A Depth-Five Lower Bound for Iterated Matrix Multiplication*

Suman K. Bera and Amit Chakrabarti

Department of Computer Science, Dartmouth College
Hanover, USA
{suman.k.bera.gr, amit.chakrabarti}@dartmouth.edu

Abstract

We prove that certain instances of the iterated matrix multiplication (IMM) family of polynomials with N variables and degree n require $N^{\Omega(\sqrt{n})}$ gates when expressed as a homogeneous depth-five $\Sigma\Pi\Sigma\Pi\Sigma$ arithmetic circuit with the bottom fan-in bounded by $N^{1/2-\varepsilon}$. By a depth-reduction result of Tavenas, this size lower bound is optimal and can be achieved by the weaker class of homogeneous depth-four $\Sigma\Pi\Sigma\Pi$ circuits.

Our result extends a recent result of Kumar and Saraf, who gave the same $N^{\Omega(\sqrt{n})}$ lower bound for homogeneous depth-four $\Sigma\Pi\Sigma\Pi$ circuits computing IMM. It is analogous to a recent result of Kayal and Saha, who gave the same lower bound for homogeneous $\Sigma\Pi\Sigma\Pi\Sigma$ circuits (over characteristic zero) with bottom fan-in at most $N^{1-\varepsilon}$, for the harder problem of computing certain polynomials defined by Nisan–Wigderson designs.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases arithmetic circuits, iterated matrix multiplication, depth five circuits, lower bound

Digital Object Identifier 10.4230/LIPIcs.CCC.2015.183

1 Introduction

The fundamental goal of algebraic complexity theory is an understanding of which polynomials can be computed efficiently. Arithmetic formulas and circuits, being the most natural and intuitive model for computing polynomials, are the basis for the notion of *complexity* of a polynomial. They are defined in an analogous way to Boolean formulas and circuits, the key difference being that the *gates* used to build them are addition (+) and multiplication (×) gates, rather than logic gates (more details appear in Section 2).

A classic result in the area is that the symbolic $n \times n$ determinant – an n^2-variate polynomial of degree n – can be computed by a poly(n)-sized arithmetic circuit over an arbitrary field [2]. A classic open problem is to prove that the symbolic $n \times n$ permanent – also n^2-variate and of degree n – cannot be so computed. In a highly influential work, Valiant [22] defined complexity classes analogous to P and NP for the algebraic world, which have since come to be called VP (polynomial-sized arithmetic circuits) and VNP (roughly, polynomial-sized arithmetic circuits with a summation quantifier; the permanent has such circuits), and hypothesized that VP $\neq$ VNP. Proving this separation is the preeminent open problem in the area.

Recent work, starting with Agarwal and Vinay [1], has shown that the VP $\neq$ VNP conjecture can be attacked by focusing on *constant-depth* circuits (equivalently, constant-

* This work was partially supported by NSF grant CCF-1217375.

30th Conference on Computational Complexity (CCC'15).
Editor: David Zuckerman; pp. 183–197
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

depth formulas). In particular, it suffices to prove certain strong size lower bounds for depth-four $\Sigma\Pi\Sigma\Pi$ circuits: these are layered circuits with four layers of gates, alternating between $+$-gates and $\times$-gates, with a $+$-gate at the top (output) level. A flurry of research over the last two years has greatly advanced our understanding of the power of such circuits. Tavenas [21] has shown that every $N^{O(1)}$-sized arithmetic circuit on N variables computing a polynomial of degree $n = N^{\Theta(1)}$ can be transformed into a depth-four $\Sigma\Pi\Sigma\Pi$ circuit of size $N^{O(\sqrt{n})}$ with bottom fan-in at most $O(\sqrt{n})$. Moreover the transformation preserves *homogeneity*: if the original circuit is homogeneous – meaning that each $+$-gate computes a homogeneous polynomial – then so is the transformed circuit.

In a recent *tour de force*, Kumar and Saraf [16] showed that depth-four homogeneous circuits for the iterated matrix multiplication (IMM) family of polynomials *require* $N^{\Omega(\sqrt{n})}$ size even without a restriction of the bottom fan-in; again N and n represent the number of variables and the degree (respectively), and their proof uses $N \approx n^{11}$. Since the IMM polynomials are easily seen to have polynomial-sized arithmetic circuits, this lower bound shows that Tavenas's depth reduction result is tight in a strong sense.

1.1 Our Results

We extend the above Kumar–Saraf theorem to obtain a similar exponential lower bound for depth-five homogeneous $\Sigma\Pi\Sigma\Pi\Sigma$ circuits, albeit with a restriction on the bottom fan-in. Namely, we consider circuits where this bottom fan-in is at most N^μ, where $\mu < 1/2$ is a constant. For each such μ, we shall consider a family of N-variate degree-n IMM polynomials, where $N = n^{\Theta(q)}$ and q is a constant depending on μ, and show that our restricted depth-five circuits require $N^{\Omega(\sqrt{n})}$ size to compute these polynomials. By Tavenas's above theorem, the bound $N^{\Omega(\sqrt{n})}$ is tight.

The IMM polynomials are defined as follows. The variables are $\{z_{i,j}^{(h)}\}_{h\in[n],\, i,j\in[m]}$, to be thought of as entries of $m \times m$ matrices $Z^{(1)}, \ldots, Z^{(n)}$: we use the standard convention that the (i,j)-entry of a matrix A is denoted $a_{i,j}$. The polynomial $\mathrm{IMM}_{n,m}$ on these variables is defined as the $(1,1)$-entry of the matrix product $Z^{(1)}Z^{(2)}\cdots Z^{(n)}$. Thus,

$$\mathrm{IMM}_{n,m}\left(z_{1,1}^{(1)}, \ldots, z_{m,m}^{(n)}\right) = \sum_{i_1,i_2,\ldots,i_{n-1}\in[m]} z_{1,i_1}^{(1)} z_{i_1,i_2}^{(2)} \cdots z_{i_{n-2},i_{n-1}}^{(n-1)} z_{i_{n-1},1}^{(n)} \,. \tag{1}$$

▶ **Theorem 1.1** (Main Theorem). *For every constant $\mu \in [0, 1/2)$, there is an integer $q > 0$ such that the following holds. With $m = n^q$, any homogeneous $\Sigma\Pi\Sigma\Pi\Sigma$ circuit with bottom fan-in N^μ that computes the N-variate degree-n polynomial $\mathrm{IMM}_{n,m}$ has size at least $N^{\Omega(\sqrt{n})}$.*

A more precise version of this theorem appears as Theorem 4.2 on page 194.

Proving a super-polynomial lower bound for homogeneous $\Sigma\Pi\Sigma\Pi\Sigma$ circuits was explicitly posed as an open problem by Nisan and Wigderson [18, Section 2.2] in their pioneering work on arithmetic circuit lower bounds. In particular, a depth-five lower bound for IMM was posed as an open problem by Kayal and Saha in a recent work [12] where they obtained such bounds for the so-called Nisan–Wigderson (NW) family of polynomials.[1] Unlike IMM, the NW polynomials are not known to have polynomial-sized arithmetic circuits. This is a strength of our result, because it applies to a potentially "easier" family of polynomials.

Another strength of our result is that, unlike the above Kayal–Saha result, it does not depend on any properties of the underlying field $\mathbb{F}$ over which the circuit is defined. Their

[1] The name "Nisan–Wigderson" for these polynomials refers to a still earlier work of Nisan and Wigderson [17] that popularized a certain kind of combinatorial design.

result crucially relies on $\mathbb{F}$ having characteristic zero. Meanwhile a weakness of our result is the $\mu < 1/2$ requirement; the analogous requirement in Kayal–Saha is that $\mu < 1$, which is still a restriction on the structure of the circuit but a weaker one.

We shall prove our depth-five lower bound for a slight restriction of the polynomial in eq. (1) obtained by setting some of its variables to 1 (clearly this only strengthens the result). Our proof will use machinery from the recent work of Kayal and Saha [12] to essentially transform a depth-five circuit into a depth-four one, while controlling the bottom fan-in.

1.2 Related Work

In a seminal work, Valiant *et al.* [23] gave the first nontrivial depth-reduction technique for general arithmetic circuits. They proved that a $\text{poly}(N)$-sized N-variate arithmetic circuit that computes a polynomial with degree $\text{poly}(N)$ can be assumed to be of $\text{poly}(\log N)$ depth without loss of generality. All subsequent depth-reduction results have built on this work. In particular, Agarwal and Vinay [1] gave a reduction to depth four and the parameters of this reduction were subsequently refined and improved by Koiran [15] and, most recently, by Tavenas [21] who gave the result described earlier.

A consequence of Tavenas's theorem is that a size lower bound of $N^{\omega(\sqrt{n})}$ for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits computing a homogeneous polynomial[2] f shows that $f \notin \mathsf{VP}$; in fact the circuits can be restricted to a bottom fan-in of $O(\sqrt{n})$. In particular, proving such a strong lower bound with f being either the permanent polynomial or the NW polynomial would imply $\mathsf{VP} \neq \mathsf{VNP}$. A number of recent works have pursued this research program and made significant progress.

This research program can be traced back to the groundbreaking work of Nisan and Wigderson [18], which introduced the idea of studying the dimension of the space of partial derivatives of a polynomial f. Lower bounds on this dimension imply lower bounds on the size of depth-three $\Sigma\Pi\Sigma$ circuits for f. In particular, this technique shows that a *homogeneous* $\Sigma\Pi\Sigma$ circuit computing the $n \times n$ symbolic determinant (over an arbitrary field) must have size $2^{\Omega(n)}$. Gupta *et al.* [7] greatly strengthened this technique by considering "shifted" partial derivatives (see Section 2), and proved that a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit with bottom fan-in at most $\sqrt{n}$ that computes either the $n \times n$ determinant or the $n \times n$ permanent must have size $2^{\Omega(\sqrt{n})}$. Kayal *et al.* [13] then proved a larger lower bound of $N^{\Omega(\sqrt{n})}$ for the same class of circuits, for the "harder" problem of computing an N-variate degree-n NW polynomial. Fournier *et al.* [4] proved the same lower bound for the problem of computing certain IMM polynomials.

The next major conceptual advance was made by Kayal *et al.* [11], who further strengthened the partial derivatives technique by adding a multilinear projection step, arriving at the "dimension of projected shifted partials" measure. Using this, and further applying well-chosen random restrictions, they removed the bottom fan-in restriction and gave an $N^{\Omega(\sqrt{n})}$ lower bound for homogeneous $\Sigma\Pi\Sigma\Pi$ circuits computing NW polynomials. However, their proof introduced a new restriction: the underlying field had to have characteristic zero. The aforementioned recent work by Kumar and Saraf [16] does not have such a restriction on the field and gives the same $N^{\Omega(\sqrt{n})}$ lower bound for certain NW polynomials as well as certain IMM polynomials. Since IMM $\in \mathsf{VP}$, this proves the tightness of Tavenas's theorem [21].

[2] Strictly speaking, one considers the complexity not of a single polynomial but of a family of polynomials $\{f_N\}_{N \in \mathcal{I}}$ for some infinite index set $\mathcal{I} \subseteq \mathbb{N}$.

Along different lines Grigoriev and Karpinski [5], and Grigoriev and Razborov [6] considered (not necessarily homogeneous) $\Sigma\Pi\Sigma$ circuits over a finite field $\mathbb{F}$ and proved that computing the MOD_q function on n variables, where $q \neq \mathrm{char}(\mathbb{F})$ is a prime, requires size $2^{\Omega(n)}$. In contrast, over a field of characteristic zero, a result of Gupta *et al.* [8] shows that a polynomial-sized N-variate arithmetic circuit can be converted to a non-homogeneous $\Sigma\Pi\Sigma$ circuit of size $N^{O(\sqrt{n})}$. Thus, another approach to proving $\mathsf{VP} \neq \mathsf{VNP}$ would be to show strong enough lower bounds for general $\Sigma\Pi\Sigma$ circuits.

Recently Kayal and Saha [12] proved that a $\Sigma\Pi\Sigma$ circuit over a field of characteristic zero computing certain N-variate degree-n polynomials – namely, NW and IMM polynomials with $N = n^{\Theta(1)}$ – must have size $N^{\Omega(\sqrt{n})}$, provided the bottom fan-in is at most $\sqrt{n}$. Their technique involves converting the $\Sigma\Pi\Sigma$ circuit into a homogeneous $\Sigma\Pi\Sigma\Pi\Sigma$ circuit with bounded bottom fan-in (precisely the class of circuits that this paper is about) and then proving lower bounds for the resulting depth-five circuits. In fact their depth-five circuits have a very special structure, which they then exploit to obtain their NW and IMM results. Without using this special structure, they are still able to obtain lower bounds for NW, but not IMM.

As noted in Theorem 1.1, this paper gives such a depth-5 lower bound for IMM. Our own proof builds on the ideas of Kayal and Saha.

The excellent survey by Shpilka and Yehudayoff [20] gives a much more detailed overview of classic and modern results on arithmetic circuits. Two new surveys by Kayal and Saptharishi [14], and Saptharishi [19] cover recent progress on constant-depth lower bounds.

2 Preliminaries and Proof Outline

All arithmetic circuits studied in this paper will be constant-depth, layered, and homogeneous, with gates of arbitrary fan-in except where noted. A layer is either a Σ-layer (consisting of $+$-gates only) or a Π-layer (consisting of $\times$-gates only). The output layer always consists of a single $+$-gate. Notation of the form $\Sigma\Pi\Sigma\Pi$ indicates the number and types of layers with the leftmost symbol corresponding to the output (a.k.a. top) layer and the rightmost symbol corresponding to the bottom layer, whose gates read only input variables.[3] Wires feeding $+$-gates are labeled with coefficients from the underlying field $\mathbb{F}$: thus a $+$-gate computes an arbitrary linear form over $\mathbb{F}$.

Following Kumar and Saraf [16], we shall consider the following restriction of the IMM polynomial defined in eq. (1). Let

$$m = n^q\,, \quad n = (B+2)k\,, \tag{2}$$

for some integers q, B, and k. Eventually, we shall take $B = \Theta(\sqrt{n})$, $k = \Theta(\sqrt{n})$, and q large enough but constant. We partition the sequence of matrices $Z^{(1)}, \ldots, Z^{(n)}$ into k contiguous subsequences, which we call *blocks*. In the hth block, we denote the first matrix as $Y^{(h)}$, the next B matrices as $X^{(h,1)}, X^{(h,2)}, \ldots X^{(h,B)}$, and the last matrix as $J^{(h)}$. We then set all entries of each $J^{(h)}$ to be 1. We shall denote the resulting polynomial, which is slightly smaller than the original and uses a different set of variable names, as

$$f_{n,q}\big(x_{1,1}^{(1,1)}, \ldots, x_{m,m}^{(k,B)}, y_{1,1}^{(1)}, \ldots, y_{m,m}^{(k)}\big)\,. \tag{3}$$

Clearly, $\deg f_{n,q} \leqslant n$ and $f_{n.q}$ is N-variate for $N = m^2(n-k)$.

[3] When studying non-homogeneous circuits, we must also allow the bottom layer gates to read the constant 1.

Our lower bound is based on the complexity measure termed "dimension of projected shifted partials" (DPSP), whose history we have recounted in Section 1.2. We now define the DPSP measure. Fix a field $\mathbb{F}$ and a set of variables $x_1, \ldots, x_N$. Consider a polynomial $f(x_1, \ldots, x_N) \in \mathbb{F}[x_1, \ldots, x_N]$. Let $\alpha = x_{i_1} \ldots x_{i_k}$ be a multilinear monomial in the same variables. We use the compact notation $\partial_\alpha f := \partial^k f / \partial x_{i_1} \cdots \partial x_{i_k}$, calling it the partial derivative of f with respect to α. Let $\mathcal{M}$ be a set of multilinear monomials and $\ell \geqslant 0$ be an integer. We define

$$\mathrm{DPSP}_{\mathcal{M},\ell}(f) := \dim \operatorname{span} \operatorname{proj} \operatorname{shift}_\ell \{\partial_\alpha f : \alpha \in \mathcal{M}\}, \tag{4}$$

where $\operatorname{shift}_\ell f = \{\beta f : \beta \text{ is a monomial of degree } \ell\}$, $\operatorname{proj} f$ is the projection of f onto the subspace of the $\mathbb{F}$-vector-space $\mathbb{F}[x_1, \ldots, x_N]$ spanned by multilinear monomials, and these operators are extended to sets of polynomials in the natural way.

For fixed choices of $\mathcal{M}$ and ℓ, the measure $\mathrm{DPSP}_{\mathcal{M},\ell}$ is easily seen to be subadditive. It is a good complexity measure because it can be nontrivially upper-bounded for "simple" circuits. Let us call a circuit *t-supported* if at most t *distinct* variables feed each bottom-level gate.

▶ **Lemma 2.1** (Essentially [10, Corollary 12]). *Let C be a t-supported degree-n homogeneous $\Sigma\Pi\Sigma\Pi$ circuit on N variables, with top fan-in at most S_0. Let $\mathcal{M}$ be a set of degree-k multilinear monomials on these N variables and let $\ell \geqslant 0$ be an integer such that $\ell + kt \leqslant N/2$. Then*

$$\mathrm{DPSP}_{\mathcal{M},\ell}(C) \leqslant S_0 \binom{2n/t+1}{k} \binom{N}{\ell + kt}. \quad ◀$$

To apply this to depth *five* circuits with small support, we proceed as in Kayal and Saha [12]: we perform a random restriction. That is, we kill (set to zero) all variables x_i lying outside a suitably randomly chosen subset V. This will simplify a polynomial f to a "smaller" polynomial, which we will denote $f|_V$. The crux of our argument is to show that a sufficiently strong restriction will, w.h.p., simplify a depth-five circuit into a depth-four circuit (the truth is a little more subtle; see Lemma 2.4). At the same time, we do not want to apply too strong a restriction, for otherwise the IMM polynomial itself might simplify too much. We desire that, w.h.p., the restricted polynomial $f_{n,q}|_V$ (see eq. (3)) still has "high" complexity, with respect to our DPSP measure.

2.1 Random Restrictions and Their Effect on IMM

Let $\mathcal{V}_{n,q}$ denote the set of variables of the polynomial $f_{n,q}$; see eq. (3). We now define a distribution over subsets of $\mathcal{V}_{n,q}$ by describing a procedure for sampling a random subset, V. The set V is a union (over h, h', and i) of random subsets $V_i^{(h,h')}$ and $V_i^{(h)}$, which are subsets of the variables in the ith row of $X^{(h,h')}$ and $Y^{(h)}$ respectively; these subsets are mutually independent. Each such subset is chosen uniformly conditioned on its size being some particular quantity, as follows (the parameters b and λ will be fixed later).

- For each h, $|V_1^{(h)}| = m^b = n^{bq}$, where $b \in (0,1)$. Further, $|V_i^{(h)}| = 0$ for $i \neq 1$.
- For each h, $|V_i^{(h,1)}| = n^\lambda$ for each i, where $\lambda \approx 2$.
- For each h and h', with $2 \leqslant h' \leqslant B - 2\log n$, $|V_i^{(h,h')}| = 2$ for each i.
- For each h and h', with $h' > B - 2\log n$, $|V_i^{(h,h')}| = 1$ for each i.

Then, as mentioned above, we set

$$V := \bigcup_{i=1}^{m} \bigcup_{h=1}^{k} \left(V_i^{(h)} \cup \bigcup_{h'=1}^{B} V_i^{(h,h')} \right). \tag{5}$$

Technically, our proof is all about studying the effects of restricting our depth-five circuits and the IMM polynomial to this random set V. This random restriction is a small generalization of the one used by Kumar and Saraf [16], where we have introduced b as a tunable parameter. Therefore, their (highly technical) analysis of the effect of this random restriction on the IMM polynomial largely carries over. We shall now explain the final outcome of this analysis.

To this end, we introduce the following key parameters:

$$k := 32\sqrt{n}\,; \qquad \text{(this then determines } B\text{)} \tag{6}$$

$$\hat{n} := Bk = n - 2k\,; \qquad \text{(the number of } X \text{ matrices)} \tag{7}$$

$$\ell := \frac{N}{2}\left(1 - \frac{\ln n}{\Gamma\sqrt{n}}\right), \qquad \text{where} \tag{8}$$

$$\Gamma := 2 + o(1) \qquad \text{is chosen such that } n^{\sqrt{n}}\left(\frac{N}{N-\ell}\right)^{\hat{n}} = \left(\frac{N}{\ell}\right)^{\hat{n}}\,; \tag{9}$$

$$\lambda := 2 - \frac{1+o(1)}{32\Gamma} \qquad \text{is chosen such that } n^{\lambda k} \cdot 2^{\hat{n}-(1+2\log n)k} = \left(\frac{N}{N-\ell}\right)^{\hat{n}}. \tag{10}$$

For each V drawn as indicated above, let $\mathcal{M}(V)$ denote the set of all monomials obtained by picking exactly one y-variable from each set $V_1^{(h)}$; the degree of each such monomial is then k.

▶ **Lemma 2.2** (Slight generalization of [16, Lemma 8.1]). *Suppose $bq > 1$. Then, for every realization of the random set V, there exists $\mathcal{M}'(V) \subseteq \mathcal{M}(V)$ such that $|\mathcal{M}'(V)| = n^{\sqrt{n}}$ and $\forall\, \alpha_1 \neq \alpha_2 \in \mathcal{M}'(V)$,*

$$|\operatorname{supp}(\alpha_1) \setminus \operatorname{supp}(\alpha_2)| = |\operatorname{supp}(\alpha_2) \setminus \operatorname{supp}(\alpha_1)| \geqslant k - \sqrt{n}\,,$$

where the support $\operatorname{supp}(\alpha)$ of a monomial α is defined as the set of variables that appear in α.

▶ **Lemma 2.3** (Essentially [16, Lemma 8.9]). *Suppose $bq > 1$. With probability at least 0.9, the above set $\mathcal{M}'(V)$ contains a subset $\mathcal{M}''(V)$ such that*

$$\mathrm{DPSP}_{\mathcal{M}''(V),\ell}\left(f_{n,q}|_V\right) \geqslant \frac{n^{\sqrt{n}}}{O(n^{\sqrt{n}/8}) \cdot n^{o(\sqrt{n})}} \left(\frac{N}{N-\ell}\right)^{\hat{n}} \binom{N-\hat{n}}{\ell}.$$

2.2 Circuit Decomposition Under Random Restrictions

To prove our depth-five lower bound using the DPSP lower bound given by Lemma 2.3, we will need to extend Lemma 2.1 as discussed right after its statement. We will analyze the random restriction defined in Section 2.1 to establish the following decomposition lemma.

▶ **Lemma 2.4** (Analogous to [12, Lemma 11]). *For each constant $\mu < 1/2$, there exists an integer $q = q(\mu)$ such that the following holds. Let C be an N^{μ}-supported degree-n homogeneous $\Sigma\Pi\Sigma\Pi\Sigma$ circuit on the variables $\mathcal{V}_{n,q}$, with size $S \leqslant n^{\varepsilon\sqrt{n}}$ for some small positive constant ε. Let the random set V be drawn as above, with b chosen such that $bq \geqslant \lambda$. Then with probability $1 - o(1)$,*

$$C|_V = C' + g\,, \tag{11}$$

where C' is a $(\sqrt{n}/64)$-supported degree-n homogeneous $\Sigma\Pi\Sigma\Pi$ circuit with top fan-in at most that of C, and g is a polynomial each of whose monomials has a variable raised to the third or higher power.

The proof of the above lemma is our main technical contribution. It occupies most of Section 3.

The projection step in the definition of DPSP ensures that the polynomial g in eq. (11) satisfies $\mathrm{DPSP}_{\mathcal{M},\ell}(g) = 0$ for every choice of $\mathcal{M}$ and ℓ. Furthermore, the bound on the bottom fan-in of C' enables us to apply Lemma 2.1. Recalling that DPSP is a subadditive measure, we then obtain the following upper bound (setting $t = \sqrt{n}/64$ in Lemma 2.1).

▶ **Lemma 2.5** (Analogous to [12, Lemma 9]). *Let μ, q, C, S, and V be as in the previous lemma. Then the following event occurs with probability $1 - o(1)$. For all sets $\mathcal{M}$ of degree-k multilinear monomials and all $\ell \geqslant 0$ such that $\ell + k\sqrt{n}/64 \leqslant N/2$, we have*

$$\mathrm{DPSP}_{\mathcal{M},\ell}(C|_V) \leqslant S\binom{128\sqrt{n}+1}{k}\binom{N}{\ell + k\sqrt{n}/64}. \qquad ◀$$

Our final lower bound – Theorem 1.1 – then follows by combining Theorems 2.3 and 2.5 and using the parameter settings in eqs. (6)–(10).

3 Proof Details

Lemmas 2.1 and 2.3 are essentially restatements of the corresponding lemmas from previous works [10, 16]. It remains to prove Lemma 2.2 and 2.4.

3.1 A Well-Spaced Collection of Derivatives

We prove the first of these lemmas, which guarantees that the set $\mathcal{M}(V)$ of monomials with respect to which we shall be taking derivatives contains a large set of pairwise far monomials.

Proof of Lemma 2.2. Recall that $|V_1^{(h)}| = n^{bq}$ for each $h \in [k]$. Therefore $\mathcal{M}(V)$ maps bijectively to $V_1^{(1)} \times \cdots \times V_1^{(k)}$ in a natural way and thence to $[n^{bq}]^k$ in an artificial way. Let $\mathbb{K}$ be the largest finite field whose order is at most n^{bq}; note that $|\mathbb{K}| \geqslant n^{bq}/2$. Then $\mathbb{K}^k$ maps injectively (artificially) into $\mathcal{M}(V)$, via an injection ι, say.

Consider a Reed–Solomon code $\mathcal{C} \subseteq \mathbb{K}^k$ where the codewords are evaluations of polynomials in $\mathbb{K}[w]$ of degree at most $\sqrt{n}$ at k distinct points in $\mathbb{K}$. Then

$$|\mathcal{C}| = |\mathbb{K}|^{\sqrt{n}+1} \geqslant (n^{bq}/2)^{\sqrt{n}+1} \geqslant n^{\sqrt{n}},$$

since $bq > 1$. Pick $\mathcal{M}'(V)$ to be an arbitrary $n^{\sqrt{n}}$-sized subset of $\iota(\mathcal{C})$. The code $\mathcal{C}$, by construction, has Hamming distance at least $k - \sqrt{n}$. This directly translates to the desired monomial distance property for $\mathcal{M}'(V)$. ◀

Proof of Lemma 2.3. As we noted while stating this lemma, it essentially restates Lemma 8.9 from Kumar and Saraf [16]. The main concern is that for small b our set $\mathcal{M}'(V)$ above could be much smaller than their corresponding set. However, an examination of the proof of their Lemma 8.9 shows that the only properties of $\mathcal{M}'(V)$ that are needed are the size bound $|\mathcal{M}'(V)| \geqslant n^{\sqrt{n}}$ and the above farness property, both of which our Lemma 2.2 guarantees. ◀

3.2 The Main Lemma: Circuit Decomposition

We prove the remaining lemma which establishes the circuit decomposition indicated by eq. (11). The following technical lemma will be useful in the analysis.

▶ **Lemma 3.1.** *Given integers $0 < t \leqslant s' \leqslant s$ and sets A and B with $|A| = s$, $|B| = t$, and $B \subseteq A$, let R be a random subset of A chosen uniformly conditioned on $|R| = s'$. Then* $\Pr[B \subseteq R] \leqslant (s'/s)^t$. ◀

To start the proof of Lemma 2.4, consider C, an arbitrary N^μ-supported degree-n homogeneous $\Sigma\Pi\Sigma\Pi\Sigma$ circuit with size $S \leqslant n^{\varepsilon\sqrt{n}}$, on the variables $\mathcal{V}_{n,q}$, that computes the IMM polynomial $f_{n,q}$. Fix this C for the rest of this section. Expanding C into a formula, we have

$$C = \sum_i \prod_j \sum_r Q_{ijr}\,, \tag{12}$$

where each Q_{ijr} is a product of linear forms, each such linear form having at most N^μ variables. The proof now splits into two cases: the *thin case*, when the bottom fan-in is below $N^{1/4}$ and the *fat case*, when the bottom fan-in is $N^{1/4}$ or more.

3.3 The Thin Case

We consider the case when $0 \leqslant \mu < \frac{1}{4}$.

Let the random set V be drawn as described in Section 2.1. A monomial survives the restriction to V iff all its variables belong to V. Now Lemma 3.1 implies the following bounds for a monomial α with $|\operatorname{supp}(\alpha)| = t = O(\sqrt{n})$.

- For each h, if α has variables only from the first row of $Y^{(h)}$, then its survival probability is at most $m^{-(1-b)t} = n^{bqt}n^{-qt}$.
- For each h, if α has variables only from the ith row of $X^{(h,1)}$, then its survival probability is at most $n^{-(q-\lambda)t} = n^{\lambda t}n^{-qt}$.
- For each h, h' and i, with $2 \leqslant h' \leqslant B - 2\log n$, if α has variables only from the ith row of $X^{(h,h')}$, then its survival probability is at most $(2/m)^t = 2^t n^{-qt} = n^{t/\log n}n^{-qt}$.
- For each h, h' and i, with $h' > B - 2\log n$, if α has variables only from the ith row of $X^{(h,h')}$, then its survival probability is at most $m^{-t} = n^{-qt}$.

The hypotheses of Lemma 2.4 include the condition $bq \geqslant \lambda$, and eq. (10) implies $\lambda > 1$. Therefore the largest of these bounds is the first one, i.e., $n^{-(1-b)qt}$.

Since all the random subsets mentioned above are mutually independent, even if α's variables are spread out arbitrarily among multiple rows of multiple matrices, its survival probability is still at most $n^{-(1-b)qt}$.

Let $C|_V = \sum_i \prod_j \sum_r Q^{'}_{ijr}$ where Q'_{ijr} is a product of linear forms. Assume for some (i, j, r) that $\deg(Q'_{ijr}) = 2t$; if $\deg(Q'_{ijk}) > 2t$, then we only consider the product of the "first" $2t$ linear forms. Then the number of monomials in $Q^{'}_{ijk}$ is at most $(N^\mu)^{2t}$. Consider the *bad monomials* in $Q^{'}_{ijk}$, defined as ones where each variable has degree at most 2. These monomials have support at least t; the event that one of them survives is a *bad event.* If not a single bad monomial survives, then the circuit C decomposes into two circuits: a $2t$-supported degree-n homogeneous $\Sigma\Pi\Sigma\Pi$ circuit C' with top fan-in at most that of C, and a circuit g wherein each monomial has a variable raised to the third or higher power. Setting $t = \sqrt{n}/128$, this is exactly the decomposition we seek.

It remains to prove that the above bad event has probability $o(1)$. The probability that a bad monomial survives the random restriction is at most $n^{-(1-b)qt}$, as noted above. By a union bound, the bad event has probability at most

$$Sn^{-(1-b)qt}(N^\mu)^{2t} \leqslant n^{\varepsilon\sqrt{n}}n^{-(1-b)qt}n^{(2q+1)2\mu t} = \left(n^{128\varepsilon-(1-b)q+(2q+1)2\mu}\right)^t.$$

Since $t = \Theta(\sqrt{n})$, we can bound this by $n^{-\Omega(\sqrt{n})}$ by ensuring

$$128\varepsilon - (1-b)q + (2q+1)2\mu < 0\,. \tag{13}$$

Clearly it suffices to ensure that

$$(1-b)q = (2q+1)2\mu + 129\varepsilon\,. \tag{14}$$

Recall that we want $bq \geqslant \lambda \approx 2$ and $b \in (0,1)$. So we need $(1-b)q = q - bq \leqslant q - \lambda$, i.e.,

$$(2q+1)2\mu + 129\varepsilon \leqslant q - \lambda \iff q \geqslant \frac{\lambda + 129\varepsilon + 2\mu}{1 - 4\mu}\,, \tag{15}$$

where we have used $\mu < 1/4$.

We set q to be the smallest integer satisfying (15), then set b to satisfy (14). Then we do have $bq \geqslant \lambda$ as well as $b \in (0,1)$ as required.

3.4 The Fat Case

We consider the remaining case, when $\frac{1}{4} \leqslant \mu < \frac{1}{2}$.

We imagine the random restriction as being performed in two phases. Phase 1 chooses "large" random subsets of each row of each matrix in the IMM polynomial (for the Y-matrices, only the first row is used). Then Phase 2 chooses smaller random subsets, of the desired target sizes as in Section 2.1. The net effect is the same as the random restriction described in Section 2.1.

Phase 1

Let a be a parameter such that $0 < b < a < 1$; its value will be fixed in the later analysis.

We now define a distribution over subsets of $\mathcal{V}_{n,q}$ for sampling a random subset, W. Similar to V, W is also a union of random subsets $W_1^{(h)}$ and $W_i^{(h,h')}$ over h, h' and i, where $W_i^{(h)}$ and $W_i^{(h,h')}$ are subsets of variables in the ith row of $Y^{(h)}$ and $X^{(h,h')}$ respectively; these subsets are mutually independent. Each subset is chosen uniformly conditioned on its size being m^a. In the first phase, we consider a restriction to W, i.e., all variables outside W are set to zero.

Consider the probability that a monomial α, with $|\operatorname{supp}(\alpha)| = t = O(\sqrt{n})$, survives Phase 1. By Lemma 3.1, if α's variables come only from the first row of $Y^{(h)}$ for some particular h, or only from the ith row of some particular $X^{(h,h')}$, then its survival probability is at most $m^{-(1-a)t} = n^{-(1-a)qt}$. Since all the random subsets are mutually independent, even if α's variables are spread out arbitrarily among multiple rows of multiple matrices, its survival probability is still at most $n^{-(1-a)qt}$.

Phase 2

In this phase, we sample $V_1^{(h)} \subseteq W_1^{(h)}$ and $V_i^{(h,h')} \subseteq W_i^{(h,h')}$, uniformly and independently, subject to the cardinality constraints given in Section 2.1. We then define V as in eq. (5).

Let α be a monomial with $|\operatorname{supp}(\alpha)| = t = O(\sqrt{n})$. If the variables in α all come from a single set $W_1^{(h)}$ or $W_i^{(h,h')}$, then we can bound the probability of α surviving this second phase exactly as in Section 3.3, by using Lemma 3.1.

- If the variables come from $W_1^{(h)}$, the survival probability is at most $m^{-(a-b)t} = n^{bqt}n^{-aqt}$.
- If the variables come from $W_i^{(h,1)}$, the survival probability is at most $n^{-(aq-\lambda)t} = n^{\lambda t}n^{-aqt}$.

- If the variables come from $W_i^{(h,h')}$, where $2 \leqslant h' \leqslant B - 2\log n$, then the survival probability is at most $(2/m^{-a})^t = 2^t n^{-aqt} = n^{t/\log n} n^{-aqt}$.
- If the variables come from $W_i^{(h,h')}$, where $h' > B - 2\log n$, then the survival probability is at most $m^{-at} = n^{-aqt}$.

Again, recalling that $bq \geqslant \lambda > 1$, we see that the largest of these bounds is the first one, i.e., $n^{-(a-b)qt}$. Since all the random subsets mentioned above are mutually independent, even if α's variables are spread out arbitrarily among multiple rows of multiple matrices, its survival probability in phase 2 is still at most $n^{-(a-b)qt}$.

Effect of Phase 1 Restriction

We now analyze the effect of the phase 1 random restriction on the circuit C. Recall the expansion in eq. (12). Let $Q_{ijr} = \prod_u L_u$ where each L_u is a linear form with (w.l.o.g.) exactly N^μ terms.

Observe that the survival probability of each variable in C is at most $n^{-(1-a)q}$. Therefore, by linearity of expectation,

$$\mathbb{E}[\text{number of surviving terms in } L_u|_W] \leqslant N^\mu n^{-(1-a)q} \leqslant n^{(2q+1)\mu-(1-a)q} =: T\,. \tag{16}$$

We would like to bound the probability that the bottom fan-in of $C|_W$ greatly exceeds this bound T. This is not a straightforward Chernoff bound because the number of surviving terms in $L_u|_W$ is a sum of *dependent* indicator random variables. However, the dependency is of a benign sort. To see this, we recall some facts from probability theory, proved in, e.g., [3, Section 3.1] and [9].

▶ **Fact 3.2.** *Negative association of random variables is closed under products. That is, if $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are two independent collections of random variables that are separately negatively associated, then the union $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ is also negatively associated.*

▶ **Fact 3.3.** *Let a subset $R \subseteq [n]$ be drawn uniformly at random, conditioned on $|R| = k$, for some $k \leqslant n$, and let X_i be an indicator for the event $i \in R$. Then the collection $X_1, \ldots, X_n$ is negatively associated.*

▶ **Fact 3.4.** *The Chernoff–Hoeffding bounds apply as is to a sum of negatively associated random variables.*

Using these facts, we see that standard Chernoff bounds may be applied to the number of surviving terms in $L_u|_W$. Doing so and applying a union bound over all linear forms L_u gives us

$$\Pr\left[\text{bottom fan-in of } C|_W \geqslant (1+\sqrt{3})T\right] \leqslant Se^{-T} \leqslant n^{\varepsilon\sqrt{n}} e^{-T}\,.$$

For this probability to be $o(1)$, it suffices to have

$$\begin{aligned} &\varepsilon\sqrt{n}\ln n - n^{(2q+1)\mu-(1-a)q} \leqslant -\omega(1) \quad \text{(using eq. (16))} \\ \Leftarrow\ &(2q+1)\mu - (1-a)q \geqslant \tfrac{1}{2} + \Theta(1)\,. \end{aligned} \tag{17}$$

Effect of Phase 2 Restriction

After phase 1, with high probability the bottom fan-in of $C|_W$ is bounded by $(1+\sqrt{3})T$. Assuming that this bound holds, we analyze the effect of phase 2 on $C|_W$. This analysis is analogous to that in the thin case.

Let $C|_W = \sum_i \prod_j \sum_k Q'_{ijk}$ where Q'_{ijk} is a product of linear forms. Assume for some i, j, k, that $\deg(Q'_{ijk}) = 2t$; if $\deg(Q'_{ijk}) > 2t$, then we only consider the product of the first $2t$ linear forms. Then the number of monomials in Q'_{ijk} is at most $(1 + \sqrt{3})^{2t} T^{2t}$. Consider the *bad monomials* in Q'_{ijk}: those where each variable has degree at most 2. By our previous analysis, the probability that such a monomial survives phase 2 is at most by $n^{-(a-b)qt}$. If no bad monomial survives, then $C|_W$ indeed decomposes into two circuits as desired: a $2t$-supported degree-n homogeneous $\Sigma\Pi\Sigma\Pi$ circuit C' with top fan-in at most that of C and a circuit g wherein each monomial has a variable raised to a power $\geqslant 3$. We set $t = \sqrt{n}/128$ to obtain the decomposition we seek.

By a union bound over the at most S bad monomials, the probability that no bad monomial survives phase 2 – which we would like to bound by $o(1)$ – is at most

$$\begin{aligned} Sn^{-(a-b)qt}(1+\sqrt{3})^{2t}T^{2t} &\leqslant n^{\varepsilon\sqrt{n}} n^{-(a-b)qt}(1+\sqrt{3})^{2t}(n^{(2q+1)\mu-(1-a)q})^{2t} \\ &\leqslant \left(n^{128\varepsilon-(a-b)q-2(1-a)q+2(2q+1)\mu}\right)^t (1+\sqrt{3})^{2t}\,. \end{aligned}$$

Since $t = \Theta(\sqrt{n})$, we can bound this by $n^{-\Omega(\sqrt{n})}$ by ensuring that

$$128\varepsilon - (a-b)q - 2(1-a)q + 2(2q+1)\mu < 0\,. \tag{18}$$

Recall that we also want a, b, q to satisfy $bq \geqslant \lambda \approx 2$ as well as the phase 1 condition (17). Moreover, for the two-phase random restriction process to make sense, we want $0 < b < a < 1$. We claim that it suffices to choose a, b, and q such that

$$(1-a)q = (2q+1)\mu - 0.51\,, \quad \text{and} \tag{19}$$

$$(a-b)q = 2 \times 0.51 + 129\varepsilon\,. \tag{20}$$

Clearly condition (17) is satisfied. By adding (19) and 2×(20), we see that condition (18) is also satisfied. We will soon set q to a positive integer, satisfying $b < a$. By adding (19) and (20), we get

$$(2q+1)\mu + 0.51 + 129\varepsilon = (1-b)q\,. \tag{21}$$

We want $(1-b)q = q - bq \leqslant q - \lambda$, i.e.,

$$(2q+1)\mu + 0.51 + 129\varepsilon \leqslant q - \lambda \iff q \geqslant \frac{\lambda + 0.51 + 129\varepsilon + \mu}{1 - 2\mu}\,, \tag{22}$$

where we used $\mu < 1/2$. We set q to be the smallest integer satisfying condition (22). Next we set a and b satisfying eq. (19) and eq. (20) respectively. Now we want $a < 1$, or equivalently

$$(2q+1)\mu - 0.51 > 0 \iff q > \frac{1}{2}\left(\frac{0.51}{\mu} - 1\right)\,. \tag{23}$$

So we want

$$\begin{aligned} &\frac{1}{2}\left(\frac{0.51}{\mu} - 1\right) < \frac{\lambda + 0.51 + 129\varepsilon + \mu}{1 - 2\mu} \\ \iff\ &(\lambda + 0.51 + 129\varepsilon)2\mu + 2\mu^2 > (1-2\mu)(0.51 - \mu) \\ \iff\ &(\lambda + 0.51 + 129\varepsilon)2\mu + 2\mu^2 > 0.51 - 2.02\mu + 2\mu^2 \\ \iff\ &\mu > \frac{0.51}{2(\lambda + 0.51 + 129\varepsilon) + 2.02}\,. \end{aligned}$$

Since, $1 < \lambda < 2$ and $\mu \geqslant 1/4$, the above inequality does hold.

This completes the proof of Lemma 2.4.

4 Final Result and Discussion

We now put together the lemmas proven so far to obtain our final lower bound on homogeneous N^{μ}-supported $\Sigma\Pi\Sigma\Pi\Sigma$ circuits for IMM.

The following estimation will be useful in our calculations.

▶ **Lemma 4.1** (See, e.g., [7, Lemma 6]). *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*

$$\frac{(a+f)!}{(a-g)!} = a^{f+g} \cdot e^{\pm O\left((f+g)^2/a\right)}. \qquad ◀$$

With this, we are ready to prove our result.

▶ **Theorem 4.2** (Precise version of Main Theorem 1.1). *For every constant $\mu \in [0, 1/2)$, there is an integer $q > 0$ such that the following holds. Let C be a homogeneous N^{μ}-supported $\Sigma\Pi\Sigma\Pi\Sigma$ circuit that computes the N-variate degree-n IMM polynomial $f_{n,q}$ mentioned in Equation (3). Then C has size at least $N^{\Omega(\sqrt{n})}$.*

Proof. Suppose C has size S. Clearly we may choose an arbitrary small constant $\varepsilon > 0$ and proceed under the assumption that $S \leqslant n^{\varepsilon\sqrt{n}}$. So we make this assumption.

Let V be a random subset of $\mathcal{V}_{n,q}$, the variable set of $f_{n,q}$, sampled according to the distribution described in Section 2.1. By Lemma 2.5, for all sets $\mathcal{M}$ of degree-k multilinear monomials and all $\ell \geqslant 0$ such that $\ell + k\sqrt{n}/64 \leqslant N/2$,

$$\mathrm{DPSP}_{\mathcal{M},\ell}(C|_V) \leqslant S\binom{128\sqrt{n}+1}{k}\binom{N}{\ell + k\sqrt{n}/64}$$

with probability $1 - o(1)$.

By Lemma 2.3, with probability at least 0.9 there exists a set $\mathcal{M}''(V)$ of degree-k multilinear monomials such that

$$\mathrm{DPSP}_{\mathcal{M}''(V),\ell}\left(f_{n,q}|_V\right) \geqslant \frac{n^{\sqrt{n}}}{O(n^{\sqrt{n}/8}) \cdot n^{o(\sqrt{n})}} \left(\frac{N}{N-\ell}\right)^{\hat{n}} \binom{N-\hat{n}}{\ell}.$$

for all $\ell > 0$. Hence with non-zero probability both these bounds hold. Comparing the above two bounds, and using parameters $k = 32\sqrt{n}$ and $\ell = \frac{N}{2}\left(1 - \frac{\ln n}{\Gamma\sqrt{n}}\right)$ from eq. (6) and (8), we get

$$\begin{aligned}
S &\geqslant \frac{\frac{n^{\sqrt{n}}}{O\left(n^{\sqrt{n}/8}\right)\cdot n^{o\left(\sqrt{n}\right)}} \cdot \left(\frac{N}{N-\ell}\right)^{\hat{n}} \cdot \binom{N-\hat{n}}{\ell}}{\binom{128\sqrt{n}+1}{32\sqrt{n}} \cdot \binom{N}{\ell+32\sqrt{n}\cdot\frac{\sqrt{n}}{64}}} \\
&= \frac{n^{\sqrt{n}} \cdot \left(\frac{N}{N-\ell}\right)^{\hat{n}}}{O\left(n^{\sqrt{n}/8}\right) \cdot n^{o\left(\sqrt{n}\right)} \cdot 2^{O\left(\sqrt{n}\right)}} \cdot \frac{\binom{N-\hat{n}}{\ell}}{\binom{N}{\ell+0.5n}} && \text{since } \binom{128\sqrt{n}+1}{32\sqrt{n}} = 2^{\Theta(\sqrt{n})} \\
&= \frac{\left(\frac{N}{\ell}\right)^{\hat{n}}}{O\left(n^{\sqrt{n}/8}\right) \cdot n^{o\left(\sqrt{n}\right)} \cdot 2^{O\left(\sqrt{n}\right)}} \cdot \frac{(N-\hat{n})!}{N!} \cdot \frac{(N-\ell-0.5n)!}{(N-\ell-\hat{n})} \cdot \frac{(\ell+0.5n)!}{\ell!} && \text{using (9)} \\
&\approx \frac{1}{O\left(n^{\sqrt{n}/8}\right) \cdot n^{o\left(\sqrt{n}\right)} \cdot 2^{O\left(\sqrt{n}\right)}} \cdot \left(\frac{N}{\ell}\right)^{\hat{n}} \cdot \frac{1}{N^{\hat{n}}} \cdot (N-\ell)^{\hat{n}-0.5n} \cdot \ell^{0.5n} && \text{by Thm 4.1} \\
&= \frac{1}{O\left(n^{\sqrt{n}/8}\right) \cdot n^{o\left(\sqrt{n}\right)} \cdot 2^{O\left(\sqrt{n}\right)}} \cdot (N-\ell)^{\hat{n}-0.5n} \cdot \ell^{0.5n-\hat{n}}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{O\left(n^{\sqrt{n}/8}\right)\cdot n^{o\left(\sqrt{n}\right)}\cdot 2^{O\left(\sqrt{n}\right)}}\cdot\left(\frac{N-\ell}{\ell}\right)^{\hat{n}-0.5n} \\
&= \frac{1}{O\left(n^{\sqrt{n}/8}\right)\cdot n^{o\left(\sqrt{n}\right)}\cdot 2^{O\left(\sqrt{n}\right)}}\cdot\left(\frac{N-\frac{N}{2}\left(1-\frac{\ln n}{\Gamma\sqrt{n}}\right)}{\frac{N}{2}\left(1-\frac{\ln n}{\Gamma\sqrt{n}}\right)}\right)^{\hat{n}-0.5n} && \text{using (8)} \\
&= \frac{1}{O\left(n^{\sqrt{n}/8}\right)\cdot n^{o\left(\sqrt{n}\right)}\cdot 2^{O\left(\sqrt{n}\right)}}\cdot\left(\frac{1+\frac{\ln n}{\Gamma\sqrt{n}}}{1-\frac{\ln n}{\Gamma\sqrt{n}}}\right)^{\hat{n}-0.5n} \\
&\approx \frac{1}{O\left(n^{\sqrt{n}/8}\right)\cdot n^{o\left(\sqrt{n}\right)}\cdot 2^{O\left(\sqrt{n}\right)}}\cdot e^{2\cdot\frac{\ln n}{\Gamma\sqrt{n}}\cdot(\hat{n}-0.5n)} \\
&= \frac{1}{O\left(n^{\sqrt{n}/8}\right)\cdot n^{o\left(\sqrt{n}\right)}\cdot 2^{O\left(\sqrt{n}\right)}}\cdot n^{\frac{2}{\Gamma\sqrt{n}}\left(n-64\sqrt{n}-0.5n\right)} && \text{using (7)} \\
&= \frac{1}{O\left(n^{\sqrt{n}/8}\right)\cdot n^{o\left(\sqrt{n}\right)}\cdot 2^{O\left(\sqrt{n}\right)}}\cdot n^{\frac{2}{\Gamma}\left(\sqrt{n}-64-0.5\sqrt{n}\right)}\,.
\end{aligned}$$

Using the estimate for Γ from (9), we obtain $S \geqslant n^{\Omega\left(\sqrt{n}\right)} = N^{\Omega(\sqrt{n})}$, as desired. ◀

4.1 Remarks and Discussion

Notably, our lower bound only applies to circuits with bottom fan-in below $\sqrt{N}$ – or rather, at most $N^{1/2-\Theta(1)}$. This is a somewhat strong restriction because in a general depth-five circuit on N variables this fan-in could have been as high as N. In particular it is a stronger restriction than in the Kayal–Saha lower bound for certain Nisan–Wigderson polynomials (NW polynomials) [12], where this bottom fan-in had to be at most $N^{1-\Theta(1)}$.

On the positive side, our lower bound works for arithmetic circuits over an arbitrary field, whereas the Kayal–Saha bound requires characteristic zero. Ultimately, this is because the technique for lower-bounding DPSP that they use (which is borrowed from Kayal *et al.* [10]) hinges on an operator-theoretic interpretation of matrix rank. In contrast, the DPSP lower bound that we use (borrowed from Kumar and Saraf [16]) is proven using counting alone.

It is worth understanding why our result hits a barrier at bottom fan-in around $N^{1/2}$. The random restriction used in this analysis retains at least one variable from almost every row of every matrix in the IMM polynomial. Therefore, it reduces the variable set from size N to size slightly more than $N^{1/2}$ (the "slightly" is in fact contingent on making q very large), and this is not a severe enough random restriction to give us the required circuit decomposition. More concretely, satisfying Equation (21), even in the extreme setting $b = 0$, forces $q \to \infty$ as $\mu \to 1/2$. It could be that an even more severe random restriction is worth considering, but proving a good DPSP lower bound for IMM polynomials so restricted seems unlikely to proceed along the lines of the Kumar–Saraf argument. Whether our size lower bound still holds with the bottom fan-in allowed to reach up to $N^{1-\Theta(1)}$, or even N (which is the general case) is the most immediate and natural open question.

Acknowledgments. The second author would like to thank Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi for several stimulating discussions.

References

1 Manindra Agrawal and V Vinay. Arithmetic circuits: A chasm at depth four. In *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 67–75, 2008.

2 Stuart J. Berkowitz. On computing the determinant in small parallel time using a small number of processors. *Inform. Process. Lett.*, 18(3):147–150, 1984.

3 Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms.* Cambridge University Press, 2009.

4 Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *Proc. 46th Annual ACM Symposium on the Theory of Computing*, pages 128–135, 2014.

5 Dima Grigoriev and Marek Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *Proc. 30th Annual ACM Symposium on the Theory of Computing*, pages 577–582, 1998.

6 Dima Grigoriev and Alexander Razborov. Exponential lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Applicable Algebra in Engineering, Communication and Computing*, 10(6):465–487, 2000.

7 Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the chasm at depth four. In *Proc. 28th Annual IEEE Conference on Computational Complexity*, pages 65–73, 2013.

8 Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth three. In *Proc. 54th Annual IEEE Symposium on Foundations of Computer Science*, pages 578–587, 2013.

9 Kumar Joag-Dev and Frank Proschan. Negative association of random variables, with applications. *Ann. Stat.*, 11(1):286–295, 1983.

10 Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An exponential lower bound for homogeneous depth four arithmetic formulas. In *Proc. 55th Annual IEEE Symposium on Foundations of Computer Science*, 2014.

11 Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. Super-polynomial lower bounds for depth-4 homogeneous arithmetic formulas. In *Proc. 46th Annual ACM Symposium on the Theory of Computing*, pages 119–127, 2014.

12 Neeraj Kayal and Chandan Saha. Lower bounds for depth three arithmetic circuits with small bottom fanin. Technical Report TR14-089, ECCC, 2014.

13 Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *Proc. 46th Annual ACM Symposium on the Theory of Computing*, pages 146–153, 2014.

14 Neeraj Kayal and Ramprasad Saptharishi. *A selection of lower bounds for arithmetic circuits*, pages 77–115. Springer Verlag, April 2014.

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

16 Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. In *Proc. 55th Annual IEEE Symposium on Foundations of Computer Science*, 2014.

17 Noam Nisan and Avi Wigderson. Hardness vs. randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994.

18 Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. In *Proc. 36th Annual IEEE Symposium on Foundations of Computer Science*, pages 16—25, 1995.

19 Ramprasad Saptharishi. Recent progress on arithmetic circuit lower bounds. *Bulletin of the EATCS*, 114, 2014.

20 Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Found. Trends Theor. Comput. Sci.*, 5(3-4):207–388, 2010.

21 Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *Proc. 38th International Symposium on Mathematical Foundations of Computer Science*, pages 813–824, 2013.

22 L. G. Valiant. Completeness classes in algebra. In *Proc. 11th Annual ACM Symposium on the Theory of Computing*, pages 249–261, 1979.

23 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.