Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 26 (2013)

# Arithmetic circuits: A chasm at depth ~~four~~ three

Ankit Gupta
Microsoft Research India
t-ankitg@microsoft.com

Pritish Kamath
Microsoft Research India
t-pritk@microsoft.com

Neeraj Kayal
Microsoft Research India
neeraka@microsoft.com

Ramprasad Saptharishi
Chennai Mathematical Institute
ramprasad@cmi.ac.in

February 27, 2013

## Abstract

We show that, over $\mathbb{Q}$, if an $n$-variate polynomial of degree $d = n^{O(1)}$ is computable by an arithmetic circuit of size $s$ (respectively by an algebraic branching program of size $s$) then it can also be computed by a depth three circuit (i.e. a $\Sigma\Pi\Sigma$-circuit) of size $\exp(O(\sqrt{d \log d \log n \log s}))$ (respectively of size $\exp(O(\sqrt{d \log n \log s}))$). In particular this yields a $\Sigma\Pi\Sigma$ circuit of size $\exp(O(\sqrt{d} \cdot \log d))$ computing the $d \times d$ determinant $\mathsf{Det}_d$. It also means that if we can prove a lower bound of $\exp(\omega(\sqrt{d} \cdot \log^{3/2} d))$ on the size of any $\Sigma\Pi\Sigma$-circuit computing the $d \times d$ permanent $\mathsf{Perm}_d$ then we get superpolynomial lower bounds for the size of any arithmetic circuit computing $\mathsf{Perm}_d$. We then give some further results pertaining to derandomizing polynomial identity testing and circuit lower bounds.

The $\Sigma\Pi\Sigma$ circuits that we construct have the property that (some of) the intermediate polynomials have degree much higher than $d$. Indeed such a counterintuitive construction is unavoidable - it is known (cf. Appendix A) that in any $\Sigma\Pi\Sigma$ circuit $C$ computing either $\mathsf{Det}_d$ or $\mathsf{Perm}_d$, if every multiplication gate has fanin at most $d$ (or any constant multiple thereof) then $C$ must have size at least $\exp(\Omega(d))$.

ISSN 1433-8092

# 1 Introduction

**Arithmetic Circuits.** The most natural way to compute a polynomial function $f(x_1, x_2, \ldots, x_n)$ starting from its inputs $x_1, x_2, \ldots, x_n$ is via a sequence of basic arithmetic operations consisting of addition, multiplication and subtraction. Such a computation can be visualized as an arithmetic circuit. We typically allow the incoming edges to a + gate to be labelled with constants from the underlying field $\mathbb{F}$ so that a + gate can in fact compute an arbitrary $\mathbb{F}$-linear combination of its inputs. Two relevant complexity measures for an arithmetic circuit are its size (the total number of arithmetic operations involved) and its depth (the maximum length of a path from an input to the output). The goal here is to understand the optimal complexity (in terms of size and depth) of computing a given polynomial family. Two closely related families of polynomials, the determinant and the permanent defined as

$$\begin{aligned}\mathsf{Det}_d &= \sum_{\sigma \in S_d} \mathrm{sign}(\sigma) \cdot \prod_{i=1}^{d} x_{i,\sigma(i)} \\ \mathsf{Perm}_d &= \sum_{\sigma \in S_d} \prod_{i=1}^{d} x_{i,\sigma(i)}\end{aligned}$$

are of particular interest as they feature in many different areas of mathematics and computer science. Although these two polynomials look very similar, they have strikingly different complexities. The determinant and permanent are in fact complete problems for the classes VP and VNP respectively, which are algebraic analogues of P and NP [Val79]. A grand challenge in this direction is to show that $\mathsf{Perm}_d$ cannot be computed by arithmetic circuits of polynomial size.

**Depth Reduction.** Circuits with low depth correspond to computations which are highly parallelizable and therefore it is natural to try to minimize the depth of a circuit while allowing the size to increase somewhat. Csanky [Csa76] showed that $\mathsf{Det}_d$ can be computed by circuits of size $d^{O(1)}$ *having only* $(\log d)^{O(1)}$ *depth.* Subsequently Valiant, Skyum, Berkowitz and Rackoff [VSBR83] discovered a remarkable generalization. They showed that if a polynomial $f$ of degree $d$ can be computed by a circuit of size $s$ then it can in fact be computed by a circuit of depth $O(\log d \cdot \log s)$ and size $s^{O(1)}$. Pushing this line of investigation of size-depth tradeoffs further, recent work has considered reduction to circuits of even smaller depth while allowing the addition and multiplication gates to have arbitrary (unbounded) fanin. In this direction, the work of Agrawal and Vinay [AV08] and a subsequent strengthening by Koiran [Koi12] showed that if $f$ has circuits of size $s = d^{O(1)}$ then $f$ can in fact be computed by depth four circuits of size $2^{O(\sqrt{d} \cdot \log^2 d)}$ [1]. Despite the large blowup, the reason that such

[1] A simple but nonconstructive counting argument shows that over any field $\mathbb{F}$, most $n$-variate polynomials of degree $d$ require circuits of size $\sqrt{\binom{n+d}{d}}$

reductions to constant depth circuits are interesting is that these reductions 'explain' the lack of progress towards lower bounds even for constant-depth arithmetic circuits. [2] Viewed optimistically, such a reduction entails that one *merely* needs to prove a (good enough) lower bound for depth four circuits in order to prove lower bounds for arbitrary circuits. Indeed, motivated by this, we recently proved lower bounds for depth four circuits which comes very close to the threshold required by Koiran[3]. In a similar spirit Raz [Raz10] showed that close-to-optimal lower bounds for small degree tensors[4] imply superpolynomial lower bounds on formula size[5].

**Depth Three Circuits.** Being the shallowest nontrivial subclass of arithmetic circuits, depth three arithmetic circuits, also denoted as $\Sigma\Pi\Sigma$ circuits, have been intensely investigated. Such a circuit $C$ computes a polynomial in the following manner:

$$C(\mathbf{x}) = \sum_{i=1}^{s} \prod_{j=1}^{d_i} \ell_{ij}(\mathbf{x}), \tag{1}$$

where each $\ell_{ij}(\mathbf{x})$ is an affine form over the input variables. $\Sigma\Pi\Sigma$ circuits (more specifically tensors) arise naturally in the investigation of the complexity of polynomial multiplication and matrix multiplication[6]. Moreover, the optimal formula/circuit for some well known families of polynomials are in fact depth three circuits. In particular, the best known circuit for computing the permanent $\mathsf{Perm}_d$ is known as Ryser's

[2] Note that in contrast to arithmetic circuits, exponential lower bounds are known for constant depth boolean circuits with $\wedge, \vee$ gates of unbounded fanin.

[3] Specifically, [GKKS13] shows a lower bound of $\exp(\sqrt{d})$ on the size of $\Sigma\Pi^{[\sqrt{d}]}\Sigma\Pi^{[\sqrt{d}]}$ computing $\mathsf{Perm}_d$. In comparison, Koiran shows that a lower bound of $\exp(\omega(\sqrt{d}\log^2 d))$ on the size of $\Sigma\Pi^{[\sqrt{d}]}\Sigma\Pi^{[\sqrt{d}]}$ computing $\mathsf{Perm}_d$ entails superpolynomial lower bounds for general circuits computing $\mathsf{Perm}_d$.

[4] a subclass of $\Sigma\Pi\Sigma$ also known as set-multilinear $\Sigma\Pi\Sigma$ circuits

[5] Specifically, Raz showed that for $d = d(n) \le \frac{\log n}{\log\log n}$, any explicit example of a tensor $A : [n]^d \mapsto \mathbb{C}$ with tensor rank $\ge n^{d(1-o(1))}$ implies an explicit superpolynomial lower bound for the size of general arithmetic formulas. Because of the restriction on the degree $d$, Raz's result does not seem to be applicable to the permanent

[6] For example it can be shown that the product of two $n \times n$ matrices can be computed with $\tilde{O}(n^\omega)$ arithmetic operations if and only if the polynomial

$$M_n = \sum_{i\in[n]} \sum_{j\in[n]} \sum_{k\in[n]} x_{ij} \cdot y_{jk} \cdot z_{ki}$$

can be computed by a $\Sigma\Pi\Sigma$ circuit where the top fanin $s$ is at most $\tilde{O}(n^\omega)$.

formula [Rys63] which is a (homogeneous[7]) depth three circuit of size $O(d^2 \cdot 2^d)$[8]. For more on $\Sigma\Pi\Sigma$ circuits, we refer the reader to the thesis of Shpilka [Shp01] and the references therein.

**Circuit Lower bounds.** While there has been significant progress in upper bounds, progress in proving lower bounds for arithmetic circuits has been much slower. This is generally considered to be one of the most challenging problems in computer science. The difficulty of the problem has led researchers to focus on natural subclasses of arithmetic circuits. We refer the interested reader to the recent surveys by Shpilka and Yehudayoff [SY10] and Chen, Kayal and Wigderson [CKW11] for more on lower bounds for various subclasses of arithmetic circuits. Bounded depth circuits being one such natural subclass has received a lot of attention. The simplest nontrivial such subclass is that of $\Sigma\Pi\Sigma$ circuits. Nisan and Wigderson [NW97] showed that over any field $\mathbb{F}$, any homogeneous $\Sigma\Pi\Sigma$ circuit computing the determinant $\mathsf{Det}_d$ must be of size $2^{\Omega(d)}$. Grigoriev and Karpinski [GK98], and Grigoriev and Razborov [GR00] showed that any $\Sigma\Pi\Sigma$ arithmetic circuit over any *fixed* finite field computing $\mathsf{Det}_d$ must be of size at least $2^{\Omega(d)}$. Raz and Yehudayoff give $2^{\Omega(d)}$ lower bounds for *multilinear* $\Sigma\Pi\Sigma$ circuits[9]. But without any restrictions, even a superpolynomial lower bound for $\Sigma\Pi\Sigma$ circuits (over an infinite field) has remained ellusive. The best known lower bound in the general $\Sigma\Pi\Sigma$ case is the quadratic lower bound due to Shpilka and Wigderson [SW01]. Avi Wigderson [Wig07] highlighted this frontier in arithmetic complexity by concluding his plenary talk on 'P, NP and mathematics' at ICM 2006 with the problem of proving superpolynomial lower bounds for $\Sigma\Pi\Sigma$ circuits computing the determinant.

**Our contribution.** The $2^{\Omega(d)}$ lower bounds for various restrictions of $\Sigma\Pi\Sigma$ circuits mentioned above seemed to suggest (at least to us) that any $\Sigma\Pi\Sigma$ circuit computing the determinant $\mathsf{Det}_d$ needs to be of size at least $2^{\Omega(d)}$. Surprisingly, we show that this is not true - there do indeed exist much smaller $\Sigma\Pi\Sigma$ circuits computing the determinant. Specifically we show that over $\mathbb{Q}$, the field of rational numbers, there exists a $\Sigma\Pi\Sigma$ circuit of size $2^{O(\sqrt{d}\cdot\log d)}$ computing the determinant. To the best of our knowledge, no $\Sigma\Pi\Sigma$ circuit of size smaller than even $2^{O(d\cdot\log d)}$ was previously known. More generally, we show:

[7] Recall that a multivariate polynomial is said to be homogeneous if all its monomials have the same total degree. An arithmetic circuit is said to be *homogeneous* if the polynomial computed at every internal node of the circuit is a homogeneous polynomial. It is a folklore result (cf. the survey by Shpilka and Yehudayoff [SY10]) that as far as computation by polynomial-sized arithmetic circuits of unbounded depth is concerned one can assume without loss of generality that the circuit is homogeneous. Specifically, if a homogeneous polynomial $f$ of degree $d$ can be computed by an (unbounded depth) arithmetic circuit of size $s$, then it can also be computed by a *homogeneous* circuit of size $O(d^2 \cdot s)$.

[8] To the best of our knowledge, no $\Sigma\Pi\Sigma$ circuit of size smaller than $2^{O(d\log d)}$ computing $\mathsf{Det}_d$ was previously known.

[9] The results of Raz and Yehudayoff are more general.

**Theorem 1.1.** *Let $f(\mathbf{x}) \in \mathbb{Q}[\mathbf{x}]$ be an $n$-variate polynomial of degree $d = n^{O(1)}$ computed by an arithmetic circuit of size $s$. Then it can also be computed by a $\Sigma\Pi\Sigma$ circuit of size $2^{O(\sqrt{d \log n \log s \log d})}$. Further, if $f$ can be computed by an ABP of size $s$ then it can also be computed by a $2^{O(\sqrt{d \log n \log s})}$ sized $\Sigma\Pi\Sigma$ circuit.*

Note that $\mathsf{Det}_d$ can be computed by an ABP of size $d^{O(1)}$ (cf. [AJMV98]) and hence we immediately get the $\Sigma\Pi\Sigma$ circuit for it as mentioned above. We note here that in particular, the above theorem shows that a $\Sigma\Pi\Sigma$ lower bound of $2^{\Omega(d \log n)}$ for any explicit $n$-variate polynomial $f$ of degree $d$ entails a $2^{\Omega(d \log n)}$ lower bound on the size of any ABP (or arithmetic formula) computing $f$.

**Comparison with prior work.** Prior work by [AV08] and [Koi12] reduce the depth to four and we build upon their work to reduce the depth even further to three. Most closely related is the work by Raz [Raz10]. While both our work and [Raz10] have the same high-level message, namely that strong enough lower bounds for $\Sigma\Pi\Sigma$ circuits imply more general superpolynomial lower bounds, we make here significant quantitative improvements[10]. We state this from the perspective of obtaining lower bounds. Firstly, Raz's result would yield only superpolynomial *formula* lower bounds while ours can yield circuit lower bounds. Secondly, Raz requires the degree of the output polynomial to be rather small: $d \leq \frac{\log n}{\log \log n}$ while our results are valid for much larger $d$, say $d = n^{\Omega(1)}$. Most importantly, Raz requires almost optimal $\Sigma\Pi\Sigma$ lower bounds of about $2^{(d \log n)\cdot(1-o(1))}$ while for us a much weaker lower bound of $2^{\omega(\sqrt{d} \log^{3/2} n)}$ will suffice. $\Sigma\Pi\Sigma$ circuits have been intensely investigated - initially for their connection to tensor rank and more recently as a special case of the polynomial identity testing problem. Given this intense investigation of $\Sigma\Pi\Sigma$ circuits, it is natural to wonder as to why these results were not obtained before. We feel that this may be because our construction is significantly counterintuitive - the intermediate terms are of degree much higher than the degree of the output polynomial and moreover we need the field to be have zero or large characteristic. Finally, we remark here that as a tool for proving lower bounds, our result is somewhat incomparable to [Koi12] - while the reduction of the depth further should facilitate the task of proving lower bounds, the fact that the resulting $\Sigma\Pi\Sigma$ circuit is nonhomogeneous hinders it. Overall, it is not clear to us as to which of these two kinds of circuits is a better starting point.

**Further Results.** In section 5 we then present some further consequences and reductions. Following [AV08] we show that a blackbox derandomization of polynomial identity testing for $\Sigma\Pi\Sigma$ circuits leads to a quasipolynomial identity test for general circuits. We also explore the relation of $\Sigma\Pi\Sigma$ circuits with a subclass of circuits that arise in our proof.

[10] The motivation of Raz's work was somewhat different: Raz was interested in giving a tight analysis of the blowup in size when a formula is homogenized.

## 2 Proof Overview

**Quick Sketch.** The depth reduction proceeds through a series of transformations to the circuit - first decreasing the depth, then replacing the multiplication ($\times$) gates with exponentiation gates at the expense of increasing the depth slightly and then decreasing the depth once again to three. We will flesh out this quick sketch after introducing some relevant notation.

**Notation.** Bounded depth arithmetic circuits consist of alternating layers of addition and multiplication gates. We will denote an arithmetic circuit of depth $d$ by a sequence of $d$ symbols wherein each symbol (either $\Sigma$ or $\Pi$) denotes the nature of the gates at the corresponding layer and the leftmost symbol indicates the nature of the gate at the output layer. For example, a $\Pi\Sigma$ circuit with $n$ input variables computes a polynomial in the following way:

$$C(\mathbf{x}) \quad = \quad \prod_{i \in [d]} \left( \sum_{j \in [n]} a_{ij} x_j + a_{i0} \right) \quad \text{where each } a_{ij} \in \mathbb{F} \text{ is a field element,}$$

while a $\Sigma\Pi\Sigma$ circuit computes a polynomial as in equation (1). Some of the intermediate circuits that we construct will have the feature that all the incoming edges to a multiplication gate come from a single gate $g$ (thus computing $g^e$, if there are $e$ wires entering the multiplication gate). We will refer to such circuits as *powering circuits* and use $\wedge$ instead of $\Pi$ to denote a layer of multiplication gates in such circuits. So for example, a $\Sigma\wedge\Sigma$ circuit computes a polynomial in the following manner:

$$C(\mathbf{x}) \quad = \quad \sum_{i \in [s]} \ell_i(\mathbf{x})^{e_i} \quad \text{where each } \ell_i \in \mathbb{F}[\mathbf{x}] \text{ is an affine form.}$$

In doing the transformations it is useful to keep track of the fanin to various gates, especially multiplication gates. Towards this end, we extend the above notation and allow integer superscripts on $\Pi$ symbols (respectively $\Sigma$ and $\wedge$ symbols) which denotes an upper bound on the fanin of any gate in the corresponding layer. So for example a $\Sigma\Pi^{[a]}\Sigma\Pi^{[b]}$ circuit computes a polynomial of the following form:

$$C \quad = \quad \sum_{i \in [s]} \prod_{j \in [a]} Q_{ij} \quad \text{with } \deg Q_{ij} \leq b \text{ for all } i \in [s] \text{ and } j \in [a]$$

while a $\Sigma\wedge^{[a]}\Sigma$ circuit computes a polynomial a following manner:

$$C(\mathbf{x}) \quad = \quad \sum_{i \in [s]} \ell_i(\mathbf{x})^{e_i} \quad \text{where each } e_i \leq a \text{ and each } \ell_i \in \mathbb{F}[\mathbf{x}] \text{ is an affine form.}$$

With this notation in place we are ready to give a more detailed overview.

**Proof Overview.** Let $f(\mathbf{x}) \in \mathbb{Q}[\mathbf{x}]$ be a polynomial of degree $d$ computed by an arithmetic circuit $C$ of size $s$. The first step is to obtain from $C$ a depth-4 circuit $C_1$ computing $f$. Specifically, $C_1$ is a $\Sigma\Pi^{[a]}\Sigma\Pi^{[d/a]}$ circuit (for a suitably chosen $a$) of size $s_1 = 2^{O(\sqrt{d \log d \log s \log n})}$. This is achieved via Koiran's [Koi12] strengthening of the Agrawal-Vinay [AV08] depth reduction.

The second step is to use $C_1$ to obtain a depth-5 *powering circuit* $C_2$. Specifically, $C_2$ is a $\Sigma\wedge^{[a]}\Sigma\wedge^{[d/a]}\Sigma$ circuit of size $s_2 = 2^{O(\sqrt{d \log d \log s \log n})}$. The main ingredient of this step is a lemma of Fischer [Fis94] showing how a monomial can be computed by a $\Sigma\wedge\Sigma$ circuit. In other words, Fischer's lemma shows to compute a product as a sum of powers of sums. By applying this lemma to every multiplication gate, we can convert the circuit $C_1$ to a $\Sigma\wedge\Sigma\wedge\Sigma$ circuit $C_2$. It is worth noting that Fischer's lemma is not true over fields of small characteristic.

The final step is to convert the $\Sigma\wedge\Sigma\wedge\Sigma$ circuit $C_2$ to a $\Sigma\Pi\Sigma$ circuit $C_3$, and this is done by invoking the "duality trick" of Saxena [Sax08] and factoring the resulting univariate polynomials to get a $\Sigma\Pi\Sigma$ circuit over an extension of $\mathbb{Q}$ of not too large degree. Finally, from this we derive the required $\Sigma\Pi\Sigma$ circuit over $\mathbb{Q}$. Overall in the last step we increase the degree of the intermediate polynomials substantially - to $2^{O(\sqrt{d \log d \log n \log s})}$. Our construction ensures that all the high ($> d$) degree monomials so generated ultimately cancel out. The resulting $\Sigma\Pi\Sigma$ circuit $C_3$ is of size $s_3 = 2^{O(\sqrt{d \log d \log n \log s})}$.

# 3 Preliminaries

This section would deal with the preliminaries required for the rest of the paper. As usual $[n]$ denotes the set of first $n$ positive integers. For the ease of book-keeping, we define the size of a circuit as the number of wires[11] in the circuit.

**Algebraic Branching Programs (ABPs).** An ABP is a layered graph with edges going from layer $i$ to $i+1$. Every edge $e$ is labelled by a linear polynomial $\ell_e$. The first layer has only one vertex called the *source* and the last layer has only one vertex called the *sink*. For any path $\gamma = (e_1, \ldots, e_d)$ from source to sink, the weight of $\gamma$ is defined as $\mathrm{wt}(\gamma) = \ell_{e_1} \ldots \ell_{e_d}$. The ABP is said to compute the polynomial $\sum_\gamma \mathrm{wt}(\gamma)$, where $\gamma$ runs over all source-sink paths.

An ABP is said to be *homogeneous* if all edge labels are homogeneous linear forms, and naturally computes a homogeneous polynomial.

[11] One could alternatively define the size of a circuit as the number of nodes but defining it this way is more natural for us as it simplifies the bookkeeping - for example in tracking the size of an exponentiation gate.

We shall say that a (possibly non-homogeneous) polynomial is computed by a homogeneous ABP if each of its homogeneous parts can be computed by a homogenous ABP. It is a well-known fact (cf. [SY10]) that if a degree $d$ polynomial can be computed by a size $s$ (possibly non-homogeneous) ABP, then it can be computed by a homogeneous ABP of size $sd^2$.

The following conversion from circuits to ABPs is attributed by Koiran [Koi12] to Malod and Portier [MP08] but can also be easily deduced directly from [VSBR83].

**Lemma 3.1.** *Let $f$ be a polynomial of degree $d$ computed by a circuit of size $s$. Then there is a homogeneous ABP of depth $d$ and size $s' = 2^{O(\log s \cdot \log d)}$ computing $f$.*

The following bound on a binomial coefficients is an easy consequence of Stirling's approximation.

**Lemma 3.2.** *(cf. [AV08], lemma 2.2 for a proof) For any $n, k$,*

$$\binom{n+k}{k} \quad = \quad 2^{O(k \log n)}$$

# 4 The depth-reduction

Let us recall the statement of our main result as applicable for ABPs.

**Theorem 1.1 (restated).** *Let $f(\mathbf{x}) \in \mathbb{Q}[\mathbf{x}]$ be an $n$-variate polynomial of degree $d = n^{O(1)}$ computed by an ABP of size $s$. Then it can also be computed by a $\Sigma\Pi\Sigma$ circuit of size $2^{O(\sqrt{d \log n \log s})}$.*

We first observe that the corresponding depth reduction for general circuits (as given in the statement of this theorem in section 1) follows immediately from the above statement via an application of lemma 3.1 - we first convert the given circuit to a slightly larger ABP and then apply the depth reduction for ABPs. The rest of this section is devoted to the proof of the above theorem. So let $f$ be an $n$-variate polynomial of degree $d$ computed by an ABP of size $s$. As mentioned in Section 2, the depth reduction roughly proceeds through the following steps:

| | | | |
|---|---|---|---|
| **Step 1:** | Algebraic Branching Programs | $\longrightarrow$ | $\Sigma\Pi^{[a]}\Sigma\Pi^{[d/a]}$ circuit (for a suitably chosen $a$) |
| **Step 2:** | $\Sigma\Pi^{[a]}\Sigma\Pi^{[d/a]}$ circuit | $\longrightarrow$ | $\Sigma\wedge^{[a]}\Sigma\wedge^{[d/a]}\Sigma$ circuit |
| **Step 3:** | $\Sigma\wedge^{[a]}\Sigma\wedge^{[d/a]}\Sigma$ circuit | $\longrightarrow$ | $\Sigma\Pi\Sigma$ circuit |

In the rest of this section we shall provide the details on how to perform each of the above steps while keeping track of the loss in size incurred at each step.

## 4.1 Step 1: ABPs to $\Sigma\Pi^{[a]}\Sigma\Pi^{[d/a]}$ circuit

The first step is a direct consequence of the depth-reduction result of Koiran [Koi12].

**Theorem 4.1** ([Koi12]). *Let $f$ be an $n$-variate polynomial of degree $d$ computed by a homogeneous ABP of size $s$. Then, for all $a$ there is an equivalent homogeneous $\Sigma\Pi^{[a]}\Sigma\Pi^{[(d/a)]}$ circuit computing $f$ of size $s^a + (s^2 d) \cdot \binom{n+(d/a)}{(d/a)}$.*

We present a proof of it here as the statement above is *slightly* different from Koiran's (though can be easily seen from his proof).

*Proof.* Since $f$ is computed by an ABP, it can be computed as an entry of a product of $d$ many $s \times s$ matrices with entries being linear functions. Let $\mathrm{MM}_{s,d}^{(i,j)}(A_1, \ldots, A_d)$ be the homogeneous degree $d$ polynomial that is the $(i,j)$-th entry of the product $A_1 \ldots A_d$ of $s \times s$ matrices. Then, $f$ is the appropriate projection $\mathrm{MM}_{s,d}^{(1,1)}$ where $A_i$ is replaced by the bipartite adjacency matrix between layer $i$ and $i+1$.

Of course, $\mathrm{MM}_{s,d}^{(i,j)}$ can be computed by a $\Sigma\Pi^{[a]}\Sigma\Pi^{[(d/a)]}$ circuit in the straightforward way: break the $d$ matrices into $a$ blocks of $(d/a)$ matrices each, multiply the matrices in each block, and finally multiply the product matrices together. Formally,

$$\mathrm{MM}_{s,d}^{(i,j)}(A_1, \ldots, A_d) \quad = \quad \mathrm{MM}_{s,a}^{(i,j)}(Q_1, \ldots, Q_a)$$

where $Q_i$ is the product matrix of the $i$-th block of $(d/a)$ matrices. Representing $\mathrm{MM}_{s,a}^{(i,j)}$ as a $\Sigma\Pi^{[a]}$ circuit, and each entry of $Q_k$ as a $\Sigma\Pi^{[d/a]}$ circuit gives the required $\Sigma\Pi^{[a]}\Sigma\Pi^{[(d/a)]}$ circuit. It is trivial to check that such a circuit has size is at most $s^a + (s^2 d)\binom{n+(d/a)}{(d/a)}$, and that it is homogeneous. □

Choosing $a = \sqrt{\frac{d \log n}{\log s}}$ (and using the bound of Lemma 3.2) gives:

**Corollary 4.2.** *Let $f$ be an $n$-variate degree $d$ polynomial computed by a homogeneous ABP of size $s$. Then, there is an equivalent homogeneous $\Sigma\Pi^{[a]}\Sigma\Pi^{[d/a]}$ $\left(\text{for } a = \sqrt{\frac{d \log n}{\log s}}\right)$ circuit $C_1$ computing $f$ of size $s_1 = \exp\left(O(\sqrt{d \log n \log s})\right)$.*

## 4.2 Step 2: $\Sigma\Pi^{[a]}\Sigma\Pi^{[(d/a)]}$ circuit to $\Sigma\wedge^{[a]}\Sigma\wedge^{[d/a]}\Sigma$ circuit

The next step relies on a construction of Fischer [Fis94] to write a monomial as a sum of powers of linear functions which we describe below.

**Lemma 4.3** ([Fis94]). *For any $n$, the monomial $x_1 \cdots x_n$ can be expressed as a linear combination of $2^{n-1}$ powers of linear forms through the following:*

$$2^{n-1} \cdot n! \cdot x_1 \dots x_n \quad = \quad \sum_{(r_2,\dots,r_n) \in \{\pm 1\}^{n-1}} \left( x_1 + \sum_{i=2}^{n} r_i x_i \right)^n \cdot (-1)^{\mathrm{wt}(\mathbf{r})}$$

*where* $\mathrm{wt}(\mathbf{r}) = |\{i \; : \; r_i = -1\}|$.

Since any degree $d$ monomial is a projection of $x_1 \dots x_d$, a corollary of the above lemma is that every degree $d$ monomial can be expressed as a sum of $2^{d-1}$ powers of linear forms. It is also worth noting that the above lemma is not true in low characteristic. For example, in a field $\mathbb{F}$ of characteristic 2, $(x + \alpha y + \beta)^2 = x^2 + \alpha^2 y^2 + \beta^2$ and hence the monomial $x \cdot y$ cannot be expressed as a sum of squares of affine forms. With this lemma, we can now proceed to the transformation in step 2.

**Lemma 4.4.** *Let $f$ be an $n$-variate degree $d$ polynomial computed by a homogeneous $\Sigma\Pi^{[a]}\Sigma\Pi^{[b]}$ circuit of size $s_1$. Then, there is an equivalent homogeneous $\Sigma\wedge^{[a]}\Sigma\wedge^{[b]}\Sigma$ circuit of size $(s_1^3 an) \cdot 2^{a+b}$ computing $f$.*

*Proof.* A $\Sigma\Pi^{[a]}\Sigma\Pi^{[b]}$ circuit $C$ of size $s_1$ computes a polynomial of the form

$$C = \sum_{i \in [s_1]} \prod_{j \in [a]} Q_{ij} \quad \text{with } Q_{ij} \text{ homogeneous of degree } b.$$

Since each of the $Q_{ij}$'s are degree $b$ polynomials having at most $s_1$ monomials, we may apply Lemma 4.3 to each monomial and express $Q_{ij}$ as a $\Sigma\wedge^{[b]}\Sigma$ circuit of size $(bs_1 n) \cdot 2^b$. Now applying Lemma 4.3 to each term $T_i = \prod_{j \in [a]} Q_{ij}$, we can express it as a sum of at most $2^a$ $a$-th powers of linear combinations of the $Q_{ij}$'s. Thus each $T_i$ can be expressed as a $\Sigma\wedge^{[a]}\Sigma\wedge^{[b]}\Sigma$ circuit of size $(a2^a) \cdot (s_1) \cdot (bs_1 n 2^b)$. Since there are at most $s_1$ terms $T_i$, we get overall a $\Sigma\wedge^{[a]}\Sigma\wedge^{[b]}\Sigma$ circuit of size at most $(s_1^3 abn) \cdot 2^{a+b}$. □

Combining this with corollary 4.2 we immediately get:

**Corollary 4.5.** *Let $f$ be an $n$-variate polynomial of degree $d$ computed by a homogeneous ABP of size $s$. Then there is an equivalent homogeneous $\Sigma\wedge^{[a]}\Sigma\wedge^{[d/a]}\Sigma$ (for $a = \sqrt{\frac{d \log n}{\log s}}$) circuit $C_2$ computing $f$ of size $s_2 = \exp\left(O(\sqrt{d \log n \log s})\right)$.*

### 4.3 Step 3: $\Sigma\wedge^{[a]}\Sigma\wedge^{[d/a]}\Sigma$ circuits to $\Sigma\Pi\Sigma$ circuits

The final step of the proof uses the *duality trick* of Saxena [Sax08]. The statement given below is slightly different from the statement in [Sax08], but can be easily inferred from the proof.

**Lemma 4.6** ([Sax08])**.** *For every $m, d > 0$ and distinct $\alpha_1, \ldots, \alpha_{md+1} \in \mathbb{Q}$, there exists $\beta_1, \ldots, \beta_{md+1} \in \mathbb{Q}$ such that*

$$(u_1 + \cdots + u_m)^d \quad = \quad \sum_{i=1}^{md+1} \beta_i \prod_{j=1}^{m} E_d(\alpha_i \cdot u_j)$$

*where $E_d(u) \overset{\text{def}}{=} 1 + \frac{u}{1!} + \ldots \frac{u^d}{d!}$.*

In order to make our exposition self-contained we reproduce the following proof from [Sax08].

*Proof.* Let $\ell \overset{\text{def}}{=} (u_1 + u_2 + \ldots + u_m)$. Then since

$$e^{\ell z} = 1 + \frac{\ell}{1!} z + \ldots + \frac{\ell^d}{d!} z^d + \ldots$$

we have that $\frac{\ell^d}{d!}$ equals the coefficient of $z^d$ in $e^{\ell z}$. Therefore

$$\begin{aligned} \ell^d &= d! \cdot \{\text{coeff of } z^d \text{ in } e^{\ell z} \ \} \\ &= d! \cdot \{\text{coeff of } z^d \text{ in } e^{u_1 z} \cdot e^{u_2 z} \cdot \ldots \cdot e^{u_m z} \ \} \\ &= d! \cdot \{\text{coeff of } z^d \text{ in } E_d(u_1 z) \cdot E_d(u_2 z) \cdot \ldots \cdot E_d(u_m z) \ \} \end{aligned}$$

Now let $F(\mathbf{u}, z) := E_d(u_1 z) \cdot E_d(u_2 z) \cdot \ldots \cdot E_d(u_m z)$. Viewing $F$ as a univariate polynomial in $z$ of degree $(md)$, we see via interpolation that the coefficient of $z^d$ in $F$ can be obtained as a linear combination of $F(\mathbf{u}, \alpha_1), \ldots, F(\mathbf{u}, \alpha_{md+1})$. That is, there exist $\delta_1, \ldots, \delta_{md+1}$ such that

$$\text{coeff of } z^d \text{ in } F(\mathbf{u}, z) = \sum_{i \in [md+1]} \delta_i F(\mathbf{u}, \alpha_i).$$

Taking $\beta_i = \delta_i \cdot d!$ then completes the proof. □

With the above lemma, we can proceed to step 3. First we obtain a $\Sigma\Pi\Sigma$ circuit over $\mathbb{C}$, the field of complex numbers. We shall subsequently convert the $\Sigma\Pi\Sigma$ circuit over $\mathbb{C}$ to one over $\mathbb{Q}$.

**Lemma 4.7.** *Let $f$ be a polynomial computed by a homogeneous $\Sigma\wedge^{[a]}\Sigma\wedge^{[b]}\Sigma$ circuit of size $s_2$ over $\mathbb{Q}$. Then, there is an equivalent $\Sigma\Pi\Sigma$ circuit $C_3$ over $\mathbb{C}$ of size $s_3 = O(s_2^3 a^2 b n)$ computing $f$. The circuit $C_3$ has top fan-in $O(s_2^2 a)$ and formal degree at most $O(s_2 a b)$.*

*Proof.* A homogeneous $\Sigma\wedge^{[a]}\Sigma\wedge^{[b]}\Sigma$ circuit $C$ computes a polynomial of the form $C = T_1 + \cdots + T_{s_2}$ where each $T_i = \left(\ell_{i1}^b + \cdots + \ell_{is_2}^b\right)^a$ for some linear forms $\ell_{ij}$'s. Applying Lemma 4.6 to each such term $T = (\ell_1^b + \cdots + \ell_{s_2}^b)^a$, we can write $T$ as

$$\begin{aligned} T &= \sum_{i=1}^{s_2 a+1} \beta_i \prod_{j=1}^{s_2} E_a(\alpha_i \cdot \ell_j^b) \\ &= \sum_{i=1}^{s_2 a+1} \prod_{j=1}^{s_2} f_i\left(\ell_j\right) \qquad \text{where } f_i(t) \overset{\text{def}}{=} E_a(\alpha_i \cdot t^b) \end{aligned}$$

Since each $f_i(t)$ is a univariate polynomial of degree at most $ba$, it splits as a product of linear factors over $\mathbb{C}$, yielding a depth-3 circuit of the form,

$$T \quad = \quad \sum_{i=1}^{sd+1} \prod_{j=1}^{s} \prod_{k=1}^{ab} (\ell_j - \gamma_{ik}) \tag{2}$$

Thus, $f$ can be computed by a $\Sigma\Pi\Sigma$ circuit of top fan-in $s_2 \cdot (s_2 a + 1) = O(s_2^2 a)$ and degree $O(s_2 ab)$, thereby yielding an overall size of

$$\begin{aligned} s_3 &= s_2 \cdot (s_2 a + 1) \cdot O(s_2 ab) \cdot (n+1) \\ &= O(s_2^3 a^2 bn) \end{aligned}$$

□

### 4.3.1 Obtaining a $\Sigma\Pi\Sigma$ circuit over rationals

The circuit thus obtained from Lemma 4.7 involve coefficients that are roots of the univariates $f_i(t) = E_a(\alpha_i \cdot t^b)$. The following lemma shows that all these coefficients come from an algebraic extension of $\mathbb{Q}$ of relatively small degree[12].

**Lemma 4.8.** *Let $\gamma_1, \ldots, \gamma_a$ be roots of $E_a(t)$, and let $\omega$ be a principal $b$-th root of unity. Then, the field $\mathbb{Q}(\gamma_1^{1/b}, \ldots, \gamma_a^{1/b}, \omega)$ contain the roots of every $E_a(\alpha \cdot t^b)$ for every $\alpha \in \mathbb{Q}$ such that $\alpha^{1/b} \in \mathbb{Q}$.*

*Proof.* The roots of $E_a(\alpha \cdot t^b)$ are precisely $\left(\frac{\gamma_i}{\alpha}\right)^{1/b} \omega^j$ for every $i \in [a], j \in [b]$. Since $\alpha^{1/a} \in \mathbb{Q}$, each of these roots are in $\mathbb{Q}(\gamma_1^{1/b}, \ldots, \gamma_a^{1/b}, \omega)$. □

Since we are free to choose the $\alpha_i$'s in Lemma 4.7, we can choose distinct $\alpha_i$'s so that $\alpha_i^{1/b} \in \mathbb{Q}$. Hence, the coefficients in the depth-3 circuit obtained from Lemma 4.7 come from $\mathbb{Q}(\gamma_1^{1/b}, \ldots, \gamma_a^{1/b}, \omega)$, which is an extension over $\mathbb{Q}$ of degree at most $(ab)^a \cdot b$ (since each $\gamma_i^{1/b}$ is a root of a polynomial over $\mathbb{Q}$ of degree at most $ab$). The following lemma below gives a generic way of converting a circuit involving coefficients from a small extension to a circuit over the base field.

[12] basic facts about field extensions, their degrees, etc. can be found in Chapter 5 of [Her75].

**Lemma 4.9.** *Let $f(\mathbf{x}) \in \mathbb{Q}[\mathbf{x}]$ be computed by a $\Sigma\Pi\Sigma$ circuit of top fanin $s_3$ and formal degree $D$ with coefficients coming from a finite extension field $\mathbb{K}/\mathbb{Q}$. Then, there is an equivalent $\Sigma\Pi\Sigma$ circuit computing $f$ of size* $\operatorname{poly}(s_3, D, [\mathbb{K} : \mathbb{Q}])$ *with coefficients coming from $\mathbb{Q}$.*

With this lemma, using the fact that the degree of the extension containing all coefficients is $2^{O(a \log d)}$ and Lemma 4.7, we immediately get our main theorem.

*Proof of Lemma 4.9.* Since $\mathbb{K}$ is a finite extension of $\mathbb{Q}$, there exists an element $\theta \in \mathbb{K}$ such that $\mathbb{K} = \mathbb{Q}(\theta)$[13]. Hence, if $m = [\mathbb{K} : \mathbb{Q}]$, then $\mathbb{K}$ is the vector space over $\mathbb{Q}$ with $\{1, \theta, \theta^2, \ldots, \theta^{m-1}\}$ as a basis and the minimum polynomial of $\theta$ has degree $m$. Therefore, any polynomial $g(\mathbf{x}) \in \mathbb{K}[\mathbf{x}]$ can be uniquely written as $g_0 + g_1\theta + \cdots + g_{m-1}\theta^{m-1}$ where each $g_i \in \mathbb{Q}[\mathbf{x}]$; we shall use $g^{[i]}$ to denote $g_i$.

If $f = T_1 + \cdots + T_{s_3}$ where each $T_i$ is a product of linear polynomials over $\mathbb{K}$, then $f = T_1^{[0]} + \cdots + T_{s_3}^{[0]}$. Hence it suffices to show that $T^{[0]}$ can be expressed as a small depth-3 circuit over $\mathbb{Q}$.

Let $T = \ell_1 \ldots \ell_D \in \mathbb{K}[\mathbf{x}]$. Then,

$$\begin{aligned} T &= \prod_{i=1}^{D} (\ell_i^{[0]} + \ell_i^{[1]}\theta + \cdots + \ell_i^{[m-1]}\theta^{m-1}) \\ &= T^{[0]} + T^{[1]}\theta + \cdots + T^{[m-1]}\theta^{m-1} \end{aligned}$$

Consider the polynomial obtained from above by replacing $\theta$ by a formal variable $y$:

$$\begin{aligned} \tilde{T}(\mathbf{x}, y) &= \prod_{i=1}^{D} (\ell_i^{[0]} + \ell_i^{[1]}y + \cdots + \ell_i^{[m-1]}y^{m-1}) \\ &= \tilde{T}_0 + \tilde{T}_1 y + \cdots + \tilde{T}_{(m-1)D}\, y^{(m-1)D} \end{aligned}$$

Therefore, using interpolation, every $\tilde{T}_i$ can be written as a linear combinations of the set $\left\{\tilde{T}(\mathbf{x}, \alpha_i) \;:\; 1 \le i \le (m-1)D + 1\right\}$. Since each such evaluation is a product of $D$ linear polynomials over $\mathbb{Q}$, we have that each $\tilde{T}_i$ can be expressed as a depth-3 circuit of top fanin $(m-1)D + 1$ and formal degree $D$.

To obtain $T^{[0]}, \ldots, T^{[m-1]}$ from $\tilde{T}_0, \ldots, \tilde{T}_{(m-1)D}$, we can express each $\theta^i$ for $m \le i \le (m-1)D$ as a linear combination of $1, \ldots, \theta^{m-1}$ to obtain that each $T^{[i]}$ is the appropriate linear combination of $\tilde{T}_0, \ldots, \tilde{T}_{(m-1)D}$. Therefore, in particular, $T^{[0]}$ can be expressed as a depth-3 circuit over $\mathbb{Q}$ of top fanin $((m-1)D) \cdot ((m-1)D + 1)$ and formal degree $D$. Hence, $f$ can be expressed as a depth-3 circuit over $\mathbb{Q}$ of top fanin $O(s_3 m^2 D^2)$ and formal degree $D$. □

[13] follows from the Primitive Element Theorem (Theorem 5.5.1 in [Her75]).

# 5 Further Consequences

## 5.1 Depth reduction to PIT lift

Any depth reduction to a class $\mathcal{C}$ of circuits provides a framework for lifting a polynomial time black-box PIT for the class $\mathcal{C}$ to black-box PITs for general circuits with *slightly worse* running time. We now present such a lift in the context of the depth reduction in Theorem 1.1 on exactly the same lines as in [AV08].

The following result roughly states that any black-box PIT for a class yields a lower bound for the same class.

**Lemma 5.1** ([HS80, Agr05]). *Let $\{\mathcal{C}_n\}$ be any subclass of arithmetic circuits computing $n$-variate degree $n$ polynomials, and suppose there is a black-box PIT running in time $n^{O(1)}$ for the circuits of size $n^2$ in $\mathcal{C}_n$. Then, there is a family of multilinear polynomials $\{q_n\}$ such that $q_n$ requires $\mathcal{C}_n$-circuits of size $2^{\Omega(n)}$. Further, $q_n$ is computable in time $2^{O(n)}$.*

The next lemma of Kabanets and Impagliazzo [KI04] states that given any family of polynomials that require exponential sized *general* circuits to compute them, one can construct a quasi-polynomial black-box PIT for *general* circuits.

**Lemma 5.2** ([KI04]). *Suppose $\{q_n\}$ is a family of multilinear polynomials computable in exponential time such that $q_n$ requires arithmetic circuits of size $2^{n^{\Omega(1)}}$. Then, there is a black-box PIT running in time $2^{(\log n)^{O(1)}}$.*

Suppose we did have a polynomial time black-box PIT for a class $\mathcal{C}_n$, then Lemma 5.1 gives a family $\{q_n\}$ that requires $2^{\Omega(n)}$-sized $\mathcal{C}_n$-circuits. Lemma 5.2, however, requires a family of polynomials $\{q_n\}$ that $2^{\Omega(n)}$-sized *general* circuits. If we could say that $\{q_n\}$ requiring $2^{\Omega(n)}$-sized $\mathcal{C}_n$-circuits *implies* $\{q_n\}$ requires $2^{n^{\Omega(1)}}$-sized general circuits, then we would be done. Such a statement is precisely the contrapositive of reducing a general circuit to a $\mathcal{C}_n$-circuit.

**Theorem 5.3.** *If there is a polynomial time black-box PIT for the class of depth-3 circuits, then there is a $2^{(\log n)^{O(1)}}$ time black-box PIT for general circuits computing a low degree polynomial.*

*Proof.* By Lemma 5.1, a polynomial time black-box PIT for depth-3 circuits imply that there is a multilinear family of polynomials $\{q_n\}$ that require $2^{\Omega(n)}$-sized depth-3 circuits. Theorem 1.1 that the family $\{q_n\}$ require general circuits of size $2^{n^{\Omega(1)}}$. Using Lemma 5.2, we obtain a $2^{(\log n)^{O(1)}}$ time black-box PIT for general circuits. □

Saha, Saptharishi and Saxena [SSS09] showed that PIT for width-2 branching programs completely capture PIT for depth-3 circuits. A corollary for the above theorem is that black-box PIT for width-2 ABPs would imply quasi-polynomial time black-box PIT for general circuits.

**Corollary 5.4.** *If there is a polynomial time black-box PIT for the class of width-2 ABPs over $\mathbb{Q}$, then there is a $2^{(\log n)^{O(1)}}$ time black-box PIT for general circuits computing a low degree polynomial.*

## 5.2 Reduction from $\Sigma\Pi\Sigma$ circuits to $\Sigma\wedge\Sigma\wedge\Sigma$ circuits

In this section we show that, for the purpose of proving super-polynomial formula lower bounds over $\mathbb{Q}$, we can equivalently work with any of $\Sigma\Pi\Sigma$ or $\Sigma\wedge\Sigma\wedge\Sigma$ circuits. We will need the following families of symmetric polynomials

$$\mathsf{Sym}_n(x_1,\ldots,x_m) \stackrel{\text{def}}{=} \sum_{\substack{S\subseteq[m]\\|S|=n}} \prod_{i\in S} x_i \qquad , \qquad \mathsf{Pow}_n(x_1,\ldots,x_m) \stackrel{\text{def}}{=} \sum_{j\in[m]} x_j^n\ .$$

Our proof relies on the following implication of Newton's identities (cf. [Lit50]).

**Lemma 5.5.** *Let $\mathsf{Sym}_n(x_1,\ldots,x_m)$ and $\mathsf{Pow}_n(x_1,\ldots,x_m)$ denote the elementary symmetric and power symmetric polynomials of degree $n$ respectively, as defined above. Then,*

$$\mathsf{Sym}_n = \frac{1}{n!}\cdot\begin{vmatrix} \mathsf{Pow}_1 & 1 & 0 & \cdots & \\ \mathsf{Pow}_2 & \mathsf{Pow}_1 & 2 & 0 & \cdots \\ \vdots & & \ddots & \ddots & \\ \mathsf{Pow}_{n-1} & \mathsf{Pow}_{n-2} & \cdots & \mathsf{Pow}_1 & n-1 \\ \mathsf{Pow}_n & \mathsf{Pow}_{n-1} & \cdots & \mathsf{Pow}_2 & \mathsf{Pow}_1 \end{vmatrix}.$$

We now give a partial converse of Lemma 4.7.

**Lemma 5.6.** *Let $f$ be an $N$-variate degree $n$ polynomial over any characteristic zero field computed by a $\Sigma\Pi\Sigma$ circuit with top fan-in $s$ and product gates with fan-in at most $d$. Then, there exists an equivalent $\Sigma\wedge\Sigma\wedge\Sigma$ circuit of size $\operatorname{poly}(N,d,s)\cdot 2^{O(\sqrt{n}\cdot\log n)}$.*

*Proof.* We are given that there exist $s\cdot d$ linear polynomials $\ell_{ij}$'s such that

$$f = \sum_{i=1}^{s}\prod_{j=1}^{d}\ell'_{ij}.$$

As we are working over an infinite field, we can assume without loss of generality[14] that each $\ell'_{ij}$ has a non-zero constant term. This is because constructing an depth-5 powering circuit for an *affine shift* of $f$ gives a depth-5 powering circuit for $f$ of polynomially larger size. Hence $f$ has an expression of the form

$$f = \sum_{i=1}^{s} \alpha_i \prod_{j=1}^{d} (1 + \ell_{ij}).$$

Let $f^{[r]}$ denote the degree-$r$ homogeneous component of $f$. Then

$$f^{[r]} = \sum_{i=1}^{s} \alpha_i \cdot \mathsf{Sym}_r(\ell_{i1}, \ldots, \ell_{id}).$$

We now focus on a summand of the form $\mathsf{Sym}_r(\ell_1, \ldots, \ell_d)$. From Lemma 5.5, there exist scalars $\beta_{\mathbf{a}}$'s such that,

$$\mathsf{Sym}_r(\ell_1, \ldots, \ell_d) = \sum_{\substack{\mathbf{a}=(a_1,\ldots,a_r)\in\mathbb{Z}^r_{\geq 0} \\ \sum_i i\cdot a_i = r}} \beta_{\mathbf{a}} \cdot \prod_{i\in[r]} \mathsf{Pow}_i^{a_i}(\ell_1, \ldots, \ell_d). \tag{3}$$

The number of solutions of $\sum_{i\in[r]} i \cdot a_i = r$ is exactly the number of ways to partition the natural number $r$ and hence is $2^{O(\sqrt{r})}$ by the Hardy-Ramanujan estimate for the partition function [HR18]. Hence the number of terms in the above summation is $2^{O(\sqrt{r})}$.

The next step is to convert the product into a powering gate, similar to step 2 in the depth reduction which used Fischer's construction [Fis94]. However, we shall need a more efficient way of converting a "low support" monomial into a sum of powers. Such a construction is provided the following lemma of Ellison [Ell69].

**Lemma 5.7** ([Ell69])**.** *Over any field of zero characteristic, any homogeneous $n$-variate degree $d$ polynomial can be expressed as a linear combination of $d$-th powers of linear forms. Further, the number of such powers of linear forms used is bounded by $(d+1)^{n-1} = 2^{O(n \log d)}$.*

If $y_i \overset{\text{def}}{=} P_i(\ell_1, \ldots, \ell_d)$, then equation (3) can be written as

$$\mathsf{Sym}_r(\ell_1, \ldots, \ell_d) = \sum_{\substack{\mathbf{a}=(a_1,\ldots,a_r)\in\mathbb{Z}^r_{\geq 0} \\ \sum_i i\cdot a_i = r}} \beta_{\mathbf{a}} \cdot \prod_{i\in[r]} y_i^{a_i}.$$

[14] by taking a random affine shift

Note that if $\sum_{i\in[r]} i \cdot a_i = r$ then at most $O(\sqrt{r})$ of the $a_i$'s are non-zero. Hence, each of the monomials (in the $y_i$'s) in the RHS of the above equation has support at most $O(\sqrt{r})$ and degree at most $r$. Hence, applying Lemma 5.7 to each of the monomials gives a representation of the form

$$\mathsf{Sym}_r(\ell_1,\ldots,\ell_d) \;=\; \sum_{\substack{\mathbf{a}=(a_1,\ldots,a_r)\in\mathbb{Z}_{\geq 0}^r \\ \sum_i i\cdot a_i = r}} \beta_{\mathbf{a}} \cdot \left( \sum_{i=1}^{2^{O(\sqrt{r}\log r)}} (\alpha_{i1}y_1 + \cdots + \alpha_{ir}y_r)^{\sum a_j} \right)$$

for some scalars $\alpha_{ij}$, which is a $\Sigma\wedge\Sigma$ circuit in the $y_i$'s of size $2^{O(\sqrt{r}\log r)}$. Substituting each $y_i$ by $P_i(\ell_1,\ldots,\ell_d)$, we get a $\Sigma\wedge\Sigma\wedge\Sigma$ circuit for $\mathsf{Sym}_r(\ell_1,\ldots,\ell_d)$ of size bounded by $\mathrm{poly}(N,d)\cdot 2^{O(\sqrt{r}\log r)}$. Since each $f^{[r]}$ is a linear combination of $s$ such terms, and $f = f^{[0]} + \cdots + f^{[n]}$, we have that $f$ can be computed by a $\Sigma\wedge\Sigma\wedge\Sigma$ circuit of size $\mathrm{poly}(N,d,s)\cdot 2^{O(\sqrt{n}\log n)}$. □

# 6 Conclusion

We saw that powering circuits are useful as an intermediate step in doing depth reduction of general circuits. We feel that proving lower bounds for powering circuits will also be useful as a good intermediate step towards proving lower bounds for general circuits. Towards this we recall a problem posed in [CKW11] in the form of a more specific conjecture.

**Conjecture 6.1.** *Over any field $\mathbb{F}$ of characteristic 0, any depth-$\Delta$ powering circuit with* $\mathrm{poly}(d)$*-bounded degree computing the monomial $x_1\cdots x_d$ must be of size at least $2^{d^{\Omega(1/\Delta)}}$ size.*

Also, step 1 and step 2 of the depth-reduction reduces any ABP to a depth-5 powering circuit with formal degree bounded by the degree of the polynomial. It is only step 3 that results in the blow-up in degree. Hence, proving a $2^{\omega(\sqrt{d}\log d)}$ lower bound for low (formal) degree depth-5 powering circuits computing $\mathsf{Perm}_d$ would yield super-polynomial formula lower bound. Proving such a lower bound for homogeneous depth-5 powering circuits might be an easier task than proving lower bounds for non-homogeneous depth-3 circuits (where we have no degree bound). We pose the following problem as the next potential step towards proving super-polynomial formula lower bounds for $\mathsf{Perm}_d$.

**Problem 6.2. Lower Bounds for $\Sigma\wedge\Sigma\wedge\Sigma$ circuits.** *Show that for any expression of the form*

$$\mathsf{Perm}_d = \sum_{i\in[s]} \left( \sum_{j\in[t]} \ell_{ij}^{\sqrt{d}} \right)^{\sqrt{d}}$$

*then one must have* $s \cdot t = 2^{\omega(\sqrt{d})}$.

We already know of a $2^{\Omega(\sqrt{d})}$ lower bound [GKKS13] on $s$, independent of how large $t$ is. Presumably using the internal structure of the polynomials, say the fact that each of the inner terms have a *low partial derivative space*, we should be able to prove better lower bounds. We believe that any technique to solve Problem 6.2 would be very insightful towards the grander goal of proving super-polynomial circuit lower bounds.

**Summary.** $\Sigma\Pi\Sigma$ circuits have been intensely investigated. Some recent work led to both structural results on $\Sigma\Pi\Sigma$ identities as well as deterministic blackbox algorithms for various subclasses of $\Sigma\Pi\Sigma$ circuits. We also note here that the Sylvester-Gallai configurations that arose in this line of work, has led to some beautiful and productive series of works [BDYW11, DSW12] in understanding such geometric configurations culminating in essentially optimal upper and lower bounds on the dimension of such configurations. Because of these developments, we are inclined towards interpreting Theorem 1.1 in an optimistic manner - of proving lower bounds for $\Sigma\Pi\Sigma$ circuits as yet another potential route to obtaining lower bounds for arbitrary arithmetic circuits.

### Acknowledgements

We are thankful to Avi Wigderson for several useful comments and suggestions on the initial draft of this write-up.

## References

[Agr05] Manindra Agrawal. Proving Lower Bounds Via Pseudo-random Generators. In *FSTTCS*, pages 92–105, 2005.

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

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

[BDYW11] B. Barak, Z. Dvir, A. Yehudayoff, and A. Wigderson. Rank bounds for design matrices with applications to combinatorial geometry and locally correctable codes. In *STOC*, pages 519–528, 2011.

[CKW11] Xi Chen, Neeraj Kayal, and Avi Wigderson. Partial derivatives in arithmetic complexity. *Foundations and Trends in Theoretical Computer Science*, 2011.

[Csa76] L. Csanky. Fast parallel inversion algorithm. *SIAM Journal on Computing*, 5:618–623, 1976.

[DSW12] Zeev Dvir, Shubhangi Saraf, and Avi Wigderson. Improved rank bounds for design matrices and a new proof of Kelly's theorem. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:138, 2012.

[Ell69] W.J. Ellison. A 'waring's problem' for homogeneous forms. *Proceedings of the Cambridge Philosophical Society*, 65:663–672, 1969.

[Fis94] I. Fischer. Sums of like powers of multivariate linear forms. *Mathematics Magazine*, 67(1):59–61, 1994.

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

[GKKS13] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the chasm at depth four. In *Proceedings of the Conference on Computational Complexity (CCC)*, 2013.

[GR00] Dima Grigoriev and Alexander A. Razborov. Exponential lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Appl. Algebra Eng. Commun. Comput.*, 10(6):465–487, 2000.

[Her75] I.N. Herstein. *Topics in Algebra.* Wiley, second edition, 1975.

[HR18] G. H. Hardy and S. Ramanujan. Asymptotic formula in combinatory analysis. *Proceedings of the London Mathematical Society*, s2-17(1):75–115, 1918.

[HS80] Joos Heintz and Claus-Peter Schnorr. Testing Polynomials which Are Easy to Compute (Extended Abstract). In *STOC*, pages 262–272, 1980.

[KI04] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004.

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

[Lit50] D.E. Littlewood. *The Theory of Group Characters and Matrix Representations of Groups.* Ams Chelsea Publishing. AMS Chelsea Pub., 2nd edition, 1950.

[MP08] Guillaume Malod and Natacha Portier. Characterizing valiant's algebraic complexity classes. *J. Complexity*, 24(1):16–38, 2008.

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

[Raz10] Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. In *STOC*, pages 659–666, 2010.

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

[Rys63] H. J. Ryser. Combinatorial mathematics. *Math. Assoc. of America*, 14, 1963.

[Sax08] Nitin Saxena. Diagonal Circuit Identity Testing and Lower Bounds. In *ICALP (1)*, pages 60–71, 2008.

[Shp01] Amir Shpilka. *Lower Bounds for Small Depth Arithmetic and Boolean Circuits*. PhD thesis, The Hebrew University, 2001.

[SSS09] Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. The Power of Depth 2 Circuits over Algebras. In *FSTTCS*, pages 371–382, 2009.

[SW01] A. Shpilka and A. Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Computational Complexity*, 10(1):1–27, 2001.

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

[Val79] Leslie G. Valiant. Completeness Classes in Algebra. In *STOC*, pages 249–261, 1979.

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

[Wig07] Avi Wigderon. P, NP and Mathematics - A computational complexity perspective. In *Proceedings of the ICM 06 (Madrid)*, volume 1, pages 665–712. EMS Publishing House, Zurich, 2007.

# A Lower bounds for *slightly* non-homogeneous $\Sigma\Pi\Sigma$ circuits

Nisan and Wigderson [NW97] showed that the partial derivative method can be used to prove lower bounds for homogeneous depth-3 circuits computing $\mathsf{Perm}_n$. Their method also works if the $\Sigma\Pi\Sigma$ circuit was *slightly non-homogeneous*, in the sense that the formal degree is $O(n^{2-\delta})$ for any constant $\delta > 0$. This observation has not been explicitly stated in the literature (to the best of our knowledge) and we note that here.

**Theorem A.1** ([NW97]). *If $C$ is a $\Sigma\Pi\Sigma$ circuit of formal degree $O(n^{2-\delta})$ computing $\mathsf{Perm}_n$, then size$(C) = \exp\left(\Omega(n^\delta)\right)$.*

Before we prove the theorem, we shall need the following lemmas for estimating binomial coefficients.

**Lemma A.2.** $\binom{a}{b} = 2^{\Theta\left(a \cdot H\left(\frac{b}{a}\right)\right)} \cdot \mathrm{poly}(ab)$ *where* $H(x) = x \ln \frac{1}{x} + (1-x) \ln \left(1 - \frac{1}{x}\right)$.

*Proof.* Follows from Stirling's approximation of $n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$ (cf. [Rom]). □

**Lemma A.3.** *For every $0 < x < 1$, we have that* $x \ln \frac{1}{x} \leq H(x) \leq x \ln \frac{1}{x} + x$ □

*Proof of Theorem A.1.* The proof shall proceed by bounding the dimension of the $k$-th order partial derivatives (for a suitably chosen parameter $k$). Let $\boldsymbol{\partial}^{=k}(f)$ denote the vector space of $k$-th order partial derivatives of $f$.

**Lower bounding** $\dim \boldsymbol{\partial}^{=k}(\mathsf{Perm}_n)$**:** Any $k$-th order derivative of $\mathsf{Perm}_n$ is just a $(n-k) \times (n-k)$ permanental minor. There are $\binom{n}{k}^2$ many such permanental minors, and all of them are of course linearly independent. Hence $\dim \boldsymbol{\partial}^{=k}(\mathsf{Perm}_n) \geq \binom{n}{k}^2$.

**Upper bounding** $\dim \boldsymbol{\partial}^{=k}(C)$**:** The circuit $C$ computes a polynomial of the form

$$C \quad = \quad T_1 + \cdots + T_s$$

where each $T_i$ is a product of at most $d = O(n^{2-\delta})$ linear polynomials. Also as $\dim \boldsymbol{\partial}^{=k}(C) \leq s \cdot \max \boldsymbol{\partial}^{=k}(T_i)$, it suffices to bound the dimension of partial derivatives of one such term.

It is easy to observe that the $k$-th order partial derivative of $T = \ell_1 \ldots \ell_d$ is generated by linear combinations of the subterms $\left\{T_S = \prod_{i \in S} \ell_i \ : \ S \subseteq [d], |S| = d-k\right\}$. Hence, $\boldsymbol{\partial}^{=k}(T) \leq \binom{d}{k}$ and hence $\dim \boldsymbol{\partial}^{=k}(C) \leq s \cdot \binom{d}{k}$.

If $C$ computes $\mathsf{Perm}_n$, then

$$\begin{aligned}
s \cdot \binom{d}{k} &\geq \binom{n}{k}^2 \\
\implies s \geq \frac{\binom{n}{k}^2}{\binom{d}{k}} &\geq \exp\left(2n \cdot H\left(\frac{k}{n}\right) - d \cdot H\left(\frac{k}{d}\right) - O(\ln n)\right) \quad \text{(Lemma A.2)} \\
&\geq O\left(2n\left(\frac{k}{n}\ln\frac{n}{k}\right) - d\left(\frac{k}{d}\ln\frac{d}{k} + \frac{k}{d}\right)\right) - O(\ln n) \quad \text{(Lemma A.3)} \\
&= O\left(k \ln\left(\frac{n^2}{kde}\right)\right) - O(\ln n) \\
&= \Omega(n^\delta) \quad \text{for } k = \epsilon n^\delta \text{ for a small enough } \epsilon > 0 \\
\text{Hence} \quad s &= \exp\left(\Omega(n^\delta)\right)
\end{aligned}$$

□