Sums of Products of Polynomials in Few Variables: Lower Bounds and Polynomial Identity Testing

Mrinal Kumar[*1] **and Shubhangi Saraf**[†2]

1 **Department of Computer Science, Rutgers University, New Brunswick, USA**
mrinal.kumar@rutgers.edu

2 **Department of Computer Science and Department of Mathematics, Rutgers University, New Brunswick, USA**
shubhangi.saraf@gmail.com

Abstract

We study the complexity of representing polynomials as a sum of products of polynomials in few variables. More precisely, we study representations of the form $P = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij}$ such that each Q_{ij} is an arbitrary polynomial that depends on at most s variables.

We prove the following results.

- Over fields of characteristic zero, for every constant μ such that $0 \leq \mu < 1$, we give an explicit family of polynomials $\{P_N\}$, where P_N is of degree n in $N = n^{O(1)}$ variables, such that any representation of the above type for P_N with $s = N^{\mu}$ requires $Td \geq n^{\Omega(\sqrt{n})}$. This strengthens a recent result of Kayal and Saha [17] which showed similar lower bounds for the model of sums of products of linear forms in few variables. It is known that any asymptotic improvement in the exponent of the lower bounds (even for $s = \sqrt{n}$) would separate VP and VNP [17].
- We obtain a deterministic subexponential time blackbox polynomial identity testing (PIT) algorithm for circuits computed by the above model when T and the individual degree of each variable in P are at most $\log^{O(1)} N$ and $s \leq N^{\mu}$ for any constant $\mu < 1/2$. We get quasipolynomial running time when $s < \log^{O(1)} N$. The PIT algorithm is obtained by combining our lower bounds with the hardness-randomness tradeoffs developed in [6, 14]. To the best of our knowledge, this is the first nontrivial PIT algorithm for this model (even for the case $s = 2$), and the first nontrivial PIT algorithm obtained from lower bounds for small depth circuits.[1]

1998 ACM Subject Classification F.2.1 Numerical Algorithms and Problems, I.1.1 Expressions and Their Representation

Keywords and phrases arithmetic circuits, lower bounds, polynomial identity testing, hardness vs randomness

Digital Object Identifier 10.4230/LIPIcs.CCC.2016.35

1 Introduction

Arithmetic circuits are the most natural model of computation for a wide variety of algebraic problems such as matrix multiplication, computing fast fourier transforms etc. The problem

* Research supported in part by NSF grants CCF-1350572 and by Simons Graduate Fellowship.
† Research supported by NSF grant CCF-1350572.

[1] In a recent independent work, Forbes [7] does blackbox identity testing for another subclass of depth four circuits using shifted partial derivative based methods. To the best of our understanding, the results in these two papers are incomparable even though both rely on similar techniques.

31st Conference on Computational Complexity (CCC 2016).
Editor: Ran Raz; Article No. 35; pp. 35:1–35:29

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

of proving lower bounds for arithmetic circuits is one of the most fundamental and interesting problems in complexity theory. Proving superpolynomial lower bounds for general arithmetic circuits would resolve the VP versus VNP conjecture [34], the algebraic analog of the P vs NP conjecture. This is one of the holy grails of complexity theory and has received a lot of attention, since it is a more structured and potentially easier question to understand and analyse than the P vs NP problem .

The intimately related problem of polynomial identity testing (PIT) is the problem of testing if a polynomial, given as an arithmetic circuit is identically zero. In the setting where the algorithm cannot look inside the circuit, but only has access to evaluations of the circuit, the problem is referred to as blackbox PIT. There is a very simple randomized algorithm for this problem - simply evaluate the polynomial at a random point from a large enough domain. With very high probability, a nonzero polynomial will have a nonzero evaluation [30, 36]. It is a very important and fundamental question to derandomize the above algorithm. In a seminal work, Kabanets and Impagliazzo [14] showed that the problem of proving lower bounds for arithmetic circuits and the problem of derandomizing identity testing are essentially equivalent[2]!

These two problems have occupied a central position in complexity theory and despite much attention, our understanding of general arithmetic circuits is still very limited. Thus there has been a great deal of effort in understanding the complexity of restricted classes of arithmetic circuits in an attempt to obtain a better understanding of the general problem. Low depth arithmetic circuits in particular are one such well studied class.

Lower bounds for homogeneous low depth arithmetic circuits

The last few years have seen a tremendous amount of exciting progress on the problems of "depth reduction" of general arithmetic circuits to low depth arithmetic circuits, and of proving lower bounds for low depth arithmetic circuits. Using depth reduction techniques [35, 1, 20, 33] it was shown that $N^{\omega(\sqrt{n})}$ lower bounds (for polynomials in N variables and of degree n) for just homogeneous depth 4 arithmetic circuits of bottom fan-in $\sqrt{n}$ would suffice to separate VP from VNP and imply superpolynomial lower bounds for general arithmetic circuits. At the same time there was a very exciting line of works proving $N^{\Omega(\sqrt{n})}$ lower bounds for the same model of arithmetic circuits (and in fact for even the more general class of homogeneous depth 4 arithmetic circuits with no restriction on bottom fan-in) [11, 10, 19, 21, 15, 22].

Lower bounds for non-homogeneous low depth arithmetic circuits

Despite all this remarkable progress, and some very strong lower bounds for homogeneous low depth arithmetic circuits, in the nonhomogenous world much less is understood. Only mild lower bounds are known when we drop the condition of homogeneity, even for very simple classes of low depth arithmetic circuits. For depth 3 circuits over fields of characteristic 0, only quadratic lower bounds known [31, 32], and there has been no progress on this question in more than a decade now.

In a beautiful depth reduction result over fields of characteristic 0, Gupta et al [13] showed that $N^{\omega(\sqrt{n})}$ lower bounds (for polynomials in N variables and of degree n) for the class of non-homogeneous *depth 3* circuits would already separate VP from VNP. It was recently observed by Kayal and Saha [17] [3] that in fact it suffices to prove such lower bounds for depth 3 circuits with bottom fan-in $\sqrt{n}$.

---

[2] They non-trivially transferred such known tradeoffs from the boolean world to the arithmetic world[25].
[3] They attribute the observation to Ramprasad Saptharishi.

Till recently (in particular till the work of [17]), the best known lower bounds for depth 3 circuits even with bottom fan-in 2 were still just quadratic. In a very nice recent result, Kayal and Saha [17] showed an exponential lower bound for depth 3 circuits over fields of characteristic 0, whose bottom fan-in is at most N^{μ}, where N is the number of variables and $0 \leq \mu < 1$ is an arbitrary constant. More precisely, they prove the following.

▶ **Theorem 1.1** (Kayal-Saha [17]). *Let $\mathbb{F}$ be a field of characteristic zero. Then, for every constant $0 \leq \mu < 1$ there is a family $\{P_N\}$ of degree n polynomials in $N = n^{O_\mu(1)}$ variables over $\mathbb{F}$ in* VNP *such that any depth three circuit of bottom fan-in at most N^{μ} computing P_N has top fan-in at least $N^{\Omega_\mu(\sqrt{n})}$.*

Our Model

In this work, we consider the model of sums of products of polynomials in few variables. More formally, we consider representations of polynomials P (degree n in $N = n^{O(1)}$ variables) in the form

$$P = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij} \tag{1}$$

where each Q_{ij} is an arbitrary polynomial (of arbitrarily high degree) in at most s variables. We call this the model of $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuits.

Observe that the model is more general than that considered in [17]. The model in [17] corresponds to sums of products of *linear forms* in few variables. In our case, the Q_{ij} no longer have to be linear forms, but can be general polynomials of arbitrarily high degree. Prior to this work, even for the case when $s = 2$, there were no nontrivial lower bounds known for this model.

$\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuits for $s \geq 2$ can also be seen as a generalization of the model of sums of products of univariate polynomials (which corresponds to $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuits with $s = 1$), which has been very well studied in the arithmetic circuit complexity literature. Lower bounds for $\Sigma\Pi\left(\Sigma\Pi\right)^{[1]}$ circuits follow from works of Nisan [24] and Saxena [29]. Over the last few years, there have been some very nice results giving quasipolynomial time blackbox identity testers for $\Sigma\Pi\left(\Sigma\Pi\right)^{[1]}$ circuits [8, 9, 3]. $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuits can also be seen as a generalization of the widely studied model of diagonal circuits, since polynomials computable by diagonal circuits can be represented as a $\Sigma\Pi\left(\Sigma\Pi\right)^{[1]}$ circuit without much blow up in the size of the representation [29].

Although $\Sigma\Pi\left(\Sigma\Pi\right)^{[1]}$ circuits seem fairly well understood from the point of view of lower bounds and derandomization of polynomial identity testing, if one considers the model of sums of products of bivariate polynomials ($\Sigma\Pi\left(\Sigma\Pi\right)^{[2]}$ circuits), then our understanding changes completely. Although only seemingly a mild generalization of $\Sigma\Pi\left(\Sigma\Pi\right)^{[1]}$ circuits, the known proof techniques for lower bounds for $\Sigma\Pi\left(\Sigma\Pi\right)^{[1]}$ circuits (which were proved using *evaluation dimension* techniques of [24, 27]) seem to completely break down in this setting. In fact, Forbes [7] was able to confirm this, showing that there is a polynomial which is a sum of product of bivariates which has exponentially large evaluation dimension under all possible partitions of variables. Thus, studying this model seems like an interesting next step towards understanding non-homogeneous small depth algebraic computation. As far as we are aware there are also (not surprisingly) no nontrivial PIT results for the model. We are now ready to state our results.

1.1 Our results

Lower bounds

We show an exponential lower bound for the model of $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$, when s is at most N^{μ} for any constant $0 \leq \mu < 1$ (N is the number of variables). More precisely, we show the following.

▶ **Theorem 1.2.** *Let $\mathbb{F}$ be a field of characteristic zero and μ be any constant such that $0 \leq \mu < 1$. There exists a family $\{P_N\}$ of polynomials over $\mathbb{F}$ in* VNP*, where P_N is of degree n in $N = n^{O_\mu(1)}$ variables, such that for any representation of P_N of the form*

$$P_N = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij}$$

where each Q_{ij} is polynomial in at most $s = N^{\mu}$ variables, it must be true that

$$T \cdot d \geq n^{\Omega_\mu(\sqrt{n})}.$$

Given the depth reduction results of [13] and the observation mentioned earlier from [17], it is known that any asymptotic improvement in the exponent of the lower bound (even for $s = O(\sqrt{n})$) would imply VNP is different from VP.

As discussed in the introduction, even though this model seems a natural generalization of the model of sums of products of univariate polynomials, our lower bound technique is very different from those used in proving lower bounds for sums of products of univariates. Our lower bound proof is based on ideas developed in the course of investigating homogeneous depth four arithmetic circuits [15, 22].

Blackbox PIT

We also consider the problem of PIT for the model of $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuits. For general sums of products of even bivariate polynomials, this question seems quite difficult, and as of now we are not even able to obtain subexponential time PIT. However, as a consequence of our lower bounds and by suitably adapting hardness randomness tradeoffs for arithmetic circuits developed in [14] and [6], we are able to obtain PIT results in the setting where the top fan-in of the circuit is bounded, and when we have the promise that the circuit computes a polynomial of low individual degree.

Our understanding of blackbox PIT for depth four circuits is very limited, and the results known are in very restricted settings. Saraf and Volkovich [28] gave blackbox PIT algorithms for multilinear depth 4 circuits with bounded top fan-in. To the best of our knowledge, the idea in [28] does not extend to the case of non-multilinear depth 4 circuits, even when the individual degree of each of the variables is at most 2. Recently, Oliveira et al [5] gave a subexponential time blackbox PIT for all depth four multilinear circuits[4]. In the non-multilinear setting, Agrawal et al. [2] gave PIT algorithms for constant depth formulas in which the number of *occurences* of each variable is bounded. Without going into the technical details, we remark that the notion of *bounded occur* is a generalization of the well studied notion of bounded reads. The most closely related results to those in this paper that we are aware of are the recent papers of Gupta [12] and Mukhopadhyay [23], which give

[4] The running time increases with the size of the circuit, and in particular, it is subexponential time for polynomial sized depth four multilinear circuits.

blackbox PIT results for sums of products of low degree polynomials, where the top sum fan-in is bounded and the circuits satisfy certain algebraic geometric restrictions.

So, the question of getting PIT results for general depth four circuits (even with bounded top and bottom fan-in) remains wide open. For instance we still do not know any nontrivial PIT results for a sum of constant many products of degree 2 polynomials. Though we still don't know how to deal with this question, when we replace the polynomials of low degree with polynomials of few variables (but of arbitrarily large degree), then we are able to obtain quasipolynomial PIT results. There is one added caveat however, that the final polynomial computed needs to be of low individual degree (as seems necessary for PIT results obtained from the known hardness-randomness tradeoffs for bounded depth circuits [6]). We now formally state the theorem.

▶ **Theorem 1.3.** *Let c and μ be arbitrary constants such that $c > 0$ and $0 \leq \mu < 1/2$, and let $\mathbb{F}$ be a field of characteristic zero. Let $\mathcal{C}$ be the set of polynomials P in N variables and individual degree at most k over $\mathbb{F}$, with the property that P can be expressed as*

$$P = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij}$$

such that

1. $T < \log^c N$
2. $k < \log^c N$
3. $d < N^c$
4. *each Q_{ij} depends on at most N^μ variables*

Then, there exists a constant $\epsilon < 1$ dependent only on c and μ, such that there is a hitting set of size $\exp(N^\epsilon)$ for $\mathcal{C}$ which can be constructed in time $\exp(N^\epsilon)$.

Moreover, from our proof, it also follows that if each of polynomial Q_{ij} depends only on $\log^{O(1)} N$ variables, then both the size of the hitting set and the time to construct it, are upper bounded by a quasipolynomial function in N.

Independent work

In a simultaneous independent work, Kayal and Saha [18] employ very similar techniques and ideas to show an analog of Theorem 1.2 for the iterated matrix multiplication polynomial (an entry in the product of n generic matrices of dimension $\text{poly}(n) \times \text{poly}(n)$) when each of the polynomials Q_{ij} depends on at most $\sqrt{n}$ variables.

Organisation of the paper

We provide an overview of the proofs in Section 2. We describe some definitions and preliminaries in Section 3. We present the proof of the lower bound in Section 4. We describe the application to blackbox PIT in Section 5 and conclude with some open problems in Section 6.

2 Proof overview

In this section, we provide an overview of the main ideas in proofs of Theorem 1.2 and Theorem 1.3.

2.1 Overview of proof of Theorem 1.2

We restate Theorem 1.2 for the sake of clarity.

▶ **Theorem 1.2** (restated). *Let $\mathbb{F}$ be a field of characteristic zero and μ be any constant such that $0 \leq \mu < 1$. There exists a family $\{P_N\}$ of polynomials over $\mathbb{F}$ in* VNP*, where P_N is of degree n in $N = n^{O_\mu(1)}$ variables, such that for any representation of P_N of the form $P_N = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij}$ where each Q_{ij} is polynomial in at most $s = N^\mu$ variables, it must be true that*

$$T \cdot d \geq n^{\Omega_\mu(\sqrt{n})}.$$

The key difference between proving the above lower bound and the lower bounds for homogeneous depth four circuits is that the formal degree of the circuit in the above case could be much larger than the degree of the polynomial, which is n. In fact, even the fan-in of the product gates at level 2, that is d could be much larger than n. Therefore, a straightforward application of homogeneous depth four circuit lower bounds does not seem to work. Our proof is in two steps and at a high level follows the strategy of the lower bound for non-homogeneous depth three circuits with bounded bottom fan-in by Kayal and Saha [17] with some key differences.

- In the first step, we obtain another representation of P_N, as

$$P_N = \sum_{i=1}^{Td2^{O(\sqrt{n})}} \prod_{j=1}^{n} Q'_{ij}$$

  where every monomial in each of the Q'_{ij} has *support*[5] at most s, although each Q'_{ij} could now depend on all the variables. The key property that we have gained from this transformation is that the fan-in of the product gates at level two is bounded by n now, which is the degree of P_N. However, we have no bound on the degree of the Q'_{ij}. Moreover, we have blown up the top fan-in a bit, but we will be able to tolerate this loss if s is small.
- In the second step, the strategy can be seen in two stages. If μ was very small, say 0.001, then we could have taken advantage of the fact that in the representation obtained in the first step above, the product fan-in is at most n and the support of every monomial in each of the Q'_{ij} is small, to prove an upper bound on the dimension of the space of projected shifted partial derivatives of the above representation. Comparing this dimension with that of our hard polynomial gives us our lower bound. For larger values of μ, we use random restrictions to ensure that all the monomials of *large support* in Q'_{ij} are set to zero. At the end of such a procedure, we are back to the low support case. This step of the proof is closely along the lines of the proof of homogeneous depth four arithmetic circuit lower bounds in [15, 22] although in the present case, formal degree of the circuit could be as large as n^2, which is much larger than the degree of the polynomial P_N. For such large formal degrees, in general we do not even know lower bounds for non-homogeneous depth three circuits.

We would like to point out that the first step of the proof above is similar to the homogenization step in the proof of lower bounds for general depth three circuits with bounded bottom fan-in by Kayal and Saha [17]. The key difference is that while the circuit they obtain at the end of

[5] A monomial is said to have support support s if it depends on at most s distinct variables.

this step is a strictly homogeneous circuit of formal degree n, we are unable to get a similar structure. The complication stems from the fact that when Q_{ij} are not affine forms, they could contain monomials of varying degrees. In this case, it seems difficult to obtain a strict homogenization with a small blow up in size. We get around this deficiency by a more subtle analysis in the second step, where we show a lower bound for a circuit which has a formal degree much larger than the degree of the polynomial being computed, but has some added structure. This step critically uses that the fact that the product fan-in at level two of these circuits is at most n, and the support of every monomial in each of the Q'_{ij} is small.

2.2 Overview of proof of Theorem 1.3

We first restate Theorem 1.3.

▶ **Theorem 1.3** (restated). *Let c and μ be arbitrary constants such that $c > 0$ and $0 \leq \mu < 1/2$, and let $\mathbb{F}$ be a field of characteristic zero. Let $\mathcal{C}$ be the set of polynomials P in N variables and individual degree at most k over $\mathbb{F}$, with the property that P can be expressed as $P = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij}$ such that*

1. $T < \log^c N$
2. $k < \log^c N$
3. $d < N^c$
4. *each Q_{ij} depends on at most N^μ variables*

Then, there exists a constant $\epsilon < 1$ dependent only on c and μ, such that there is a hitting set of size $\exp(N^\epsilon)$ for $\mathcal{C}$ which can be constructed in time $\exp(N^\epsilon)$.

The construction of the hitting set is based on the well known idea of using hard functions for derandomization. Our goal is to reduce the number of variables from N to at most N^δ for some constant $\delta < 1$, while maintaining the zeroness/nonzeroness of the polynomial being tested [14, 6]. Once we have done this, we take a brute force hitting set of size $(\text{Degree} + 1)^{\text{Number of variables}}$ as given by Lemma 5.5. To reduce the number of variables, we use the framework introduced by Kabanets and Impagliazzo [14].

The key technical step of the proof is to show that for a non-zero polynomial P as defined above, if there exists a polynomial $f \in \mathbb{F}[X_1, X_2, \ldots, X_{i-1}, X_{i+1}, X_{i+2}, \ldots, X_N]$ such that $X_i - f$ divides P, then f can also be expressed as a sum of products of polynomials in few variables of reasonably small size. This step crucially uses a statement about complexity of roots of polynomials computed by low depth circuits from [6]. Therefore, if f is a polynomial which does not have a small representation as a sum of products of polynomials in *few* variables, then $X_i - f$ does not divide P. This observation guarantees that the construction of hitting sets from hard polynomials given by [14] works for this class of circuits.

3 Notation and Preliminaries

We now introduce some notation and preliminary notions that we use in the rest of the paper.

Computational model

In this work, we consider the model of sums of products of polynomials in few variables. More formally, we consider representations of polynomials P (degree n in $N = n^{O(1)}$ variables) in

the form

$$P = \sum_{i=1}^{T} \alpha_i \cdot \prod_{j=1}^{d} Q_{ij} \tag{2}$$

where each Q_{ij} is an arbitrary polynomial (of arbitrarily high degree) in at most s variables and each α_i is a field constant. We call this the model of $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuits. We use the quantity Td as a measure of the size of a $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuit. Without loss of generality, we can assume that the degree zero term in each of the Q_{ij} is either zero or one. If it is a non-zero constant other than 1, we can extract it out and absorb it in α_i. For each of the product gates, the fan-in could be different, but we can assume without loss of generality that all the product fan-ins are equal to d. Observe that the d could be much larger than the degree of the polynomial P. Throughout this paper, we will be working over a field of characteristic zero.

Some basic notations

1. For an integer i, we denote the set $\{1, 2, \ldots, i\}$ by $[i]$.
2. By $\overline{X}$, we mean the set of variables $\{X_1, X_2, \ldots, X_N\}$.
3. For a polynomial P and a positive integer i, we represent by $\mathsf{Hom}^i[P]$, the homogeneous component of P of degree equal to i. By $\mathsf{Hom}^{\leq i}[P]$ and $\mathsf{Hom}^{\geq i}[P]$, we represent the component of P of degree at most i and at least i respectively.
4. The support of a monomial α is the set of variables which appear with a non-zero exponent in α. We denote the size of the support of α by $\mathrm{Supp}(\alpha)$.
5. Throughout the paper, we say that a function $f(N)$ is subexponential in N if there exists a positive real number ϵ, such that $\epsilon < 1$ and for all N sufficiently large, $f(N) < \exp(N^{\epsilon})$.
6. We say that a function $f(N)$ is quasipolynomial in N if there exists a positive absolute constant c, such that for all N sufficiently large, $f(N) < \exp(\log^c N)$.
7. In this paper, we only consider layered arithmetic circuits and we will be counting levels from top to bottom, starting with the output gates being at level one.
8. By a $\Sigma\Pi\Sigma\wedge$ circuit, we refer to a depth four circuit with all the product gates at the lowest level being replaced by powering ($\wedge$) gates. Similarly, by a $\Sigma\Pi\Sigma \wedge \Sigma\Pi$ circuit, we mean a depth six circuit all of whose product gates at level four from the top are powering gates.

Hitting set

Let $\mathcal{C}$ be a set of polynomials in N variables over a field $\mathbb{F}$. Then, a set $\mathcal{H} \subseteq \mathbb{F}^N$ is said to be a *hitting set* for the class $\mathcal{C}$, if for every polynomial $P \in \mathcal{C}$ such that P is not the identically zero polynomial, there exists a $p \in \mathcal{H}$ such that $P(p) \neq 0$.

Elementary symmetric polynomials

For variables $\overline{X} = \{X_1, X_2, \ldots, X_N\}$ and any integer $0 \leq l \leq N$, the elementary symmetric polynomial of degree l on variables $\overline{X}$ is defined as

$$\mathsf{ESYM}_l(\overline{X}) = \sum_{S \subseteq [N], |S| = l} \prod_{j \in S} X_j \,.$$

Projected shifted partial derivatives

A key idea behind the recent progress on lower bounds is the notion of *shifted partial derivatives* introduced in [16]. In this paper, we use a variant of the measure, called projected shifted partial derivatives introduced in [15] and subsequently used in [22]. Although we never explicitly do any calculations with the measure in this paper, we provide a brief introduction to it below since the bounds are based on it.

For a polynomial P and a monomial γ, $\partial_\gamma(P)$ is the partial derivative of P with respect to γ. For every polynomial P and a set of monomials $\mathcal{M}$, $\partial_{\mathcal{M}}(P)$ is the set of partial derivatives of P with respect to monomials in $\mathcal{M}$. The space of $(\mathcal{M}, m)$-projected shifted partial derivatives of a polynomial P is defined below.

▶ **Definition 3.1** ($(\mathcal{M}, m)$-projected shifted partial derivatives)**.** For an N variate polynomial $P \in \mathbb{F}[X_1, X_2, \ldots, X_N]$, set of monomials $\mathcal{M}$ and a positive integer $m \geq 0$, the space of $(\mathcal{M}, m)$-projected shifted partial derivatives of P is defined as

$$\langle \partial_{\mathcal{M}}(P) \rangle_m \stackrel{def}{=} \mathbb{F}\text{-}span\{\sigma(\prod_{i \in S} X_i \cdot g) : g \in \partial_{\mathcal{M}}(P), S \subseteq [N], |S| = m\} \tag{3}$$

Here, $\sigma(P)$ of a polynomial P is the projection of P on the multilinear monomials in its support. The measure of complexity of a polynomial that we use in this paper, is the dimension of projected shifted partial derivative space of P with respect to some set of monomials $\mathcal{M}$ and a parameter m. Formally,

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

From the definitions, it is straight forward to see that the measure is subadditive.

▶ **Lemma 3.2** (Sub-additivity)**.** *Let P and Q be any two multivariate polynomials in $\mathbb{F}[X_1, X_2, \ldots, X_N]$. Let $\mathcal{M}$ be any set of monomials and m be any positive integer. Then, for all scalars α and β*

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

Approximations

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

▶ **Lemma 3.3** ([11])**.** *Let $a(n), f(n), g(n) : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be integer valued functions such that $(f + g) = o(a)$. Then,*

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

In the proofs in this paper, we use Lemma 3.3 only in situations where $(f+g)^2$ will be $O(a)$. In this case, the error term will be bounded by an absolute constant. So, up to constant factors, $\frac{(a+f)!}{(a-g)!} = a^{(f+g)}$. We use the symbol $\approx$ to indicate equality up to constant factors.

Complexity of coefficients and homogeneous components

We now summarise two simple lemmas which are useful for our proof. The first lemma summarises that given a circuit C for a polynomial $P \in \mathbb{F}[X1, X2, \ldots, X_N, Y]$ of degree at most d, for every $0 \leq i \leq d$, the coefficient of Y^i in P (when viewing P as a polynomial in $\mathbb{F}[X_1, X_2, \ldots, X_N][Y]$) can also be computed by a circuit of size not much larger than the size of C.

▶ **Lemma 3.4.** *Let $P \in \mathbb{F}[X_1, X_2, \ldots, X_N, Y]$ be a polynomial of degree at most d in Y over a field $\mathbb{F}$ of characteristic zero, such that P is computable by an arithmetic circuit C of size $|C|$. Let*

$$P = \sum_{i=0}^{d} Q_i(X_1, X_2, \ldots, X_N) \cdot Y^i$$

for polynomials $Q_i(X_1, X_2, \ldots, X_N) \in \mathbb{F}[X_1, X_2, \ldots, X_N]$. Then, for every i such that $0 \leq i \leq d$, the polynomial Q_i can be computed by an arithmetic circuit C' of size at most $|C| \cdot (d+1)$. Moreover, if the output gate of C is a $+$ gate, then the depth of C' is equal to the depth of C. Else, the depth of C' is at most 1 more than the depth of C.

Proof. We can view P as a univariate polynomial of degree at most d in Y with the coefficients coming from $\mathbb{F}(\overline{X})$. From the classical Lagrange interpolation, we know that the coefficient of Y^i in P can be written as an $\mathbb{F}(\overline{X})$ linear combination of the evaluations of P at $d+1$ distinct values of Y taken from $\mathbb{F}(\overline{X})$. In fact, more strongly, we can evaluate P at $d+1$ values of Y all chosen from $\mathbb{F}$ itself, in which case the constants in the linear combination are also from $\mathbb{F}$. So, Q_i can be computed by a circuit obtained from taking $d+1$ circuits each obtained from P by substituting Y by a scalar in $\mathbb{F}$, and taking their linear combination. Let this circuit be C'. Clearly the size of C' is at most $(d+1)$ times the size of C. If the output gate of C was an addition gate, then the outer addition for the linear combination can be absorbed into it, and the depth remains the same. Else, the depth increases by one. ◀

The second lemma stated below essentially says that the circuit complexity of homogeneous components of a polynomial is not much larger than the circuit complexity of the polynomial itself.

▶ **Lemma 3.5.** *Let P be a polynomial of degree at most d in N variables over a field $\mathbb{F}$ of characteristic zero, such that P is computable by an arithmetic circuit C of size $|C|$. Then, for every i such that $0 \leq i \leq d$, the homogeneous component of degree i of P can be computed by an arithmetic circuit C' of size at most $|C| \cdot (d+1)$. Moreover, if the output gate of C is a $+$ gate, then the depth of C' is equal to the depth of C. Else, the depth of C' is at most 1 more than the depth of C.*

Proof. Let $P'(t)$ be the polynomial obtained from P by replacing every variable X in P by $X \cdot t$ for a new variable t. We can view P' to be a univariate polynomial of degree at most d in t with the coefficients coming from $\mathbb{F}(\overline{X})$. Observe that for every i such that $0 \leq i \leq d$, the homogeneous component of P of degree equal to i is equal to the coefficient of t^i in P'. The proof now follows from Lemma 3.4. ◀

4 Proof of the lower bound

In this section, we give the proof of Theorem 1.2. We prove the lower bound for a variant of the well known family of Nisan-Wigderson polynomials defined by Kayal and Saha [17].

4.1 Target polynomials for the lower bound

We now define the family of polynomials of degree n in N variables for which we prove the lower bounds. The family is a variant of the Nisan-Wigderson polynomials which were introduced by Kayal et al in [19] in the context of lower bounds for homogeneous depth four circuits. The particular variant we use in the paper is due to Kayal and Saha [17].

The tradeoff between the number of variables N and the degree n will be parameterized by the parameter μ where $0 \leq \mu < 1$. First we need some parameters, which we define below.

1. $\delta = (1-\mu)/2$ is a positive real number such that $\mu + \delta < 1$.
2. $\gamma = \frac{2(\mu+\delta)+1}{1-\mu-\delta}$.
3. N is chosen such that N/n is a prime number between $n^{1+\gamma}$ and $2n^{1+\gamma}$. Such a prime number always exists from the Bertrand-Chebychev theorem. Without loss of generality, we pick the smallest one.
4. $\rho = (\mu+\delta)\frac{\log N}{\log n}$
5. $D = \frac{\gamma+\rho}{2(1+\gamma)} \cdot n$, where $D-1$ is the degree of the underlying univariate polynomials in the definition of $NW_{n,\mu}$.

Let ψ be the prime number equalling N/n. We are now ready to restate the definition of $NW_{n,\mu}$ from [17].

▶ **Definition 4.1** (Nisan-Wigderson Polynomials [17]). Let μ be a real number such that $0 \leq \mu < 1$. For a given μ and n, let N, D, ψ be as defined above. For the set of N variables $\{X_{ij} : i \in [n], j \in [\psi]\}$, we define the degree n homogeneous polynomial $NW_{n,\mu}$ as

$$NW_{n,\mu} = \sum_{\substack{f(z)\in\mathbb{F}_\psi[z] \\ deg(f)\leq D-1}} \prod_{i\in[n]} X_{if(i)} \,.$$

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

1. The number of monomials in $NW_{n,\mu}$ is exactly $\psi^D = n^{O(D)}$.
2. Each of the monomials in $NW_{n,\mu}$ is multilinear.
3. Each monomial corresponds to evaluations of a univariate polynomial of degree at most $D-1$ at all points of $\mathbb{F}_\psi$. Thus, any two distinct monomials agree in at most $D-1$ variables in their support.

We will also need the following lemma in our proof.

▶ **Lemma 4.2.** *Let μ be a non-negative real number less than 1. Given $q \in \mathbb{F}^N$, μ, n, we can evaluate the polynomial $NW_{n,\mu}$ at q in time $N^{O(n)}$.*

Proof. Given n and μ, we first find D, ψ as given by the choice of parameters. Once we have D, we iterate through every monomial α of degree n in the $\overline{X}$ variables which is supported on all the rows of the variable matrix and check if it is in the polynomial $NW_{n,\mu}$ by trying to find a univariate polynomial $f(z) \in \mathbb{F}_\psi[z]$ such that degree of f is at most $D-1$ and $\prod_{i\in[n]} X_{if(i)} = \alpha$. The interpolation takes only $\mathrm{Poly}(n)$ time, and the total number of monomials to try is at most N^n. So, we get the lemma. ◀

We now proceed with the proof as outlined in Section 2.1.

4.2 Reducing the product fan-in at level two

Let P be a homogeneous polynomial in N variables of degree n which has a $\Sigma\Pi\,(\Sigma\Pi)^{[s]}$ circuit of top fan-in T and product fan-in d at the second level. In other words, there exist polynomials $\{Q_{ij} : i \in [T], j \in [d]\}$ in at most s variables each, such that

$$P = \sum_{i=1}^{T} \alpha_i \cdot \prod_{j=1}^{d} Q_{ij} \,. \tag{4}$$

Recall that without loss of generality, we can assume that the constant term in each of the Q_{ij} is either 0 or 1. We have the following lemma.

▶ **Lemma 4.3.** *Let $\mathbb{F}$ be a field of characteristic zero. Let P be a homogeneous polynomial of degree n in N variables over $\mathbb{F}$ as defined above. For each i, $1 \leq i \leq T$ define the set*

$$S_i = \{j : 1 \leq j \leq d \text{ and } \mathsf{Hom}^0[Q_{ij}] = 1\}.$$

Then,

$$P = \sum_{i=1}^{T} \alpha_i \cdot \mathsf{Hom}^n \left[\prod_{j \notin S_i} Q_{ij} \times \sum_{l=0}^{n} \mathsf{ESYM}_l(\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : j \in S_i\}) \right]. \tag{5}$$

Proof. To prove the lemma, we will try to extract out the homogeneous part of degree n of each product gate $\prod_{j=1}^{d} Q_{ij}$. Together with the fact that the polynomial P is homogeneous of degree n, we get the lemma. Every Q_{ij} with a non-zero constant term can be written as $\mathsf{Hom}^{\geq 1}[Q_{ij}] + 1$, since the constant term in each Q_{ij} is either 0 or 1. Now,

$$\prod_{j=1}^{d} Q_{ij} = \prod_{j \notin S_i} Q_{ij} \times \prod_{j \in S_i} (\mathsf{Hom}^{\geq 1}[Q_{ij}] + 1). \tag{6}$$

Decomposing the product $\prod_{j \in S_i} (\mathsf{Hom}^{\geq 1}[Q_{ij}] + 1)$ further, we have

$$\prod_{j \in S_i} (\mathsf{Hom}^{\geq 1}[Q_{ij}] + 1] = \sum_{l=0}^{|S_i|} \sum_{U \subseteq S_i : |U| = l} \prod_{j \in U} \mathsf{Hom}^{\geq 1}[Q_{ij}]. \tag{7}$$

Now, observe that the degree of every monomial in $\prod_{j \in U} \mathsf{Hom}^{\geq 1}[Q_{ij}]$ is at least as large as the size of U. So, for every subset U of size larger than n, $\prod_{j \in U} \mathsf{Hom}^{\geq 1}[Q_{ij}]$ is a polynomial of degree strictly larger than n. Also, for any fixed l, the expression $\sum_{U \subseteq S_i : |U| = l} \prod_{j \in U} \mathsf{Hom}^{\geq 1}[Q_{ij}]$ is precisely the elementary symmetric polynomial of degree l in the set of variables $\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : j \in S_i\}$. Therefore,

$$\mathsf{Hom}^{\leq n} \left[\prod_{j \in S_i} (\mathsf{Hom}^{\geq 1}[Q_{ij}] + 1) \right] = \mathsf{Hom}^{\leq n} \left[\sum_{l=0}^{n} \mathsf{ESYM}_l(\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : j \in S_i\}) \right]. \tag{8}$$

Therefore,

$$\mathsf{Hom}^n \left[\prod_{j=1}^{d} Q_{ij} \right] = \mathsf{Hom}^n \left[\prod_{j \notin S_i} Q_{ij} \times \sum_{l=0}^{n} \mathsf{ESYM}_l(\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : j \in S_i\}) \right]. \tag{9}$$

Summing up for all i, we get the lemma. ◀

The lemma above has in some sense helped us locate the monomials of degree n in the circuit, which otherwise has a much higher formal degree. We now combine the above lemma with the well known fact that elementary symmetric polynomial of degree l in k variables can be computed by homogeneous $\Sigma\Pi\Sigma\wedge$ circuits of size at most $k2^{O(\sqrt{l})}$ to obtain a $\Sigma\Pi\Sigma \wedge \Sigma\Pi$ circut C' such that the fan-in of the product gates at level two is at most n. We use the following theorem (Theorem 5.2) by Shpilka and Wigderson [31].

▶ **Theorem 4.4** (Shpilka-Wigderson [31])**.** *For every set of variables $\{Y_1, Y_2, \ldots, Y_m\}$ and a positive integer l, $\mathsf{ESYM}_l(\{Y_1, Y_2, \ldots, Y_m\})$ can be computed by a homogeneous $\Sigma\Pi\Sigma\wedge$ circuit of size $m2^{O(\sqrt{l})}$.*

We now prove the following lemma.

▶ **Lemma 4.5.** *Let $\mathbb{F}$ be a field of characteristic zero. Let P be a polynomial of degree n in N variables over $\mathbb{F}$ which is computable by an $\Sigma\Pi(\Sigma\Pi)^{[s]}$ circuit C of top fan-in T and the degree of product gates at level two being d. So, P can be represented as*

$$P = \sum_{i=1}^{T} \alpha_i \cdot \prod_{j=1}^{d} Q_{ij} .$$

Then, P can be represented as the homogeneous component of degree n of a polynomial computed by a $\Sigma\Pi\Sigma \wedge \Sigma\Pi$ circuit C'' with the following properties :

1. *The inputs to the $\wedge$ gates are the polynomials $\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : 1 \leq i \leq T, 1 \leq j \leq d\}$*
2. *The fan-in of the $\times$ gates at the second level from the top is at most n*
3. *The top fan-in of C'' is at most $Tdn2^{O(\sqrt{n})}$.*

Proof. From Lemma 4.3, we know that for the set S_i defined as

$$S_i = \{j : 1 \leq j \leq d \text{ and } \mathsf{Hom}^0[Q_{ij}] = 1\}$$

the polynomial P can be written as

$$P = \sum_{i=1}^{T} \alpha_i \cdot \mathsf{Hom}^n \left[\prod_{j \notin S_i} Q_{ij} \times \sum_{l=0}^{n} \mathsf{ESYM}_l(\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : j \in S_i\}) \right]$$

which is the same as

$$P = \mathsf{Hom}^n \left[\sum_{i=1}^{T} \alpha_i \cdot \prod_{j \notin S_i} Q_{ij} \times \sum_{l=0}^{n} \mathsf{ESYM}_l(\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : j \in S_i\}) \right] .$$

Observe that the polynomial $\prod_{j \notin S_i} Q_{ij}$ has degree at least $d - |S_i|$. We remark that if $d - |S_i|$ is larger than n, then such product gates do not contribute anything to the degree n component of the polynomial and hence can be discarded without loss of generality; hence we assume $n - (d - |S_i|) > 0$. So, we could confine the inner sum from $l = 0$ to $l = n - (d - |S_i|)$, and still preserve the degree n part of the polynomial, which is what we are interested in. From Theorem 4.4, we know that for every $0 \leq l \leq n$, we can compute the polynomial $\mathsf{ESYM}_l(\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : j \in S_i\})$ by a $\Sigma\Pi\Sigma\wedge$ circuit of top fan-in at most $d \times 2^{O(\sqrt{l})}$ which takes as input the polynomials $\{\mathsf{Hom}^{\geq 1}(Q_{ij}) : 1 \leq j \leq d\}$. From the homogeneity of the circuits given by Theorem 4.4, it follows that the product gates at level two of these circuits have fan-in at most the degree of polynomial they compute, which is at most $n - (d - |S_i|)$. So, it follows that the polynomial

$$\tilde{P} = \left(\sum_{i=1}^{T} \alpha_i \cdot \prod_{j \notin S_i} Q_{ij} \times \sum_{l=0}^{n-(d-|S_i|)} \mathsf{ESYM}_l(\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : j \in S_i\}) \right)$$

can be computed by a $\Sigma\Pi\Sigma \wedge \Sigma\Pi$ circuit, with top fan-in at most $Tdn \cdot 2^{O(\sqrt{n})}$, which satisfies the conditions in the lemma. ◀

Finally, given the circuit C'' constructed above, we can construct a circuit which computes the polynomial P as given by Lemma 3.5. For this, observe that the monomials of degree strictly larger than n in any of the Q_{ij} do not contribute to degree n part of $\tilde{P}$. So, we can

drop them, while still preserving the degree n part of $\tilde{P}$. Therefore, the degree of $\tilde{P}$ can be upper bounded by n^2d. We can recover the degree n part of $\tilde{P}$ by interpolation which blows up the top fan-in by a factor of at most n^2d.

In this process, the fan-in of the product gates at level two remains unchanged. Strictly speaking, inputs to the powering gate $\wedge$ at level four may no longer be the polynomials $\mathsf{Hom}^{\geq 1}[Q_{ij}]$, since in the process of interpolation, we replaced every variable X_i by $X_i.t$ in $\tilde{P}$ and looked at the resulting polynomial $\tilde{P}'$ as a univariate polynomial in t over the function field $\mathbb{F}(\overline{X})$. We then evaluated $\tilde{P}'$ at sufficiently many values of $t \in \mathbb{F}$ and then took their $\mathbb{F}$ linear combination. So, each of the polynomials $\mathsf{Hom}^{\geq 1}[Q_{ij}]$ gives rise to many other polynomials, one each for different values of t. We will call them the *siblings* of $\mathsf{Hom}^{\geq 1}[Q_{ij}]$. The key observation for our proof is that the set of variables in the siblings of $\mathsf{Hom}^{\geq 1}[Q_{ij}]$ is the same as the set of variables in $\mathsf{Hom}^{\geq 1}[Q_{ij}]$. From the lemma and the discussion above, we have the following corollary.

▶ **Corollary 4.6.** *Let $\mathbb{F}$ be a field of characteristic zero. Let P be a polynomial of degree n in N variables over $\mathbb{F}$ which is computable by an $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuit C of top fan-in T and the degree of product gates at level two being d. So, P can be represented as*

$$P = \sum_{i=1}^{T} \alpha_i \cdot \prod_{j=1}^{d} Q_{ij}\,.$$

Then, P can be computed by a $\Sigma\Pi\Sigma \wedge \Sigma\Pi$ circuit C'' with the following properties :

1. *The inputs to the $\wedge$ gates are the siblings of polynomials $\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : 1 \leq i \leq T, 1 \leq j \leq d\}$*
2. *The fan-in of the $\times$ gates at the second level from the top is at most n*
3. *The top fan-in of C'' is at most $Td^2n^3 2^{O(\sqrt{n})}$.*

4.3 Random Restrictions

From the definition, it follows that the total number of variables in $NW_{n,\mu}$ is N. Let the set of all these variables be $\mathcal{V}$. We now define our random restriction procedure by defining a distribution $\mathcal{D}$ over subsets $V \subset \mathcal{V}$. The random restriction procedure will sample $V \leftarrow \mathcal{D}$ and then keep only those variables "alive" that come from V and set the rest to zero. We will denote the restriction of the polynomial obtained by such a restriction as $NW_{n,\mu}|_V$. Observe that a random restriction also results in a distribution over all circuits computing the polynomial $NW_{n,\mu}$. We denote by $C|_V$ the restriction of a circuit C obtained by setting every input gate in C which is labelled by a variable outside V to 0.

The distribution $\mathcal{D}_p$: Each variable in $\mathcal{V}$ is independently kept alive with a probability p. We will choose the value of p based on the parameter μ.

4.4 Analysing the circuit under random restrictions

Let C be a $\Sigma\Pi\left(\Sigma\Pi\right)^{[N^\mu]}$ circuit computing the polynomial $NW_{n,\mu}$. Let the top fan-in of C be T and the product fan-in at the second level be d. So, we have the following expression:

$$NW_{n,\mu} = \sum_{i=1}^{T} \alpha_i \cdot \prod_{j=1}^{d} Q_{ij}$$

where each Q_{ij} depends on at most N^μ variables.

Recall that from the choice of parameters $\delta = (1-\mu)/2$. Let s be a parameter, which we later set such that $s = \Theta(\sqrt{n})$. If $T \cdot d \geq N^{\frac{\delta}{4}s}$, then we already have the desired lower bound of $n^{\Omega(\sqrt{n})}$ on the size of C and we are done. Therefore, for the rest of this discussion, we will assume that $T \cdot d \leq N^{\frac{\delta}{4}s}$. We now apply the transformation to C given by Corollary 4.6 to obtain a $\Sigma\Pi\Sigma \wedge \Sigma\Pi$ circuit C'', which has the following properties:

1. The inputs to the $\wedge$ gates are the siblings of polynomials $\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : 1 \leq i \leq T, 1 \leq j \leq d\}$
2. The fan-in of the $\times$ gates at the second level from the top is at most n
3. The top fan-in of C'' is at most $Td^2n^3 2^{O(\sqrt{n})}$.

We now analyse the effect of the random restrictions on the circuit C''. We will choose a parameter $p = N^{-\mu-\delta}$ and keep every variable alive with a probability p. The circuit C'' can be represented as

$$C'' = \sum_u \prod_v D_{uv} \,.$$

Here, each D_{uv} is a sum of powers of the siblings of $\mathsf{Hom}^{\geq 1}[Q_{ij}]$. Our goal is to argue that under random restrictions, all the monomials in each of the D_{uv} are of small support (support at most s).

For any polynomial P in N^μ variables and any integers t, t_0 such that $t_0 < t$, observe that P^t can be written as

$$P^t = P_0 + \sum_\alpha \alpha \cdot P_\alpha$$

where P_0 is the part of P consisting of monomials of support strictly less than t_0. The inner sum is over all multilinear monomials α of support equal to t_0. Such a decomposition may not be unique, but for this application, it would suffice to work with any one such decomposition. The number of such monomials α is at most $\binom{N^\mu}{t_0}$. The probability that one such monomial survives the random restriction procedure is equal to p^{t_0}. So, the expected number of such multilinear monomials α surviving the random restriction procedure is at most $\binom{N^\mu}{t_0} \cdot p^{t_0}$. The crucial observation is that if no such monomials survive, then only the monomials in P_0 survive, all of which have support at most $t_0 - 1$.

Now, observe that each of the D_{uv} are a sum of powers of the siblings of polynomials in the set $\{\mathsf{Hom}^{\geq 1}[Q_{ij}] : 1 \leq i \leq T, 1 \leq j \leq d\}$. Define $\mathcal{B}$ to be the set of all multilinear monomials of support equal to s, supported entirely on variables in any of the polynomials Q_{ij} for some $1 \leq i \leq T, 1 \leq j \leq d$. From the discussion in the paragraph above, the following observation follows.

▶ **Observation 4.7.** *Let the polynomials D_{uv}, Q_{ij} and the set $\mathcal{B}$ be as defined above. Then,*

- $|\mathcal{B}| \leq T \cdot d \cdot \binom{N^\mu}{s}$
- *If none of the monomials in $\mathcal{B}$ survive under some random restrictions, then each of the polynomials D'_{uv} obtained as a restriction of D_{uv} has all monomials of support at most s.*

Proof. The bound on the size trivially follows from the fact that each of the Q_{ij} depends on at most N^μ variables. For the second item, observe that each of the D_{uv} is a sum of powers of siblings of the $\mathsf{Hom}^{\geq 1}[Q_{ij}]$ and all the siblings are supported on the same set of variables. If all the monomials in the set $\mathcal{B}$ are set to zero, then the surviving monomials in any power of any of the siblings of $\mathsf{Hom}^{\geq 1}[Q_{ij}]$ has support at most s. ◀

We now estimate the probability that at least one of the monomials in the set $\mathcal{B}$ survives the random restriction procedure. We have the following lemma.

▶ **Lemma 4.8.** *Let δ be a positive real number such that $\delta = (1-\mu)/2$ and let $p = N^{-\mu-\delta}$. Then*

$$Pr_{V \leftarrow \mathcal{D}_p}\left[|\mathcal{B}|_V| \geq 1\right] \leq N^{-3/4\cdot\delta\cdot s}\,.$$

Proof. We know that

$$|\mathcal{B}| \leq T \cdot d \cdot \binom{N^{\mu}}{s}$$

and the probability that any fixed monomial in $\mathcal{B}$ survives the random restriction procedure is at most p^s. So

$$\mathbb{E}_{V \leftarrow \mathcal{D}_p}[|\mathcal{B}_V|] \leq T \cdot d \cdot \binom{N^{\mu}}{s} \cdot p^s\,.$$

Now, observing that the value of $T \cdot d$ is at most $N^{\frac{\delta}{4}s}$ and $p = N^{-\mu-\delta}$, the expected value is at most

$$N^{\frac{\delta}{4}s}\binom{N^{\mu}}{s} \cdot N^{-(\mu+\delta)s} \leq N^{-3/4\cdot\delta\cdot s}\,.$$

The lemma then follows by Markov's inequality. ◀

As a corollary of Lemma 4.8 and Observation 4.7, we get the following lemma.

▶ **Lemma 4.9.** *Let δ be a positive real number such that $\delta = (1-\mu)/2$ and let $p = N^{-\mu-\delta}$. Then with probability at least $1 - N^{-3/4\cdot\delta\cdot s}$ over random restrictions $V \leftarrow \mathcal{D}_p$, the polynomial computed by the circuit $C''|_V$ can be written as $\sum_{u=1}^{T'}\prod_{v=1}^{n} D'_{uv}$, where each of the monomials in each of the polynomials D'_{uv} has support at most s.*

4.5 Upper bound on the complexity of C

In order to upper bound the dimension of the projected shifted partial derivatives (under random restrictions) of the $\Sigma\Pi\,(\Sigma\Pi)^{[s]}$ circuit C, Corollary 4.6 implies that it suffices to upper bound the dimension of the space of projected shifted partial derivatives of the $\Sigma\Pi\Sigma \wedge \Sigma\Pi$ circuit C'' given by Corollary 4.6. In some sense, C'' is more structured than C and this lets us prove a better upper bound.

Recall that we are under the assumption that for the circuit C, the product of the top fan-in and the product fan-in at level two is at most $N^{\frac{\delta}{4}\cdot s}$, else we are already done. From Lemma 4.9, we know that with a high probability, under random restrictions, we are left with a circuit of the form $\sum_{u=1}^{T'}\prod_{v=1}^{n} D'_{uv}$ where each of the monomials in each of the polynomials D'_{uv} has support at most s. The upper bound on the complexity of the projected shifted partial derivatives of $\sum_{u=1}^{T'}\prod_{v=1}^{n} D'_{uv}$ then just follows from the upper bound for homogeneous depth four circuits of bounded bottom support proved in [15, 22]. We restate the bound from [22].

▶ **Lemma 4.10.** *Let C be a depth 4 circuit with the fan-in or product gates at level two bounded by n, the bottom support bounded by s and computing a polynomial in N variables. Let $\mathcal{M}$ be a set of monomials of degree equal to r and let m be a positive integer. Then,*

$$\Phi_{\mathcal{M},m}(C) \leq \textit{Top fan-in}(C)\binom{n+r}{r}\binom{N}{m+rs}$$

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

The upper bound for $\Sigma\Pi(\Sigma\Pi)^{[N^\mu]}$ circuits, follows easily form the above lemma after random restrictions, and we formalize this in the lemma below.

▶ **Lemma 4.11.** *Let μ be a positive real number such that $0 \leq \mu < 1$. Let $\delta = (1-\mu)/2$ and let $p = N^{-\mu-\delta}$ and let $\mathbb{F}$ be a field of characteristic zero. Let P be a polynomial of degree n in N variables over $\mathbb{F}$ which is computed by an $\Sigma\Pi(\Sigma\Pi)^{[N^\mu]}$ circuit C of top fan-in T and degree of product gates at level two at most d, i.e P can represented as*

$$P = \sum_{i=1}^{T} \alpha_i \cdot \prod_{j=1}^{d} Q_{ij}$$

where α_i are field constants. Let m and r be positive integers satisfying $m + rs \leq N/2$ and $\mathcal{M}$ be any subset of multilinear monomials of degree equal to r. If $Td \leq N^{\frac{s\cdot\delta}{4}}$, then with probability at least $1 - N^{-3/4\cdot\delta\cdot s}$ over random restrictions $V \leftarrow \mathcal{D}_p$,

$$\Phi_{\mathcal{M},m}(C|_V) \leq Td^2n^3 \cdot rs \cdot 2^{O(\sqrt{n})} \cdot \binom{N}{m+rs} \cdot \binom{n+r}{r}.$$

Proof. The lemma follows immediately from Corollary 4.6, Lemma 4.9 and Lemma 4.10. ◀

4.6 Nisan-Wigderson polynomial under random restrictions

To complete the proof of Theorem 1.2, we need a lower bound on the dimension of the space of projected shifted partial derivatives of the polynomial $NW_{n,\mu}$, under random restrictions. To this end, we will use the lower bound proved by Kayal and Saha [17]. We first enumerate our choice of parameters. Recall that $\delta = (1-\mu)/2$ is a positive real number.

1. $\gamma = \frac{2(\mu+\delta)+1}{1-\mu-\delta}$
2. N is such that N/n is set equal to the smallest prime number between $n^{1+\gamma}$ and $2n^{1+\gamma}$.
3. $\rho = (\mu+\delta)\frac{\log N}{\log n}$
4. $D = \frac{\gamma+\rho}{2(1+\gamma)} \cdot n$, where $D-1$ is the degree of the underlying univariate polynomials in the definition of $NW_{n,\mu}$.
5. r, s which are the order of derivative and the bound on bottom support of the circuit after random restrictions respectively, are chosen such that $r = \epsilon_1 \cdot \sqrt{n}, s = \epsilon_2 \cdot \sqrt{n}$. Here, ϵ_1 and ϵ_2 are small enough positive real numbers satisfying $\epsilon_1 \cdot \epsilon_2 = 0.001n$.
6. $m = \frac{N}{2}(1 - r\frac{\ln n}{n})$ is the degree of the shifts.
7. $p = N^{-(\mu+\delta)}$ is the probability with which each variable is independently kept alive.
8. $\mathcal{M}$ is the set of all multilinear monomials of degree r. We take partial derivatives with respect to monomials in this set.

We are now ready to state the lower bound on the dimension of projected shifted partial derivatives as in [17].

▶ **Lemma 4.12** (Kayal-Saha [17]). *Let $NW_{n,\mu}$ be Nisan-Wigderson polynomials as defined in Definition 4.1. Let $\mathbb{F}$ be any field of characteristic zero. Then, for the choice of parameters defined above*

$$\Phi_{\mathcal{M},m}(NW_{n,\mu}|_V) \geq \frac{1}{n^{O(1)}} min\left(\frac{p^r}{4^r} \cdot \binom{N}{r} \cdot \binom{N}{m}, \binom{N}{m+n-r}\right)$$

with probability at least $1 - \frac{1}{n^{\theta(1)}}$ over random restrictions $V \leftarrow \mathcal{D}_p$.

4.7 Wrapping up the proof of Theorem 1.2

From Lemma 4.12 and Lemma 4.9, we know that with a non-zero probability over the random restrictions V from the distribution $\mathcal{D}_p$, the following two conditions hold.

1. $\Phi_{\mathcal{M},m}(NW_{n,\mu}|_V) \geq \frac{1}{n^{O(1)}} \min\left(\frac{p^r}{4^r} \cdot \binom{N}{r} \cdot \binom{N}{m}, \binom{N}{m+n-r}\right)$.
2. $\Phi_{\mathcal{M},m}(C|_V) \leq Td^2n^3 \cdot rs \cdot 2^{O(\sqrt{n})} \cdot \binom{N}{m+rs} \cdot \binom{n+r}{r}$.

If C computed the polynomial $NW_{n,\mu}$, then

$$Td^2n^3 \cdot rs \geq \frac{\frac{1}{n^{O(1)}} \min\left(\frac{p^r}{4^r} \cdot \binom{N}{r} \cdot \binom{N}{m}, \binom{N}{m+n-r}\right)}{2^{O(\sqrt{n})} \cdot \binom{N}{m+rs} \cdot \binom{n+r}{r}} .$$

From the calculations in Appendix A, it follows that for our choice of parameters, the ratio is at least $\exp(\sqrt{n}\log n)$. So, we have the following theorem.

▶ **Theorem 4.13.** *Let μ be an absolute constant such that $0 \geq \mu < 1$ and $\mathbb{F}$ be a field of characteristic zero. For $1 \leq i \leq T$ and $1 \leq j \leq d$, if there exist polynomials Q_{ij}, each dependent on only $s = N^{\mu}$ variables, such that*

$$NW_{n,\mu} = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij} .$$

Then

$$T \cdot d \geq n^{\Omega_{\mu}(\sqrt{n})} .$$

As a remark, we mention here that the lower bound above also holds for any translation $NW_{n,\mu}(\overline{X} + \overline{a})$ of the polynomial $NW_{n,\mu}(\overline{X})$. This is because the highest degree term of $NW_{n,\mu}(\overline{X} + \overline{a})$ equals the polynomial $NW_{n,\mu}(\overline{X})$ and from Lemma 3.5, the homogeneous components of a polynomial computable by small sized $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuits also have small sized $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuits. We leave the details to the interested reader.

5 Application to polynomial identity testing

In this section, we prove Theorem 1.3. We are interested in identity testing for $\Sigma\Pi\left(\Sigma\Pi\right)^{[s]}$ circuits, i.e for polynomials in N variables $\{X_1, X_2, \ldots, X_N\}$ which can be expressed in the form

$$P = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij}$$

such that

1. The individual degree in P of every variable is at most k
2. Each Q_{ij} depends on at most s variables

For the case of this application, we will think of k, T being polynomial in $(\log N)$ and s being $N^{1/2-\epsilon}$ for a positive constant ϵ. Observe that the bound on individual degree lets us upper bound the total degree of the polynomials by Nk.

We describe the construction of the hitting set in Section 5.2 and prove its correctness in Section 5.3. We go over some preliminaries that we need in our proof in the next section.

5.1 Some preliminaries

In the following lemma, we prove some properties of the model of $\Sigma\Pi(\Sigma\Pi)^{[s]}$ circuits, which will be useful in the proof of the identity testing result.

▶ **Lemma 5.1.** *Let $\mathbb{F}$ be a field of characteristic zero. Let P be a non-zero polynomial in N variables and individual degree at most k over $\mathbb{F}$, which is computed by a $\Sigma\Pi(\Sigma\Pi)^{[s]}$ circuit C of top fan-in T and product fan-in d at level two, i.e P can be expressed as*

$$P = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij}$$

such that for each $i \in [T]$ and $j \in [d]$, Q_{ij} depends on at most s variables. Then, the following are true.

1. *For every variable y and integer $1 \leq j \leq k$, $\frac{\partial^j P}{\partial y^j}$ can be computed by a circuit of the form*

$$\frac{\partial^j P}{\partial y^j} = \sum_{i=1}^{T'} \prod_{j=1}^{d} Q'_{ij}$$

   where $T' \leq T \cdot (k+1)^2$ and each of the polynomials Q'_{ij} depends on at most s variables.
2. *For any $a \in \mathbb{F}^N$, $P(\overline{X} + \overline{a})$ can be computed by a circuit of the form*

$$P(\overline{X} + \overline{a}) = \sum_{i=1}^{T} \prod_{j=1}^{d} Q''_{ij}$$

   where each of the polynomials Q''_{ij} depends on at most s variables.

Proof. The proof of the second item is immediate from the definitions. The only thing that changes due to a translation is the number of monomials in the Q_{ij}. The number of variables that each Q_{ij} depends on remains unchanged, and so does the fan-in of the top sum gate and the product gates at level two.

We now prove the first item. Let the set of variables in P be $\overline{X} = \overline{X'} \cup \{y\}$ where X' is of size $N-1$. Since the individual degree of P is at most k, we can write $P = \sum_{i=0}^{k} C_i(\overline{X'}) \cdot y^i$. Here, $C_i(\overline{X'})$ are polynomials only in the X' variables and are the coefficient of y^i, when viewing P as an element of $\mathbb{F}[\overline{X'}][y]$. Now, for every $0 \leq i \leq k$, we can compute each of C_i by a $\Sigma\Pi(\Sigma\Pi)^{[s]}$ circuit with top fan-in at most $T \cdot (k+1)$ by interpolation as given by Lemma 3.4. All the partial derivatives of P with respect to y are linear combinations of the terms of the form $C_{j_1} \cdot y^{j_2}$. And so, the result follows. ◀

We will also need the following simple fact about polynomials.

▶ **Lemma 5.2.** *Let $\mathbb{F}$ be a field of characteristic zero. Let $R \in \mathbb{F}[y]$ be a non-zero polynomial of degree at most t over the field $\mathbb{F}$. Then, for every $a \in \mathbb{F}$ such that $R(a) = 0$, there exists a j such that $0 \leq j \leq t-1$ and $\frac{\partial^j R}{\partial y^j}(a) = 0$ and $\frac{\partial^{j+1} R}{\partial y^{j+1}}(a) \neq 0$.*

Proof. Let the degree of R in y be equal to t'. This means that the coefficient of highest degree term $y^{t'}$ in R is non-zero. Let us call the coefficient of $y^{t'}$ in $R(y)$ as $C_{t'}$. We know that $C_{t'}$ is nonzero. Consider $j = t' - 1$. The lemma immediately follows. ◀

We will crucially use the following result of Dvir, Shpilka, Yehudayoff [6] in the analysis of the hitting set constructed in this paper.

▶ **Lemma 5.3** (Dvir, Shpilka, Yehudayoff [6]). *For a field $\mathbb{F}$, let $P \in \mathbb{F}[X_1, X_2, \ldots, X_N, Y]$ be a non-zero polynomial of degree at most k in Y. Let $f \in \mathbb{F}[X_1, X_2, \ldots, X_N]$ be a polynomial such that $P(X_1, X_2, \ldots, X_N, f) = 0$ and $\frac{\partial P}{\partial Y}(0, 0, \ldots, 0, f(0, 0, \ldots, 0)) \neq 0$. Let*

$$P = \sum_{i=0}^{k} C_i(X_1, X_2, \ldots, X_N) \cdot y^i \,.$$

Then, for every $t \geq 0$, there exists a polynomial $R_t \in \mathbb{F}[Z_1, Z_2, \ldots, Z_{k+1}]$ of degree at most t such that

$$\mathsf{Hom}^{\leq t}[f(X_1, X_2, \ldots, X_N)] = \mathsf{Hom}^{\leq t}[R_t(C_0, C_1, \ldots, C_k)] \,.$$

A key technical idea in the proof will be the notion of Nisan-Wigderson designs introduced in [26]. We will use the following lemma.

▶ **Lemma 5.4** (Nisan-Wigderson [26]). *For every $a, b \in \mathbb{N}$, $b < 2^a$, there exists a family of sets $S_1, S_2, \ldots, S_b \subseteq \{1, 2, \ldots, l\}$ such that*
1. *$l \in O(a^2 / \log b)$*
2. *for all i, $|S_i| = a$*
3. *for all $i \neq j$, $|S_i \cap S_j| \leq \log b$*

Moreover, such a set family can be constructed in time polynomial in b and 2^l.

We will also use the following lemma of Alon [4] very crucially in our proof.

▶ **Lemma 5.5** (Combinatorial Nullstellensatz [4]). *Let P be a non-zero polynomial of individual degree at most d in N variables over a large enough field $\mathbb{F}$. Let S be an arbitrary subset of $\mathbb{F}$ of size $d + 1$. Then, there exists a point p in S^N such that $P(p) \neq 0$.*

5.2 Blackbox PIT for $\Sigma\Pi\,(\Sigma\Pi)^{[s]}$ circuits

In this section, we prove the following theorem.

▶ **Theorem 5.6.** *Let c and μ be arbitrary constants such that $c > 0$ and $0 \leq \mu < 1/2$, and let $\mathbb{F}$ be a field of characteristic zero. Let $\mathcal{C}$ be the set of polynomials P in N variables and individual degree at most k over $\mathbb{F}$, with the property that P can be expressed as*

$$P = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij}$$

such that
1. *$T < \log^c N$*
2. *$k < \log^c N$*
3. *$d < N^c$*
4. *each Q_{ij} depends on at most N^μ variables*

Then, there exists a constant $\epsilon < 1$ dependent only on c and μ, such that there is a hitting set of size $\exp(N^\epsilon)$ for $\mathcal{C}$ which can be constructed in time $\exp(N^\epsilon)$.

From our proof, it also follows that if each of polynomial Q_{ij} depends only on $\log^{O(1)} N$ variables, then both the size of the hitting set and the time to construct it, are upper bounded by a quasipolynomial function in N. In the rest of the section, we prove Theorem 5.6. We start by describing the construction of the hitting set $\mathcal{H}$.

5.2.1 Construction of hitting sets for $\Sigma\Pi(\Sigma\Pi)^{[N^\mu]}$ circuits for $0 \leq \mu < 1/2$

Given μ such that $0 \leq \mu < 1/2$, we pick the parameter μ' such that $0 < \mu' < 1$ and $\frac{2\mu}{\mu'}$ is a positive constant strictly smaller than 1. We construct a family of Nisan-Wigderson designs as described in Lemma 5.4 with the following parameters:

1. b, the number of sets is set equal to N.
2. a, the size of each of the sets S_i is set equal to $N^{\frac{\mu}{\mu'}} \log^{\frac{1}{\mu'}} N$.
3. l, the size of the universe is chosen large enough in order to satisfy the hypothesis of Lemma 5.4. From Lemma 5.4, it follows that we can pick l which is not too large ($l \in O(a^2/\log b)$). For the above chosen values of a, b, there is a choice of l such that l is at most $N^{\frac{2\mu}{\mu'}} \log^{\frac{2}{\mu'}-1} N$.

Recall that our goal is to construct a hitting set for $\Sigma\Pi(\Sigma\Pi)^{[N^\mu]}$ circuits. Observe that the choice of parameters l, a, b satisfy the hypothesis of Lemma 5.4. So, we get a collection of N subsets $S_1, S_2, \ldots, S_N$ of $\{1, 2, 3, \ldots, l\}$ satisfying

1. for all $1 \leq i \leq N$, $|S_i| = a$
2. for all $1 \leq i < j \leq N$, $|S_i \cap S_j| \leq \log N$

Moreover, these sets can be constructed in time polynomial in b and 2^l. We identify the set $\{1, 2, 3, \ldots, l\}$ with the set of new variables $\overline{Y} = \{Y_1, Y_2, \ldots, Y_l\}$. Before we proceed further, we need some notation. We will pick $\delta = (1 - \mu')/2$ to be a non-negative constant. Given, a, μ', δ, we define $\gamma = \frac{2(\mu'+\delta)+1}{1-(\mu'+\delta)}$. Then, we define q to be the smallest prime number between $(a/2)^{\frac{1+\gamma}{2+\gamma}}$ and $2 \cdot (a/2)^{\frac{1+\gamma}{2+\gamma}}$. Also, we set a' to be equal to $(a/2)^{\frac{1}{2+\gamma}}$. Observe that $a/2 \leq a'q \leq a$.

For each i, such that $1 \leq i \leq N$, let $S_i{}'$ be an arbitrary subset of S_i of size equal to $a'q$. For brevity, we rename the sets S_i' as S_i [6]. Let $\rho = (\mu' + \delta)\frac{\log a'q}{\log a'}$ and $D = \frac{\gamma+\rho}{2(1+\gamma)} \cdot a'$.

Often for the ease of notation we will identify the set S_i of $\{1, 2, \ldots, l\}$ with the set of variables $\{Y_j : j \in S_i\}$. We will think of the variables $\{Y_j : j \in S_i\}$ to be arranged in a $a' \times q$ matrix $V(i)$, with the variables placed in the matrix in some order. For every $i \in \{1, 2, 3, \ldots, N\}$, we define $NW_{a',\mu'}(S_i)$ as

$$NW_{a',\mu'}(S_i) = \sum_{\substack{f(z) \in \mathbb{F}_q[z] \\ deg(f) \leq D-1}} \prod_{j \in [a']} V(i)_{jf(j)}.$$

For a point $p = (p_1, p_2, \ldots, p_l) \in \mathbb{F}^l$, we denote by $NW_{a',\mu'}(S_i)|p$, the evaluation of $NW_{a',\mu'}(S_i)$ when the variable Y_j is set to p_j.

Let G be an arbitrary subset of $\mathbb{F}$ of size $Nka' + 1$. We define the hitting set $\mathcal{H}$ as follows.

▶ **Definition 5.7** (Definition of the hitting set $\mathcal{H}$).
$\mathcal{H} = \left\{ (NW_{a',\mu'}(S_1) \,|\, p, NW_{a',\mu'}(S_2) \,|\, p, \ldots, NW_{a',\mu'}(S_N) \,|\, p) : p \in G^l \right\}$.

We now proceed to prove the correctness of the construction. We first prove the following lemma which shows that $\mathcal{H}$ is explicit and has the correct size as per Theorem 5.6.

▶ **Lemma 5.8.** *The set $\mathcal{H}$ as defined in Definition 5.7 has size at most $(Nka' + 1)^l$ and all its elements can be enumerated in time $a^{a'} \cdot (Nka' + 1)^l \cdot N^{O(1)}$.*

[6] We have replaced the family $\{S_1, S_2, \ldots, S_N\}$ by the set family $\{S_1', S_2', \ldots, S_N'\}$ such that for each $i \in [N]$, $S_i' \subseteq S_i$. Observe that the design based properties of the original system continue to hold. The only thing that changes is that the size of S_i' could be smaller than the size of S_i, by at most a factor 2.

Proof. The size of the set $\mathcal{H}$ is equal to $|G|^l = (Nka' + 1)^l$. The set $\mathcal{H}$ can be enumerated by enumerating through the points p in G^l in some natural order (say lexicographic order) and evaluating the tuple $(NW_{a',\mu'}(S_1)|p, NW_{a',\mu'}(S_2)|p, \ldots, NW_{a',\mu'}(S_N)|p)$ at each of these points. For every point p and subset S_i, the polynomial $NW_{a',\mu'}(S_i)$ can be evaluated in time at most $a^{a'} \times \mathrm{Poly}(N)$ from Lemma 4.2. So, the second part of the lemma follows. ◀

Observe that for our choice of parameters, the above bounds on the size and the time of enumeration are bounded by a function which is subexponential in N.

We now show that for every non-zero polynomial P in the class $\mathcal{C}$, as defined in the statement of Theorem 5.6, there exists a point $p \in \mathcal{H}$, such that $P(p)$ is non-zero. We show this in Lemma 5.9 below. That will complete the proof of Theorem 5.6.

5.3 Correctness of the construction

For the rest of this section, we denote N^μ by s.

▶ **Lemma 5.9.** *Let P be a non-zero polynomial in the set $\mathcal{C}$ as defined in the statement of Theorem 5.6, and let $\mathcal{H}$ be the set defined in Definition 5.7. Then, there is a point p in the set $\mathcal{H}$ such that $P(p) \neq 0$.*

Proof. We define

$$P_i(\overline{X}, \overline{Y}) := P(NW_{a',\mu'}(S_1), NW_{a',\mu'}(S_2), \ldots, NW_{a',\mu'}(S_i), X_{i+1}, X_{i+2}, \ldots, X_N)$$

to be the polynomial obtained from P by substituting the variables X_j by $NW_{a',\mu'}(S_j)$, for every $1 \leq j \leq i$.

From the construction of our hitting set, it follows that it would suffice to argue that the polynomial $P_N(\overline{X}, \overline{Y})$ is non-zero. If this was true, then the lemma above will follow from Lemma 5.5, since the degree of any variable $P(\overline{X}, \overline{Y})$ is at most Nka'.

We proceed via contradiction. If possible, let $P_N(\overline{X}, \overline{Y})$ be identically zero. Since $P = P_0(\overline{X}, \overline{Y})$ is non-zero to start with, by a hybrid argument it follows that there is an index i, such that $P_i(\overline{X}, \overline{Y})$ is non-zero while $P_{i+1}(\overline{X}, \overline{Y})$ is identically zero. Observe that P_i is a polynomial in the variables $\overline{Y}$ and $X_{i+1}, X_{i+2}, \ldots, X_N$. In going from P_i to P_{i+1}, we substituted the variable X_{i+1} by the polynomial $NW_{a',\mu'}(S_{i+1})$. Since $P_i(\overline{X}, \overline{Y})$ is non-zero by assumption above, there exists a substitution $\overline{c}$ of all variables apart from $\{Y_j : j \in S_{i+1}\}$ and X_{i+1}, which keeps the polynomial non-zero. Let the polynomial resulting after this substitution be P_i'. From the definitions, it follows that

$$P_i' = P(NW_{a',\mu'}(S_1)|\overline{c}, NW_{a',\mu'}(S_2)|\overline{c}, \ldots, NW_{a',\mu'}(S_i)|\overline{c}, X_{i+1}, X_{i+2}|\overline{c}, \ldots, X_N|\overline{c})\,.$$

Observe that each of the polynomials $NW_{a',\mu'}(S_j)|\overline{c}$ depends only on the variables in the set $S_j \cap S_{i+1}$. From the properties of Nisan-Wigderson designs, and the choice of parameters, the size of this intersection is at most $\log N$. From the definition of P_i and the choice of $\overline{c}$, P_i' is not identically zero. We will think of P_i' as a polynomial in X_{i+1} with the coefficients being polynomials in the variables in the set $\{Y_j : j \in S_{i+1}\}$. Now, we know that the the polynomial P_{i+1}' obtained by substituting X_{i+1} by $NW_{a',\mu'}(S_{i+1})$ is identically zero. Hence, it must be the case that $X_{i+1} - NW_{a',\mu'}(S_{i+1})$ is a factor of P_i'.

To proceed further, we need the following claim.

▶ **Claim 5.10.** *P_i' as defined above can be represented as*

$$P_i' = \sum_{r=1}^{T} \prod_{j=1}^{d} Q_{rj}'$$

such that each of the polynomials Q_{rj}' depends on at most $s \log N$ variables.

Proof. Recall that P can be represented as

$$P = \sum_{i=1}^{T} \prod_{j=1}^{d} Q_{ij}$$

where each Q_{ij} is a polynomial in at most $s = N^{\mu}$ variables. In going from P to P_i', we have substituted each of the variables outside the set $\{Y_j : j \in S_{i+1}\} \cup \{X_{i+1}\}$ by either a constant or by the polynomial $NW_{a',\mu'}(S_j)|\bar{c}$ (which is a polynomial in at most $|S_j \cap S_{i+1}| \leq \log N$ variables) for some j. In either case, after substitution, the polynomials Q'_{rj} obtained from Q_{rj} depends on at most $s \log N$ variables, since Q_{rj} depended on at most s variables. This completes the proof of the claim. ◀

Moreover, since the individual degree of variables in P is at most k, the individual degree of X_{i+1} in P_i' is at most k. The goal now is to invoke Lemma 5.3, which would imply that $NW_{a',\mu'}(S_{i+1})$ also has a small circuit as a sum of product of polynomials in *few* variables, and together with the lower bound from Theorem 4.13, this would lead to a contradiction. We essentially follow this outline. Formally, we use the following claim to complete the proof of Lemma 5.9. We defer the proof of the claim to the end.

▶ **Claim 5.11.** *If $(X_{i+1} - NW_{a',\mu'}(S_{i+1}))$ divides P_i', then $NW_{a',\mu'}(S_{i+1})$ can be written as*

$$NW_{a',\mu'}(S_{i+1}) = \sum_{r=1}^{I'} \prod_{j=1}^{d'} \Gamma_{rj}$$

where

1. $I' \leq (da'^2 + 1) \cdot \binom{k+a'+1}{k+1} \times \binom{T\cdot(k+1)^3+a'}{a'}^{k+1}$,
2. $d' \leq d \cdot a'$, *and*
3. *each Γ_{rj} depends on at most $s \log N$ variables.*

From our choice of parameters, recall that

$$a = N^{\mu/\mu'} \cdot \log^{1/\mu'} N$$

and

$$s = N^{\mu}.$$

Therefore, $s \log N \leq N^{\mu} \cdot \log N \leq a^{\mu'}$. To complete the proof, we observe that by Theorem 4.13, we must have

$$I'd' \geq (a')^{\Omega(\sqrt{a'})}.$$

But, for our choice of parameters,

1. $I' \leq (da'^2 + 1) \cdot \binom{k+a'}{k} \times \binom{T\cdot(k+1)^3+a'}{a'}^{k+1} \leq da^{O(Tk^4)} \leq da'^{O(Tk^4)}$ (since a and a' are polynomially related)
2. $d' \leq da'$

This implies that $I'd' \leq d^2 a^{O(Tk^4)}$. From our choice of parameters, $s \log N < a^{\mu'}$ and $Tk^4 + 2\log d \in o(\sqrt{a'})$. This contradicts that $I'd' \geq (a')^{\Omega(\sqrt{a'})}$. This completes the proof of Lemma 5.9 assuming Claim 5.11. ◀

We now give a proof of Claim 5.11.

Proof of Claim 5.11. From Claim 5.10, we know that

$$P'_i = \sum_{r=1}^{T} \prod_{j=1}^{d} Q'_{rj}$$

such that each Q'_{rj} depends on at most $s \log N$ variables. Since P'_i is not identically zero and $NW_{a',\mu'}(S_{i+1})$ is a root of P'_i, it follows from Lemma 5.2 that there is an integer λ such that $0 \leq \lambda \leq k-1$ and,

$$\frac{\partial^{\lambda} P'_i}{\partial X_{i+1}^{\lambda}}(NW_{a',\mu'}(S_{i+1})) = 0$$

and

$$\frac{\partial^{\lambda+1} P'_i}{\partial X_{i+1}^{\lambda+1}}(NW_{a',\mu'}(S_{i+1})) \neq 0\,.$$

From Lemma 5.1 it follows that $\tilde{P}'_i = \frac{\partial^{\lambda} P'_i}{\partial X_{i+1}^{\lambda}}$ can also be expressed as

$$\tilde{P}'_i = \sum_{r=1}^{T'} \prod_{j=1}^{d} \tilde{Q}_{ij}$$

where $T' \leq T \cdot (k+1)^2$ and each of the $\tilde{Q}_{rj}$ depends on at most $s \log N$ variables.

Observe that, $\tilde{P}'_i$ vanishes when $NW_{a',\mu'}(S_{i+1})$ is substituted for X_{i+1}, while its derivative with respect to X_{i+1} does not vanish identically at $X_{i+1} = NW_{a',\mu'}(S_{i+1})$. So, in particular, there is a substitution of the Y variables where the derivative $\frac{\partial \tilde{P}'_i}{\partial X_{i+1}}$ is nonzero. Since the class of $\Sigma\Pi\,(\Sigma\Pi)^{[s]}$ circuits is closed under translations of variables (from item 2 in Lemma 5.1), we can assume without loss of generality that the derivative is nonzero when all the variables in $\overline{Y}$ are set to zero. Also observe that by this variable translation, we have actually obtained a polynomial $NW'_{a',\mu'}(S_{i+1})$ from $NW_{a',\mu'}(S_{i+1})$. Moreover, the degree of $NW'_{a',\mu'}(S_{i+1})$ is equal to a' and the homogeneous component of degree a' of $NW'_{a',\mu'}(S_{i+1})$ is equal to $NW_{a',\mu'}(S_{i+1})$. Let the polynomial obtained after the variable translation from $\tilde{P}'_i$ as $\tilde{P}''_i$. At this point, the hypothesis of Lemma 5.3 is satisfied by $\tilde{P}''_i$.

Let $\tilde{P}''_i = \sum_{j=0}^{k} C_j(\overline{Y}) \cdot X_{i+1}^j$. Here, $C_j(\overline{Y})$ is a polynomial only in the Y variables and is the coefficient of X_{i+1}^j, when viewing $\tilde{P}''_i$ as an element of $\mathbb{F}[\overline{Y}][X_{i+1}]$. From Lemma 3.4, we know that each of the polynomials C_j can be expressed as a polynomial of the form

$$C_j = \sum_{r=1}^{T_j} \prod_{l=1}^{d} Q''_{rl}$$

where $T_j \leq T' \cdot (k+1) \leq T \cdot (k+1)^3$ and each Q''_{rl} depends on at most $s \log N$ variables.

Hence, by Lemma 5.3, for every $t \geq 0$, there exists a polynomial $R_t \in \mathbb{F}[Z_1, Z_2, \ldots, Z_{k+1}]$ of degree at most t such that

$$\mathsf{Hom}^{\leq t}[NW'_{a',\mu'}(S_{i+1})] = \mathsf{Hom}^{\leq t}[R_t(C_0, C_1, \ldots, C_k)]\,.$$

The goal now is to obtain a representation of $NW_{a',\mu'}(S_{i+1})$ as a sum of products of polynomials in few variables and show that this contradicts the lower bound in Theorem 4.13.

$NW'_{a',\mu'}(S_{i+1})$ is a polynomial of degree at most a'. So, there is a polynomial $R_{a'}$ of degree at most a' in $k+1$ variables such that

$$NW'_{a',\mu'}(S_{i+1}) = \mathsf{Hom}^{\leq a'}[R_{a'}(C_0, C_1, \ldots, C_k)]\,.$$

From the discussion on the relation between $NW'_{a',\mu'}(S_{i+1})$ from $NW_{a',\mu'}(S_{i+1})$, we also know that

$$NW_{a',\mu'}(S_{i+1}) = \mathsf{Hom}^{a'}[NW'_{a',\mu'}(S_{i+1})] = \mathsf{Hom}^{a'}[R_{a'}(C_0, C_1, \ldots, C_k)]\,.$$

Since $R_{a'}$ is a polynomial in $k+1$ variables of degree a', the number of monomials in $R_{a'}$ is at most $\binom{a'+k+1}{k+1}$. Therefore, we can represent $R_{a'}(C_0, C_1, \ldots, C_k)$ as a sum of products of the C_j's, with the sum fan-in at most $\binom{a'+k+1}{k+1}$ and the product fan-in at most a'. Moreover, each of the product gates in this representation takes the polynomials C_j's as inputs. We know that each C_j can be written as

$$C_j = \sum_{r=1}^{T_j} \prod_{l=1}^{d} Q''_{rl}$$

where each Q''_{rl} is a polynomial in at most $s \log N$ variables, and the top sum fan-in T_j is at most $T \cdot (k+1)^3$. For any t, the polynomial C_j^t, has a similar representation with the top sum fan-in at most $\binom{T\cdot(k+1)^3+t}{t}$. Therefore, any product of fan-in at most a' in the C_j's can be written as a sum of product of polynomials in at most $s \log N$ variables, with top fan-in at most

$$\binom{T \cdot (k+1)^3 + a'}{a'}^{k+1}$$

since each C_j is raised to a power of at most a' and there are $k+1$ such C_j's. Therefore, $R_{a'}(C_0, C_1, \ldots, C_k)$ can be written as

$$R_{a'}(C_0, C_1, \ldots, C_k) = \sum_{r=1}^{I} \prod_{j=1}^{d'} \Gamma'_{rj}$$

such that

1. $I \leq \binom{k+a'+1}{k+1} \times \binom{T\cdot(k+1)^3+a'}{a'}^{k+1}$
2. $d' \leq d \cdot a'$
3. Each Γ'_{rj} depends on at most $s \log N$ variables

We would now like to extract the homogeneous part of degree a' of $R_{a'}(C_0, C_1, \ldots, C_k)$, which we know is equal to $NW_{a',\mu'}(S_{i+1})$. We do this by a standard application of Lemma 3.5. Since we are interested only in the homogeneous part of degree a', we can assume without loss of generality that each of the polynomials Γ'_{rj} is of degree at most a' (we can discard all monomials of degree larger than a' in each of the Γ'_{rj}, since they do not contribute to the homogeneous component of degree a' of $R_{a'}(C_0, C_1, \ldots, C_k)$). Hence, the degree of $R_{a'}(C_0, C_1, \ldots, C_k)$ is upper bounded by $da' \cdot a'$. So, from Lemma 3.5, we can extract the homogeneous component of degree a' of $R_{a'}(C_0, C_1, \ldots, C_k)$ by blowing up the top fan-in by a factor of at most $da'^2 + 1$. Hence, $NW_{a',\mu'}(S_{i+1})$ can be expressed as

$$NW_{a',\mu'}(S_{i+1}) = \sum_{r=1}^{I'} \prod_{j=1}^{d'} \Gamma_{rj}$$

where

1. $I' \leq (da'^2+1) \cdot \binom{k+a'+1}{k+1} \times \binom{T\cdot(k+1)^3+a'}{a'}^{k+1}$,
2. $d' \leq d \cdot a'$, and
3. each Γ_{rj} depends on at most $s \log N$ variables. ◀

We remark that if the value of s was $\log^{O(1)} N$ to start with, the same proof as above goes through with l and a being set to polynomials of sufficiently high degree in $\log N$. The size of the hitting set and the time to construct it in this case are upper bounded by a quasipolynomial function in N.

6 Open problems

We conclude with some open problems.

1. An intriguing open question is to obtain PIT for $\Sigma\Pi(\Sigma\Pi)^{[s]}$ circuits without the restriction on the individual degree. The strategy in this paper relies on hardness randomness tradeoffs for bounded depth circuits [6]. The tradeoffs in [6] crucially use the fact that the individual degree is bounded.
2. Another related question would be to get any non-trivial PIT (even subexponential) for the sum of constant many products of degree two polynomials.
3. A related question of interest is to obtain non-trivial PIT for sums of products of polynomials in few variables with bounded individual degree but without a restriction on the top fan-in.
4. It would also be interesting to understand if one could obtain any non-trivial PIT for slightly non-multilinear depth four circuits (say individual degree at most 2) with bounded top fan-in. A natural strategy for this question would be to reduce it to the case of $\Sigma\Pi(\Sigma\Pi)^{[s]}$ circuits by either expanding out the polynomials Q_{ij} which depend on too many variables or use a partial derivative like trick, as in [5]. The immediate challenge in this case is that the top fan-in seems to increase by any of these tricks and the calculations in this paper seem to not work out.

Acknowledgements. We would like to thank Rafael Oliveira for many helpful discussions regarding hardness-randomness tradeoffs for bounded depth arithmetic circuits at the early stages of this work. We are thankful to the anonymous reviewers at CCC-2016, whose comments helped improve the presentation in the paper.

References

1 M. Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *FOCS*, 2008.
2 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 *Proceedings of the 44th ACM symposium on Theory of computing*, pages 599–614, 2012.
3 Manindra Agrawal, Chandan Saha, and Nitin Saxena. Quasi-polynomial hitting-set for set-depth-Δ formulas. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC'13, pages 321–330, New York, NY, USA, 2013. ACM. doi:10.1145/2488608.2488649.
4 Noga Alon. Combinatorial nullstellensatz. *Combinatorics, Probability and Computing*, 8, 1999.
5 Rafael Mendes de Oliveira, Amir Shpilka, and Ben Lee Volk. Subexponential size hitting sets for bounded depth multilinear formulas. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:157, 2014.

6 Zeev Dvir, Amir Shpilka, and Amir Yehudayoff. Hardness-randomness tradeoffs for bounded depth arithmetic circuits. *SIAM J. Comput.*, 39(4):1279–1293, 2009. doi:10.1137/080735850.

7 Michael Forbes. Deterministic divisibility testing via shifted partial derivatives. In *FOCS*, 2015.

8 Michael A. Forbes and Amir Shpilka. On identity testing of tensors, low-rank recovery and compressed sensing. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19–22, 2012*, pages 163–172, 2012.

9 Michael A. Forbes and Amir Shpilka. Quasipolynomial-time identity testing of non-commutative and read-once oblivious algebraic branching programs. *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, 0:243–252, 2013.

10 H. Fournier, N. Limaye, G. Malod, and S. Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *STOC*, 2014.

11 A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi. Approaching the chasm at depth four. In *CCC*, 2013.

12 Ankit Gupta. Algebraic geometric techniques for depth-4 PIT & sylvester-gallai conjectures for varieties. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:130, 2014. URL: http://eccc.hpi-web.de/report/2014/130.

13 Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth three. In *FOCS*, 2013.

14 V. Kabanets and R. Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004.

15 N. Kayal, N. Limaye, C. Saha, and S. Srinivasan. An exponential lower bound for homogeneous depth four arithmetic formulas. In *FOCS*, 2014.

16 Neeraj Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. *ECCC*, 19:81, 2012. URL: http://eccc.hpi-web.de/report/2012/081.

17 Neeraj Kayal and Chandan Saha. Lower bounds for depth three arithmetic circuits with small bottom fanin. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:89, 2014. URL: http://eccc.hpi-web.de/report/2014/089.

18 Neeraj Kayal and Chandan Saha. Lower bounds for sums of products of low arity polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 2014.

19 Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *STOC*, 2014.

20 P. Koiran. Arithmetic circuits : The chasm at depth four gets wider. *TCS*, 2012.

21 Mrinal Kumar and Shubhangi Saraf. The limits of depth reduction for arithmetic formulas: It's all about the top fan-in. In *STOC*, 2014.

22 Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. *FOCS*, 2014.

23 Partha Mukhopadhyay. Depth-4 identity testing and noether's normalization lemma. *Electronic Colloquium on Computational Complexity (ECCC)*, 2015.

24 Noam Nisan. Lower bounds for non-commutative computation (extended abstract). In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, May 5-8, 1991, New Orleans, Louisiana, USA*, pages 410–418, 1991.

25 Noam Nisan and Avi Wigderson. Hardness vs randomness. *Journal of Computer and System Sciences*, 49(2):149–167, 1994.

26 Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994. doi:10.1016/S0022-0000(05)80043-1.

27 Ran Raz. Separation of multilinear circuit and formula size. *Theory of Computing*, 2(1):121–135, 2006. doi:10.4086/toc.2006.v002a006.

28 S. Saraf and I. Volkovich. Black-box identity testing of depth-4 multilinear circuits. In *Proceedings of the 43rd Annual STOC*, pages 421–430, 2011.

29 Nitin Saxena. Diagonal circuit identity testing and lower bounds. *Electronic Colloquium on Computational Complexity (ECCC)*, 14(124), 2007.

30 J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of ACM*, 27(4):701–717, 1980.

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

32 Amir Shpilka. Affine projections of symmetric polynomials. In *Proceedings of the 16th Annual Conference on Computational Complexity*, CCC'01, pages 160–, Washington, DC, USA, 2001. IEEE Computer Society.

33 Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *MFCS*, pages 813–824, 2013. doi:10.1007/978-3-642-40313-2_71.

34 L. G. Valiant. Completeness classes in algebra. In *STOC*, 1979.

35 Leslie G. Valiant, Sven Skyum, S. Berkowitz, and Charles Rackoff. Fast parallel computation of polynomials using few processors. *SIAM Journal of Computation*, 12(4):641–644, 1983. doi:10.1137/0212043.

36 R. Zippel. Probabilistic algorithms for sparse polynomials. *Symbolic and algebraic computation*, pages 216–226, 1979.

A Calculations

$$Td^2n^3 \cdot rs \geq \frac{\frac{1}{n^{O(1)}}\min\left(\frac{p^r}{4^r}\cdot\binom{N}{r}\cdot\binom{N}{m},\binom{N}{m+n-r}\right)}{2^{O(\sqrt{n})}\cdot\binom{N}{m+rs}\cdot\binom{n+r}{r}}.$$

We first estimate the ratio $\frac{\binom{N}{m+n-r}}{\binom{N}{m+rs}\cdot\binom{n+r}{r}}$:

$$\frac{\binom{N}{m+n-r}}{\binom{N}{m+rs}\cdot\binom{n+r}{r}} \geq \frac{(m+rs)!}{(m+n-r)!}\frac{(N-m-rs)!}{(N-m-(n-r))!}\cdot\left(\frac{r}{e(n+r)}\right)^r$$

Here we use the fact that $\binom{n+r}{r} \leq \left(\frac{e(n+r)}{r}\right)^r$. Now, approximating the ratios using Lemma 3.3 and substituting $m = \frac{N}{2}(1 - r\frac{\ln n}{n})$, we get

$$\begin{aligned}\frac{\binom{N}{m+n-r}}{\binom{N}{m+rs}\cdot\binom{n+r}{r}} &\geq \left(\frac{N-m}{m}\right)^{n-r-rs}\cdot\left(\frac{r}{e(n+r)}\right)^r \\ &\geq \exp\left(\frac{r\ln n}{n}\cdot(n-r-rs) - r\ln\frac{e(n+r)}{r}\right)\end{aligned}$$

Since $r = \Theta(\sqrt{n})$, we get that the ratio is at least $\exp\left(r\ln n((n-r-rs)/n - \frac{1}{2} + o(1))\right)$, which is $\exp(\Omega(\sqrt{n}\ln n))$.

Next we estimate the ratio $\frac{\left(\frac{p^r}{4^r}\cdot\binom{N}{r}\cdot\binom{N}{m}\right)}{\binom{N}{m+rs}\cdot\binom{n+r}{r}}$:

$$\begin{aligned}
\frac{\left(\frac{p^r}{4^r}\cdot\binom{N}{r}\cdot\binom{N}{m}\right)}{\binom{N}{m+rs}\cdot\binom{n+r}{r}} &\geq \frac{p^r}{4^r}\cdot\frac{(m+rs)!}{m!}\cdot\frac{(N-m-rs)!}{(N-m)!}\cdot\frac{N!}{(N-r)!}\cdot\frac{n!}{(n+r)!}\\
&\geq \frac{p^r}{4^r}\cdot\left(\frac{m}{N-m}\right)^{rs}\cdot\left(\frac{N}{n}\right)^r\\
&\geq \frac{p^r}{4^r}\cdot\left(1-2.01r\frac{\ln n}{n}\right)^{rs}\cdot\left(\frac{N}{n}\right)^r\\
&\geq \frac{1}{4^r}\exp\left(-r(\mu+\delta)\ln N-2.01r^2s\frac{\ln n}{n}+r\ln(N/n)\right)
\end{aligned}$$

Here, we used Lemma 3.3 in the second step and substituted $p = N^{-(\delta+\mu)}$ in the last step. Now, substituting $2n^{2+\gamma} \geq N \geq n^{2+\gamma}$, the exponent is at least

$$r\ln n(-(\mu+\delta)(2+\gamma)-2.01rs/n+(1+\gamma))\,.$$

This is at least

$$r\ln n(-(\mu+\delta)(2+\gamma)-2.01rs/n+(1+\gamma))\,.$$

Now, plugging back the value of γ, the exponent is at least $(2-2.01rs/n)r\ln n$. We have chosen rs such that $rs/n < 0.001$. Therefore, the ratio we set out to lower bound is at least $\exp(\Omega(\sqrt{n}\ln n))$.