ to denote the set of all set-multilinear monomials over X.
To analyze the above, decompose Q_A further as

$$Q_A = Q_A^{nsm} + \sum_{B\subseteq[d]} Q_A^B$$

where Q_A^{nsm} is the sum of all the *non*-set-multilinear monomials in Q_A (with the same coefficients) and Q_A^B (for each $B \subseteq [d]$) is a linear-combination of set-multilinear monomials $\mathbf{x^{i_1}}$ appearing in Q_A such that $\mathrm{MSupp}(\mathbf{x^{i_1}}) = B$.
Since non-set-multilinear monomials lie in the kernel of π_{S_1} we have for any $\mathbf{x^j} \in \mathcal{M}_X^{sm}$,

$$\pi_{S_1}(Q_A \cdot \mathbf{x^j}) = \pi_{S_1}(Q_A^{nsm} \cdot \mathbf{x^j}) + \sum_{B\subseteq[d]} \pi_{S_1}(Q_A^B \cdot \mathbf{x^j}) = 0 + \sum_{B\subseteq[d]} \pi_{S_1}(Q_A^B \cdot \mathbf{x^j}) \tag{14}$$

What follows is a crucial observation: for any $B \subseteq [d]$ and any $\mathbf{x^j} \in \mathcal{M}_X^{sm}$,

$$\pi_{S_1}(Q_A^B \cdot \mathbf{x^j}) = \begin{cases} 0, & \text{if } B \cap \{p_1, \ldots, p_k\} \neq \emptyset, \\ 0, & \text{if } \mathrm{MSupp}(\mathbf{x^j}) \cap \{p_1, \ldots, p_k\} \neq \emptyset, \\ 0, & \text{if } \mathrm{MSupp}(\mathbf{x^j}) \cap B \neq \emptyset, \\ Q_A^B \cdot \mathbf{x^j}, & \text{otherwise.} \end{cases}$$

In particular, along with (14), this implies that for any $\mathbf{x^j} \in \mathcal{M}_X^{sm}$, the polynomial $\pi_{S_1}(Q_A \cdot \mathbf{x^j})$ lies in $\mathrm{span}_{\mathbb{F}} \left\{Q_A^B \cdot \mathbf{x^j} \;\middle|\; B \subseteq [d]\right\}$. Plugging this into (13)

$$\left\{\pi_{S_1}(\partial_I(C_i)) \;\middle|\; I \in [n]^{2k}\right\} \subseteq \mathrm{span}_{\mathbb{F}} \left\{Q_A^B \cdot \mathbf{x^j} \;\middle|\; |A| = D-k, |\mathbf{j}| \le (t+1)k, B \subseteq [d]\right\}$$

We are now ready to bound $\mu_{k,\ell}(C_i)$. By linearity once more, we have

$$\begin{aligned}
\mathcal{V}_{k,\ell}(C_i) &= \left\{ \mathbf{x}^{\mathbf{i}} \cdot \pi_{S_1}(\partial_I(C_i)) \;\middle|\; I \in [n]^{2k}, |\mathbf{i}| \leq \ell \right\} \\
&\subseteq \mathrm{span}_{\mathbb{F}} \left\{ Q_A^B \cdot \mathbf{x}^{\mathbf{i}+\mathbf{j}} \;\middle|\; |A| = D-k, |\mathbf{j}| \leq (t+1)k, B \subseteq [d], |\mathbf{i}| \leq \ell \right\} \\
&= \mathrm{span}_{\mathbb{F}} \left\{ Q_A^B \cdot \mathbf{x}^{\mathbf{i}} \;\middle|\; |A| = D-k, |\mathbf{i}| \leq \ell + (t+1)k, B \subseteq [d] \right\}
\end{aligned}$$

Therefore, by the definition of $\mu_{k,\ell}$, we get

$$\begin{aligned}
\mu_{k,\ell}(C_i) &= \dim(\mathcal{V}_{k,\ell}(C_i)) \\
&\leq \left| \left\{ Q_A^B \cdot \mathbf{x}^{\mathbf{i}} \;\middle|\; |A| = D-k, |\mathbf{i}| \leq \ell + (t+1)k, B \subseteq [d] \right\} \right| \\
&\leq (\#\text{ of choices for } B) \cdot (\#\text{ of choices for } A) \cdot (\#\text{ of monomials of degee} \leq \ell + (t+1)k) \\
&= 2^d \cdot \binom{D}{k} \cdot \binom{N + \ell + (t+1)k}{\ell + (t+1)k}
\end{aligned}$$

This finishes the proof of Lemma 42.