

Separated borders: Exponential-gap fanin-hierarchy theorem for approximative depth-3 circuits*

PRANJAL DUTTA, College of Computing & Data Science, Nanyang Technological University, Singapore

NITIN SAXENA, Department of CSE, Indian Institute of Technology Kanpur, India

Mulmuley and Sohoni (2001) proposed an ambitious program, called the *Geometric Complexity Theory* (GCT), to prove $P \neq NP$ and related conjectures using algebraic geometry and representation theory. Gradually, GCT has introduced new structures and questions in complexity theory. GCT tries to capture an algebraic/geometric notion of ‘approximation’ by defining the *border* classes. Surprisingly, (Kumar ToCT’20) proved the universal power of the border of top-fanin-2 depth-3 circuits ($\overline{\Sigma^{[2]}\Pi\Sigma}$); which is in complete contrast to its classical model. Recently, (Dutta,Dwivedi,Saxena, FOCS’21) put an upper bound, by showing that bounded top-fanin border depth-3 circuits ($\overline{\Sigma^{[k]}\Pi\Sigma}$ for constant k) can be computed by polynomial-size algebraic branching programs (ABPs). It was left open to show an *exponential* separation between the class of ABPs and $\overline{\Sigma^{[k]}\Pi\Sigma}$.

In this article, we show a strong exponential separation between any two consecutive border classes, $\overline{\Sigma^{[k]}\Pi\Sigma}$ and $\overline{\Sigma^{[k+1]}\Pi\Sigma}$, thus establishing an *optimal hierarchy* of constant top-fanin border depth-3 circuits. In the language of GCT, we prove an exponential hierarchy for *padded-k-th-secant-varieties* of the Chow variety of \mathbb{F}^{n+1} . This positively answers [Open question 2 of Dutta,Dwivedi,Saxena FOCS’21] and [Problem 8.10 with constant r , of Landsberg, Annal.Ferrara’15].

CCS Concepts: • **Theory of computation** → **Problems, reductions and completeness; Algebraic complexity theory; Complexity classes.**

Additional Key Words and Phrases: approximative, border, depth-3, hierarchy, formula, GCT, secant variety, padded, ABP, ROABP, ARO, VF, inhomogeneous.

ACM Reference Format:

Pranjal Dutta and Nitin Saxena. 2026. Separated borders: Exponential-gap fanin-hierarchy theorem for approximative depth-3 circuits. 1, 1 (January 2026), 39 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

CONTENTS

Abstract	1
Contents	1
1 Introduction	2
1.1 GCT and border complexity	4
1.2 The Chow variety and lower bounds in GCT	5

*A preliminary version was presented in FOCS 2022.

Authors’ addresses: Pranjal Dutta, duttpranjal@gmail.com, College of Computing & Data Science, Nanyang Technological University, Singapore; Nitin Saxena, nitin@cse.iitk.ac.in, Department of CSE, Indian Institute of Technology Kanpur, India.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2026 Association for Computing Machinery.

Manuscript submitted to ACM

Manuscript submitted to ACM

53	1.3	Our results: The fanin-hierarchy theorem	7
54	1.4	Known depth-3 lower bounds and their limitations	9
55	2	Preliminaries I	10
56	3	Preliminaries II.	11
57	4	Detailed proof of Theorem 2 for $k \leq 3$	14
58	4.1	$k = 1$ case	14
59	4.2	$k = 2$ case: The basic induction idea	15
60	4.3	$k = 3$ case: The bloated induction idea	19
61	4.4	Brief pathway to general k	22
62	4.5	(im)Possibility of a ‘simpler’ proof, and comparison with Dutta, Dwivedi and Saxena [DDS21]	23
63	5	Hardness lies in all-non-homogeneity	24
64	5.1	Reduced problem	25
65	6	Proving Theorem 17: The main lower bound	26
66	6.1	Main fanin reduction lemma and formulating the induction hypothesis	26
67	6.2	Proof of the main induction lemma	28
68	6.3	The effect of fan-in reduction on the circuit size	31
69	6.4	Proving lower bound for ratio of depth-3 diagonal circuits	32
70	6.5	Tying the pieces together: Proof of Theorem 2	33
71	7	Conclusion	33
72		References	34
73	A	Basics of algebraic complexity	37
74	B	Technical lemmas	37
75	C	Limitations of bounded fan-in depth-3 circuits	39

1 INTRODUCTION

The main aim of computational complexity theory is to understand, as precisely as possible, the amount of computational resources required to perform computational tasks. These resources could be of various kinds depending on the computational model under consideration— e.g., time/ space for Turing machines; size, depth, and fanin for boolean and algebraic circuits; and so on. A fundamental question in this context is “Does more of the same resource lead to more computational power?”. Classical theorems in Computational Complexity such as the Time Hierarchy Theorem [HS65] and Space Hierarchy Theorem [SHL65] answer this question (affirmatively) for the resources of time respectively space on multitape Turing machines. In this paper, we consider an analogous question for algebraic circuits.

A polynomial $f \in \mathbb{F}[x_1, \dots, x_n]$ over a field \mathbb{F} is said to be computable by an algebraic circuit of size s , if there exists a directed acyclic graph of size at most s , whose leaf nodes are labeled by variables or field constants, internal nodes are labeled by field operations ($+$ and \times), and the polynomial computed at the root node is f . For a polynomial f , $\text{size}(f)$ denotes the size of the *smallest circuit* computing it. Another important complexity parameter is the *depth* – the length of the longest path in the circuit from a leaf to the root. In this paper, we

are interested in *depth-3 circuits* $\Sigma^{[k]}\Pi\Sigma$; they compute polynomials of the form $\sum_{i \in [k]} \Pi_j \ell_{ij}$, where ℓ_{ij} are affine linear polynomials.

We consider the question of proving a *top-fanin-hierarchy* theorem for algebraic circuits (in the border setting). Informally, we ask the following.

Question 1 (The fanin-hierarchy question). *For fixed $k \geq 1$, are there explicit families of polynomials $P_n \in \mathbb{F}[x_1, \dots, x_n]$, such that P_n can be ‘approximated’ by a small $\Sigma^{[k+1]}\Pi\Sigma$ -circuit but cannot be approximated by a small $\Sigma^{[k]}\Pi\Sigma$ -circuit?*

For the definitions of *explicit* and *approximation*, see Section 2. Question 1 is also related to fundamental variety constructions in geometric complexity theory (GCT); we will discuss more on GCT and its importance in the depth-3 setting in Section 1.2. If one considers the above question in the classical setting, i.e. without approximation, there is an easy *impossibility* result known, which shows that the inner product polynomial, $IP_{k+1} := \sum_{i \in [k+1]} x_i \cdot y_i$, for $k \geq 1$, cannot be computed by $\Sigma^{[k]}\Pi\Sigma$ -circuit, regardless of the size ([Kum20, CGJ⁺18]). We sketched a proof for completeness, see Theorem 33.

However, the unexpected *universality* of border depth-3 fanin-2 circuits by Kumar [Kum20] shows that the classical impossibility result breaks down in the border for $k \geq 2$; for a detailed statement, see Section 1.2. Therefore, an exponential separation between the class of border depth-3 fanin-2 circuits and any other reasonable algebraic class becomes a nontrivial and intriguing question to study. In this work, we affirmatively answer Question 1 as stated below; for a formal statement, see Theorem 2 and its remarks.

THEOREM 1 (INFORMAL). *There are explicit polynomials that can be computed by a small $\Sigma^{[k+1]}\Pi\Sigma$ -circuit but cannot be approximated by a small $\Sigma^{[k]}\Pi\Sigma$ -circuit.*

In the above, our candidate hard polynomial will be a degree d version of the inner product polynomial IP_{k+1} ; see subsection 1.3. But, Question 1 is far from obvious, and we will employ quite powerful models like ABP, ARO, $\text{Gen}(k, s)$ in the proofs; for respective definitions, see Appendix A and Section 6. This work builds upon the DiDIL technique introduced by Dutta, Dwivedi, and Saxena [DDS21], however, there are several technical differences, which we will elaborate in great detail in Section 4.5.

Basic models and classes. We quickly go through a few important models of computations that appear in our proofs. For example, if in a circuit, the out-degree of each internal node is 1, then it is a *formula*. Any formula can be converted into a layered graph called *algebraic branching program* (ABP) with a polynomial blowup in size. With different models come different complexity classes, which accordingly classify polynomials. For example, VP contains polynomial families $(f_n)_n$, where f_n is a $\text{poly}(n)$ -variate $\text{poly}(n)$ -degree polynomial, computable by circuits of size $\text{poly}(n)$. Similarly, one can define VBP and VF, wrt the models ABPs and formulas, respectively. Finally, the class VNP contains polynomial families $(f_n)_n$, where each f_n can be expressed as $\sum_{\mathbf{y} \in \{0,1\}^{p(n)}} g_n(\mathbf{x}, \mathbf{y})$, with the polynomial family g_n in VP, and $p(\cdot)$ being some polynomial function. Valiant [Val79] conjectured that VBP \neq VNP (respectively VP \neq VNP), as an algebraic analog of the P vs. NP problem. For details, see [SY10, Mah13].

1.1 GCT and border complexity

Mulmuley and Sohoni introduced the Geometric Complexity Theory (GCT) program to prove algebraic variants of the $P \neq NP$ conjecture using algebraic geometry and representation theory [MS01, MS08]. Let $\det_n := \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^n x_{i,\sigma(i)}$ be the determinant polynomial, and let $\text{per}_m := \sum_{\sigma \in S_m} \prod_{i=1}^m x_{i,\sigma(i)}$ be the permanent polynomial. Valiant's *determinant vs. permanent* conjecture, or the $VNP \not\subseteq VBP$ question states that the smallest size of a matrix A whose entries are affine linear polynomials such that $\det(A) = \text{per}_m$, is *not polynomially* bounded in m . Mulmuley and Sohoni strengthened the conjecture by allowing the permanent to be approximated arbitrarily closely coefficientwise instead of being computed exactly.

More formally, the Mulmuley–Sohoni conjecture states that when $n = \text{poly}(m)$, it must be that $\ell^{n-m} \text{per}_m \notin \overline{\text{GL}_{n^2} \det_n}$. Here $\text{GL}_{n^2} := \text{GL}(\mathbb{C}^{n \times n})$ acts on the space of homogeneous degree n polynomials in n^2 many variables by invertible linear transformations on the variables¹, and ℓ is some homogeneous linear polynomial; without loss of generality one can assume $\ell := x_{1,1}$. Moreover, the orbit closure can be taken as the Zariski closure; see below for the definition. The polynomial $\ell^{m-n} \text{per}_n$ is called the ‘padded permanent’, and the phenomenon of multiplying with a power of a linear form is called *padding*. Note here that the action of GL_{n^2} replaces variables by *homogeneous* linear polynomials.

GCT has intimate connections to many areas in mathematics and computer science including designing matrix multiplication algorithms [Str74, Bin80, BCRL79, CW90, LO15], computational invariant theory [FS13, Mul12b, GGOW16, BGO⁺18, IQS18], algebraic natural proofs [GKSS17, BIL⁺21, CKR⁺20, KRST20], lower bounds [BI13, Gro15, LO15], optimization [AZGL⁺18, BFG⁺19], derandomization [Mul12a, Muk16, DDS21], and many more. We refer to [BLMW11, Mul12b, Mul12a] for expository references.

Approximative closure. Borrowing the definition from [Bür04, Bür20], we say that a polynomial $f(x) \in \mathbb{F}[x_1, \dots, x_n]$ is *approximated* by a polynomial $g(x, \epsilon) \in \mathbb{F}(\epsilon)[x]$, over the field $\mathbb{F}(\epsilon)$, if there exists a polynomial $S(x, \epsilon) \in \mathbb{F}[\epsilon][x]$ such that $g = f + \epsilon S$. We can think analytically that $\lim_{\epsilon \rightarrow 0} g = f$. The class $\overline{\mathcal{C}}$, the *approximative closure* of a complexity class \mathcal{C} , can be defined in the usual way. Note that, arbitrary ϵ -power is allowed in the circuit approximating g , since computing constants is cost-free. Often, we will use the underlying computation over $\mathbb{F}[\epsilon^{\pm 1}]$ ², instead of $\mathbb{F}(\epsilon)$, since they are known to be *equivalent* while talking about border complexity; see [BCS13, Lem. 15.22] and [Bür00] for details. Throughout the article, we will often use f being approximated by a circuit from \mathcal{C} , and f being computed by a $\overline{\mathcal{C}}$ circuit interchangeably.

Euclidean closure. A polynomial family $(f_n)_n$ is in the Euclidean closure of \mathcal{C} over \mathbb{F} if, for every n , there exists an infinite sequence of polynomials $g_{n,i}$ in \mathcal{C} over \mathbb{F} such that the limit point of the sequence of coefficient vectors corresponding to $g_{n,i}$ is the coefficient vector of f_n .

Zariski closure. For a circuit class \mathcal{C} , consider the system of all polynomial equations which are satisfied by the coefficient vector corresponding to each polynomial in \mathcal{C} . Then, the Zariski closure of \mathcal{C} consists of all the polynomials such that the corresponding coefficient vectors satisfy assignments of the system of polynomial equations.

¹For a homogeneous polynomial p and $g \in \text{GL}_{n^2}$ define the homogeneous polynomial gp via $(gp)(x) := p(g^t x)$. The orbit is defined as $\text{GL}_{n^2} p := \{gp \mid g \in \text{GL}_{n^2}\}$.

² $\mathbb{F}[\epsilon^{\pm 1}] = \mathbb{F}[\epsilon, \epsilon^{-1}]$, i.e. the elements are of the form $\sum_{i=-M}^N a_i \epsilon^i$, for some $M, N \geq 0$

209 Interestingly, these notions are known to be equivalent over the algebraically closed field \mathbb{C} [Mum95, S2.C]
 210 and [Kra84, AI.7.2 Folgerung]. Since these definitions are equivalent, without loss of generality, we will work
 211 with the approximation closure throughout our paper.
 212

213 **Border complexity measure.** Border complexity measures are defined via limits, so that any function that
 214 can be approximated arbitrarily closely by low complexity functions itself has low border complexity. The
 215 *approximative* (or *border*) complexity of f , denoted $\overline{\text{size}}(f)$, is the minimum size of the circuit computing g , over
 216 $\mathbb{F}(\epsilon)$. It is easy to see that $\overline{\text{size}}(f) \leq \text{size}(f)$. Due to the arbitrary $(1/\epsilon)$ -power terms in the circuit computing
 217 g , evaluation at $\epsilon = 0$ is not necessarily valid, although the limit still exists. Hence, any lower bound on the
 218 border complexity of f in terms of the *exact* complexity of f is unclear.
 219

220 If $g = f + \epsilon \cdot S$, where $S \in \mathbb{F}[\epsilon][x]$, and we have a circuit of size s computing g , where the constants come
 221 from $\mathbb{F}(\epsilon)$, then Bürgisser [Bür04, Theorem 5.7] showed that there exists a nonnegative integer $M \leq \exp(s)$,
 222 such that $\hat{g} = \epsilon^r \cdot f + \epsilon^{r+1} \cdot \hat{f}$, for some $r \geq 0$, where $\hat{g}, \hat{f} \in \mathbb{F}[\epsilon, x]$, and $\deg_\epsilon(\hat{g}) \leq M$; furthermore \hat{g} has an
 223 s -size circuit over $\mathbb{F}(\epsilon)$. Therefore, one could use the trick of interpolating, by setting $(M + 1)$ -many random
 224 ϵ -values from \mathbb{F} to get f ; we need random values of ϵ so that we do not make the circuit of \hat{g} undefined.
 225 Therefore, the following relation is the best-one known: $\overline{\text{size}}(f) \leq \text{size}(f) \leq \exp(\overline{\text{size}}(f))$.
 226
 227
 228

229 **De-bordering.** De-bordering is the task of proving an upper bound on some non-border complexity measure
 230 in terms of a border complexity measure, thus getting rid of limits. De-bordering is at the heart of under-
 231 standing the difference between Valiant's determinant vs permanent conjecture, and Mulmuley and Sohoni's
 232 strengthened version of the same conjecture. The de-bordering of matrix multiplication tensors by Bini [Bin80]
 233 played a pivotal role in the development of efficient matrix multiplication algorithms. In particular, the re-
 234 sults of [CHI⁺18] shows that the matrix multiplication exponent $\omega = \lim_{n \rightarrow \infty} \log_n \overline{\text{WR}}(\text{trace}(X_n^3))$, where
 235 $X_n = (x_{ij})_{i,j \in [n]}$ is a matrix of variables, and $\overline{\text{WR}}(\cdot)$ denotes the border Waring rank. Border Waring rank of
 236 f , denoted $\overline{\text{WR}}(f) \leq s$, means that $f = \lim_\epsilon \sum_{i=1}^s \ell_i^d(\epsilon)$ for some linear forms ℓ_i ; thus, it captures a restricted
 237 version of border depth-3 circuits. This makes the problem of de-bordering depth-3 circuits already a com-
 238 pelling subject of study. We refer to the recent survey on debordering by Dutta and Lysikov [DL25] for further
 239 importance and detailed exposition.
 240
 241
 242
 243

244 1.2 The Chow variety and lower bounds in GCT

245 **Border depth-3 circuits: An algebraic view.** Since, depth-2 circuits are *closed* under taking limit, i.e. $\overline{\Pi\Sigma} = \Pi\Sigma$
 246 and $\overline{\Sigma\Pi} = \Sigma\Pi$, it is natural to study one step further and understand the border of depth-3 circuits. It is not
 247 hard to show that $\overline{\Pi\Sigma\Pi} = \Pi\Sigma\Pi$, which leaves us to understand $\overline{\Sigma\Pi\Sigma}$. As mentioned before, Kumar [Kum20]
 248 showed that border depth-3 fanin-2 circuits are 'universal'³; i.e. *any* d -degree, n -variate polynomial can be
 249 approximated by $\Sigma^{[2]}\Pi^{[D]}\Sigma$ circuits over $\mathbb{C}(\epsilon)$. For a generic polynomial, $D \approx \binom{n+d}{d}$. In the case when D is
 250 small, [DDS21] proved that $\overline{\Sigma^{[k]}\Pi\Sigma} \subseteq \text{VBP}$; that is for any $f \in \overline{\Sigma^{[k]}\Pi\Sigma}$, of approximative circuit size s , can
 251 be exactly computed by an ABP of size $s^{\exp(k)}$. This raised a basic open question of whether the containment
 252 is strict, i.e. $\overline{\Sigma^{[k]}\Pi\Sigma} \subsetneq \text{VBP}$ (even, $\overline{\Sigma^{[k]}\Pi\Sigma} \subsetneq \text{VNP}$ was *unclear!*). We settle these questions and in fact, show
 253 exponential separations between these classes in this work.
 254
 255
 256
 257

258 ³[Kum20] states the result for homogeneous polynomials. But trivially, the proof also works for any non-homogeneous polynomial, by
 259 homogenizing it using a new variable x_0 , and finally de-homogenizing it by setting $x_0 = 1$.
 260

Border depth-3 circuits: A geometric view. Most of the times, theoretical computer scientists are interested in proving robust lower bounds, i.e. lower bound techniques which would also work under the limit. This could be roughly translated into asymptotic geometry terms as follows:

Given a sequence of some ‘nice’ vector spaces V_n , and sequences of points and groups, does the inclusion⁴ fail for every $n \geq n_0$, for some n_0 ?

The *Chow variety* is one of the simplest varieties studied in the field of algebraic geometry; it is believed to be a good testing ground for GCT [Lan15, Section 2]. Interestingly, this goes back to 19th-century mathematics, studied independently by Hermite (1854) and Hadamard (1897).

Informally, specializing to the group of diagonal matrices and taking the orbit closure, one obtains the famous *Chow variety*, $\text{Ch}_d(W) \subset \mathbb{P}S^d W$; usually $W = \mathbb{C}^n$ (or, for $\det_n / \text{perm}_n, \mathbb{C}^{n^2}$). We denote by $S^d W$, the space of polynomials of degree d on W^* , $\mathbb{P}V$ denotes the projective space, and we denote $[v]$ as a corresponding point. Then, formally,

$$\text{Ch}_d(W) := \{[z] \in \mathbb{P}S^d W \mid z = w_1 \dots w_d, \text{ for } w_j \in W\}.$$

Therefore, one can define the *Chow rank* of a homogeneous polynomial f of degree d , denoted $\text{rank}_{\text{Ch}}(f)$, to be the minimum k such that $f = \sum_{i=1}^k \prod_{j=1}^d \ell_{ij}$, where ℓ_{ij} are linear forms. Often in the literature, $\text{rank}_{\text{Ch}}(f) = k$ is equivalently expressed as: the smallest k such that f (as a point) is in $\sigma_k^0(\text{Ch}_d(W))$ (= set of points with $\text{Ch}_d(W)$ -rank at most k).

Moving to the approximative setting, one defines *Chow border rank*, $\overline{\text{rank}}_{\text{Ch}}(f)$, as the border analogue of the Chow rank. In other words, $\overline{\text{rank}}_{\text{Ch}}(f) = k \iff f \in \sigma_k(\text{Ch}_d(W)) = \overline{\sigma_k(\text{Ch}_d(W))}$ the Zariski closure in $\mathbb{P}S^d W$ of $\sigma_k^0(\text{Ch}_d(W))$; it is called the k -th *secant variety*⁵ of the Chow variety of W . For details, refer to [Lan15, Lan17]. These two ranks happen to exactly coincide with the depth-3 respectively border depth-3 homogeneous circuits of f , with the smallest fanin k .

In general, from an algebraic complexity perspective, we are interested in the *non-homogeneous* setting. For instance, Kumar’s expression [Kum20, Section 3.1] is non-homogeneous. However, with suitable padding (& $W = \mathbb{C}^{n+1}$), Kumar’s result translates into geometric terms: *For any degree- d homogeneous polynomial f , there exists a linear form ℓ such that $\ell^{m-d} f \in \sigma_2(\text{Ch}_m(W))$, for $m = \exp(n, d)$, or equivalently, $\overline{\text{rank}}_{\text{Ch}}(\ell^{m-d} f) = 2$.* On the other hand, if one restricts $m = \text{poly}(n, d)$, Dutta, Dwivedi and Saxena [DDS21] showed that $\sigma_2(\text{Ch}_m(W)) \subseteq \text{VBP}$; the same result holds if one replaces σ_2 by σ_k , as long as k is a constant.

Quest for lower bounds. Most polynomials require exponential-size circuits to compute or approximate. However, proving an explicit lower bound remains a major challenge in algebraic complexity. The celebrated result of Baur and Strassen [Str73, BS83] shows the existence of an explicit n -variate degree- d polynomial which requires circuits of size at least $\Omega(n \cdot \log d)$. Though $d = \text{poly}(n)$ gives *superlinear* lower bound, ideally one would hope for at least a *superpolynomial* lower bound, optimistically even an exponential one!

Similarly, if one restricts the model of computation, we have better lower bounds; e.g. Kalorkoti [Kal85] showed $\Omega(n^2 / \log n)$ formula lower bound for an explicit n -variate polynomial, which could be easily extended to the approximative setting. We refer the reader to [Sap19] for a comprehensive survey of the lower bounds in algebraic complexity.

⁴by inclusion, we mean the points under the group action in V_n

⁵For a given variety X , s -th secant variety of X is Zariski closure of the union of all secant $(s - 1)$ -planes to X .

The situation in GCT is no way better and near to what was expected at its inception. The linchpin of the GCT program was that the permanent and determinant are both *uniquely* characterized (up to a constant factor) by their symmetries; and moreover, lower bounds are equivalent to orbit closure containment [Gro12, Section 3.3.2]. Therefore, the simplest way of proving a lower bound would be to find *occurrence obstructions*, i.e. finding an irreducible representation with multiplicity for the permanent larger than that of the determinant. The *no-go theorem* of [IP17, BIP19] proved that this is impossible since [MS01] uses the padded formulation of the Mulmuley-Sohoni conjecture. We note that no analogous result is known when the determinant is replaced, for example, by the iterated matrix multiplication polynomial – correspondingly replacing determinantal complexity with the algebraic branching program width, since the padded-free or the homogeneous setting is a much nicer setup to deal with. The potential of multiplicity obstructions is explored in [DIP20, IK20, DGI⁺22, DGI⁺24, Bür24].

On the other hand, a recent breakthrough result by Limaye, Srinivasan & Tavenas [LST21] showed the first superpolynomial lower bound against general algebraic formulas of *constant*-depth, over all fields of characteristic 0 or large. Since their method is linear-rank-based, the proofs can be lifted in the border classes analogously [Gro15, AF22]. This gives us a stronger urge to continue the quest to show *exponential* separation, in the constant-depth regime. This current work can be seen as a further step in that pursuit.

1.3 Our results: The fanin-hierarchy theorem

We state our result formally now. Our result holds for any field of characteristic 0/large characteristic.

THEOREM 2 (FANIN-HIERARCHY EXP-GAP). *Fix any constant $k \geq 1$. There is an explicit n -variate, degree $< n$ polynomial f such that f can be computed by a $\Sigma^{[k+1]}\Pi\Sigma$ circuit of size $O(n)$, but f requires $2^{\Omega(n)}$ -size $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuits.*

Remarks. 1. Interestingly, **Theorem 2** is *optimal* for $k \geq 2$. To see this, note that $2^{\Theta(n)}$ is also an upper bound on the size of $\overline{\Sigma^{[2]}\Pi\Sigma}$ circuits, for n -variate polynomials computed $\Sigma^{[k+1]}\Pi\Sigma$ circuits of size $O(n)$; see [Kum20]. Thus, we have completely characterized the gap between two constant-top-fanin border depth-3 circuits.

2. Consider the following degree- d (> 2) polynomial on $3d$ -variables $f(x) := x_1 \cdots x_d + x_{d+1} \cdots x_{2d} + x_{2d+1} \cdots x_{3d}$. Note that $f(x)$ has a trivial $\Sigma^{[3]}\Pi\Sigma$ circuit of size $O(d)$. Our proof method shows that f requires $2^{\Theta(d)}$ -size $\overline{\Sigma^{[2]}\Pi\Sigma}$ circuits.

A word on the polynomial family. Our candidate polynomial family, for a fixed k , is the *sum-product* polynomial⁶ $\mathcal{P}_{k+1} := (P_{k+1,d})_d$, where

$$P_{k+1,d} := \sum_{i \in [k+1]} \prod_{j=1}^d x_{(i-1)d+j}.$$

This is a strict generalization of the *inner product* polynomial IP_{k+1} , defined on page 2. Using [Kum20, Theorem 1.3], it follows that $IP_{k+1} \in \overline{\Sigma^{[2]}\Pi^{[O(k)]}\Sigma}$. Therefore, we cannot expect to show an exponential lower bound against IP_{k+1} . This can be mitigated by working with general d -degree monomials (instead of quadratic ones). Clearly, $P_{k+1,d}$ is a multilinear degree- d polynomial on $(k+1) \cdot d$ -variables. Lower bounds for \mathcal{P}_2 have been studied by Shpilka [Shp02], in a different context. This polynomial is closely related to the Trace-Iterated-Matrix-Multiplication, $\text{Tr-IMM}_{k+1,d}$, which is the trace of the product of d -many $(k+1) \times (k+1)$ symbolic matrices X_r , $\text{tr}(\prod_{r \in [d]} X_r)$, where the (i, j) -th entry of the matrix X_r is $x_{ij}^{(r)}$. In particular, take diagonal matrices

⁶This term has been borrowed from [Lan17, Section 8.12.2].

365 X_r with (i, i) -th entry being the variable $x_{(i-1)d+r}$. Note that $P_{k+1,d} = \text{tr}(X_1 \cdots X_d)$. Since this is a restriction
 366 on X_r , our proof also holds for $\text{Tr-IMM}_{k+1,d}$ as well.
 367

368 Non-triviality and implications.

- 370 (1) **Homogeneous vs non-homogeneous:** If we restrict ourselves to only the *homogeneous* setting, then it is
 371 easy to argue that $P_{k+1,d}$, *cannot* be computed by a homogeneous $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuit. Informally speaking,
 372 the proof exploits the fact that if indeed $P_{k+1,d}$ were computed by a $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuit C , then picking k
 373 many linear forms, one from each product gate of C , and simultaneously substituting them to 0, makes
 374 C a zero polynomial, while $P_{k+1,d}$ remains nonzero even after such a substitution; for a formal proof
 375 see [Corollary 16](#).
 376 However, this impossibility result no longer holds when we allow *non-homogeneity* [[Kum20](#)]. The
 377 naive proof method as described above fails miserably, as setting many affine linear functions to zero
 378 may lead to *inconsistency*. For example, both $x_1 - \epsilon x_2$ and $x_1 - \epsilon x_3 - 1$ can never be zero in the border
 379 (as $\epsilon \rightarrow 0$); while at the same time, x_2, x_3 contribute in the computation due to the use of the powers
 380 of $1/\epsilon$. This is precisely what grants the seemingly innocuous border computing model $\overline{\Sigma^{[2]}\Pi\Sigma}$ its
 381 universal expressive power!
 382 (2) **PIT to lower bound:** Efficient PIT algorithms are known for $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuits [[LST21](#), [DDS21](#)]. However,
 383 they are *not known* to imply strong lower bounds in the same model. The derandomization to lower
 384 bound techniques are insensitive to fanin- k [[HS80](#), [KI03](#), [AGS19](#), [KS19](#)]; as they require the ability to
 385 interpolate, which blows-up the top-fanin of the model.
 386 (3) **Separating $\overline{\Sigma^{[k]}\Pi\Sigma}$ from VF, VBP, VNP:** It was asked in [[DDS21](#)] whether $\overline{\Sigma^{[k]}\Pi\Sigma} \neq \text{VBP}$ or not. Our
 387 proof of [Theorem 2](#) shows an exponential separation between $\overline{\Sigma^{[k]}\Pi\Sigma}$ and $\Sigma^{[k+1]}\Pi\Sigma$. Consequently,
 388 we have exponential separation between $\overline{\Sigma^{[k]}\Pi\Sigma}$ and VF (respectively VBP, and VNP).
 389 (4) **Non-rank based approach:** Ours is probably the first strong depth-3 lower bound in the border setting,
 390 where the basic argument is *non-rank*-based, and does not care about the upper bound on the ϵ -power
 391 that hides in the expression.
 392 (5) **Solving Landsberg's open question [[Lan15](#), [Problem 8.10](#)]:** The same proof can be adapted for \det_d and
 393 perm_d analogously. Interestingly, [Theorem 2](#) answers a restricted version of a question asked in [[Lan15](#),
 394 [Problem 8.10](#)]; namely, $\ell^{m-d} \cdot \det_d \notin \sigma_r(\text{Ch}_m(W))$, for any $m \leq 2^{o(d)}$ and constant r . Geometrically,
 395 [Theorem 2](#) exponentially separates the padded k -th-secant varieties of the Chow variety of $W = \mathbb{C}^{n+1}$,
 396 for all constants k .
 397
 398
 399
 400
 401
 402
 403
 404

405 **Comparing our parameters with [[LST21](#)] breakthrough.** Our proof method works with both the polynomi-
 406 als $\text{IMM}_{k+1,d}$ (where instead of trace, one is interested in $(1,1)$ -th entry), and the trace version $\text{Tr-IMM}_{k+1,d}$,
 407 already defined before. The linear-rank-based lower bound method by Limaye, Srinivasan & Tavenas [[LST21](#)]
 408 shows only a superpolynomial separation between depth-3 (unbounded fanin) circuits and IMM , with different
 409 parameters. Since this method can be extended to the *border* setting, the same work yields a *superpolynomial*
 410 separation between $\overline{\Sigma^{[k]}\Pi\Sigma}$ and VBP. Their dominating variable is n (with $d = o(\log n)$) and they showed an
 411 $n^{\sqrt{d}}$ lower bound (which further weakens above depth-4); which is not optimal, as the optimal lower bound
 412 would be n^d . Since we take $k = O(1)$ and our dominating variable is d , we show an optimal lower bound of
 413 $\approx \binom{(k+1)d+d}{d} \approx 2^{\Omega(d)}$, thus establishing an *exponential* separation between $\overline{\Sigma^{[k]}\Pi\Sigma}$ and VBP. More significantly,
 414
 415

[LST21] shows a *superpolynomial-gap* depth-hierarchy; while our result is about an exponential-gap fanin-hierarchy inside the depth-3 regime. In summary, the parameters and the results in [LST21] are incomparable with ours, and their method is very different since our method is *non-linear* at a very fundamental level.

1.4 Known depth-3 lower bounds and their limitations

In this section, we briefly discuss the well-known lower bounds for depth-3 circuits (mostly in the classical setting), their techniques, and why they fail to yield our result in the border setting.

In a very influential piece of work, Nisan and Wigderson [NW96] showed that over any field \mathbb{F} , every *homogeneous* $\Sigma\Pi\Sigma$ circuit computing the determinant \det_d must be of size $2^{\Omega(d)}$. They employed a method based on partial derivatives, which can be readily adapted to the border setting. We remark that the lower bound is actually on the top-fanin and thus for constant top-fanin k , \det_d cannot even be computed by a homogeneous $\overline{\Sigma^{[k]}}\Pi\Sigma$ circuit! For completeness, we sketched an alternative non-rank-based proof, see Lemma 15.

Unfortunately, the partial derivative measure performs poorly in the non-homogeneous setting, primarily because the degree bound no longer holds, causing the argument to break down entirely. Furthermore, when considering $P_{k+1,d}$ instead of \det_d or $\text{IMM}_{k+1,d}$, the exponential separation between k and $k+1$ cannot be demonstrated using rank-based methods, as straightforward estimates of both upper and lower bounds fail to produce meaningful results. For analogous reasons, techniques developed by Grigoriev and Karpinski [GK98], and by Grigoriev and Razborov [GR00], over a *fixed* finite field, also do not apply in this context.

There has been substantial progress on lower bounds for restricted depth-3 circuits, where the restrictions involve bounded independence, bounded read/occur, bounded bottom fan-in, and related properties. A central theme in these proofs involves a sequence of reductions: one first reduces general (non-homogeneous) depth-3 circuits of top fan-in s to a subclass of depth-5 circuits with top fan-in increased to $s \cdot \exp(\sqrt{d})$ [SW01, GKKS16]; then, via a random restriction, obtains a (homogeneous) depth-4 decomposition; and finally applies a variant of the shifted partial derivative measure to prove lower bounds [KLSS17, KS17, KS16]. While these techniques can be adapted to the border setting, they fall short of establishing any meaningful hierarchy theorem or exponential separation for non-homogeneous $\overline{\Sigma^{[k]}}\Pi\Sigma$ circuits.

Even for showing lower bound for $\overline{\Sigma^{[2]}}\Pi\Sigma$ circuits, we probably cannot hope to use factorization or Chinese remaindering (CRT) based ideas, since $\text{mod } \langle e^M \rangle$, we get non-unique, usually exponentially many, factors. For example, $x^2 \equiv (x - a \cdot e^{M/2}) \cdot (x + a \cdot e^{M/2}) \text{ mod } \langle e^M \rangle$; for all $a \in \mathbb{F}$. In this case, there are, in fact, infinitely many factorizations. Moreover, $\lim_{\epsilon \rightarrow 0} 1/\epsilon^M \cdot \left(x^2 - (x - a \cdot \epsilon^{M/2}) \cdot (x + a \cdot \epsilon^{M/2}) \right) = a^2$. Therefore, infinitely many factorizations may give infinitely many limits, and thus the analysis becomes much more intricate.

In the classical affine setting, it is known that computing the immanant (which includes determinant and permanent) requires exponential-size $\Sigma^{[k]}\Pi\Sigma$ circuits [ASSS16, Theorem 1.7]. These lower bounds were obtained using the Jacobian. However, it is not clear how it behaves with respect to $\lim_{\epsilon \rightarrow 0}$, or whether we can come up with a meaningful working definition. For example, let $f_1 = x_1 + e^M \cdot x_2$, and $f_2 = x_1$, where M is arbitrarily large. Then, the underlying Jacobian $J(f_1, f_2) = e^M$ is nonzero, but it tends to zero in the limit. Seemingly, this causes the entire Jacobian-based framework to break down in the border setting, as it no longer facilitates variable reduction for border models – for instance, one must retain faithfulness to both x_1 and x_2 in the example above.

Though not exactly comparable, but [BIZ18] showed a counter-intuitive ‘collapse’ in the border setting: $\overline{\text{VBP}}_2 = \overline{\text{VBP}}_3 = \dots = \overline{\text{VBP}}_k$, for any constant k , whereas it is also known that $\text{VBP}_2 \subsetneq \text{VF} = \text{VBP}_3 = \dots =$

469 VBP_k [BOC92, AW16]. While these works succeed in demonstrating collapse results, the underlying techniques
 470 differ significantly from ours and do not directly contribute to proving a strict hierarchy-type separation, as we
 471 aim to establish in the current work.
 472

473 Lastly, in subsection 1.3, we have already discussed the similarities and differences with [LST21]. This
 474 concludes our discussion of prior work and its limitations.
 475

476 2 PRELIMINARIES I

477
 478 In this section, we describe some of the assumptions and notations used throughout the paper.
 479

480 **Notation.** Denote $[n] = \{1, \dots, n\}$, and $\mathbf{x} = (x_1, \dots, x_n)$. We use $\mathbb{F}[[x]]$, to denote the ring of formal power
 481 series over \mathbb{F} . Formally, $f = \sum_{i \geq 0} c_i x^i$, with $c_i \in \mathbb{F}$, is an element in $\mathbb{F}[[x]]$. Further, $\mathbb{F}(x)$ denotes the function
 482 field, where the elements are of the form f/g , where $f, g \in \mathbb{F}[x]$ ($g \neq 0$).
 483

484 We call $\ell =: a_1 x_1 + \dots + a_n x_n$, a linear polynomial without the constant term, as a *linear form*.

485 Throughout the paper, by $\text{rank}(\ell_1, \dots, \ell_n)$, we mean $\text{rank}_{\mathbb{F}}(\ell_1, \dots, \ell_n)$, i.e. the dimension of the linear space
 486 generated by the linear forms $\{\ell_1, \dots, \ell_n\}$.
 487

488 **Explicit.** A family $\{P_n \in \mathbb{F}[x_1, \dots, x_n] \mid n \geq 1\}$, is said to be *explicit* if there is a deterministic algorithm that
 489 given as input 1^n and a monomial m over the variables x_1, \dots, x_n , computes the coefficient of the monomial m
 490 in P_n in time $\text{poly}(n)$.
 491
 492

493 **Logarithmic derivative.** Over a ring R and a variable y , the *logarithmic derivative* $\text{dlog}_y : R[y] \rightarrow R(y)$ is
 494 defined as $\text{dlog}_y(f) := \partial_y f / f$; here ∂_y denotes the partial derivative with respect to variable y . One important
 495 property of dlog is that it is *additive* over a product as $\text{dlog}_y(f \cdot g) = \partial_y(fg) / (fg) = (f \cdot \partial_y g + g \cdot \partial_y f) / (fg) =$
 496 $\text{dlog}_y(f) + \text{dlog}_y(g)$. [*dlog linearizes product*]
 497
 498

499 **Circuit size.** Some of the complexity parameters of a circuit are – 1) *size*, the total number of nodes and edges,
 500 2) *depth*, the number of layers, 3) *degree*, the maximum degree polynomial computed by any node, and 4) *fanin*,
 501 the maximum number of inputs to a node.
 502
 503

504 **Depth-2 and Depth-3 circuits.** Product depth-2 circuits, denoted as $\Pi\Sigma$, compute polynomials of the form $\prod_i \ell_i$,
 505 where ℓ_i are affine linear polynomials. Depth-3 circuits with top-fanin k are denoted as $\Sigma^{[k]}\Pi\Sigma$; they compute
 506 polynomials of the form $\sum_{i \in [k]} \prod_j \ell_{ij}$, where ℓ_{ij} are affine linear functions. Also, depth-3 diagonal circuits are
 507 denoted as $\Sigma^{[s]} \wedge \Sigma$; they compute polynomials of the form $\sum_{i=1}^s \ell_i^{e_i}$, for linear polynomials ℓ_i . When we write
 508 $\Sigma \wedge \Sigma$, it means the top-fanin is *unbounded* and one can roughly consider its size to be $O(snd)$, where d is the
 509 maximum power in the exponent of a linear polynomial.
 510
 511

512 **Ideal generated by linear forms.** For given n -variate linear forms L_1, \dots, L_r , we denote $\langle L_1, \dots, L_r \rangle$, the ideal
 513 generated by L_i , for $i \in [r]$, which contains elements of the form $\sum_{i \in [r]} a_i \cdot L_i$, for $a_i \in \mathbb{F}[x]$.
 514

515 **Operation on Complexity Classes.** For class \mathcal{C} and \mathcal{D} defined over ring R , our bloated model is any combination
 516 of sum, product, and division of polynomials from respective classes. For instance, $\mathcal{C}/\mathcal{D} = \{f/g : f \in \mathcal{C}, 0 \neq$
 517 $g \in \mathcal{D}\}$, similarly $\mathcal{C} \cdot \mathcal{D}$ for products, $\mathcal{C} + \mathcal{D}$ for sum, and other possible combinations. Also we use \mathcal{C}_R to denote
 518 the basic ring R over which \mathcal{C} is being computed.
 519
 520

Valuation. Valuation is a map $\text{val}_y : R[y] \rightarrow \mathbb{Z}_{\geq 0}$, over a ring R , such that $\text{val}_y(\cdot)$ is defined to be the maximum power of y dividing the element. It can be easily extended to localized/fraction ring $R(y)$, by defining $\text{val}_y(p/q) := \text{val}_y(p) - \text{val}_y(q)$; where the integer value can be negative.

Field. We denote the underlying field as \mathbb{F} and assume that it is of characteristic 0 (eg. the field of rationals \mathbb{Q} , the field of reals \mathbb{R} , the field of p -adics \mathbb{Q}_p etc.). All our results hold for other fields (eg. \mathbb{F}_{p^e} etc.) of large characteristic p . We also denote \mathbb{F}^* , as the multiplicative group of the field \mathbb{F} .

We will often explicitly use the following two facts. The first fact is quite standard and we just state it without any proof.

Fact 1. Let R be any arbitrary ring. Let $g, h \in R[z]$ such that $\text{val}_z(g) \geq \text{val}_z(h)$. Then, $g/h \in R[[z]]$, and $\text{val}_z(g/h) = \text{val}_z(g) - \text{val}_z(h)$.

Fact 2. Let $g, h \in R(z, \epsilon)$ such that $\text{val}_\epsilon(g) = \text{val}_\epsilon(h) = 0$. Then, $\lim_{\epsilon \rightarrow 0} \partial_z(g/h) = \partial_z(g_0/h_0)$, where $g_0 := g|_{\epsilon=0}$, and $h_0 := h|_{\epsilon=0}$.

PROOF SKETCH. Let $g = \sum_{i \geq 0} g_i \epsilon^i$, and $h = \sum_{i \geq 0} h_i \epsilon^i$. Then, it is easy to see that

$$g/h = (g_0/h_0) + \epsilon \cdot S, \text{ for some } S \in R(z)[[\epsilon]].$$

The conclusion follows easily after taking the derivative and a limit. \square

Approximative closure. For an algebraic complexity class \mathcal{C} , the ‘approximation’ is formally modeled as follows [BIZ18, Definition 2.1].

Definition 1 (Approximative closure of a class). Let $\mathcal{C}_{\mathbb{F}}$ be a class of polynomials defined over a field \mathbb{F} . Then, $f(\mathbf{x}) \in \mathbb{F}[x_1, \dots, x_n]$ is said to be in approximative closure $\overline{\mathcal{C}}$ if and only if there exists polynomial $Q \in \mathbb{F}[\epsilon, \mathbf{x}]$ such that $\mathcal{C}_{\mathbb{F}(\epsilon)} \ni g(\mathbf{x}, \epsilon) = f(\mathbf{x}) + \epsilon \cdot Q(\mathbf{x}, \epsilon)$. In short, we denote $\lim_{\epsilon \rightarrow 0} g(\mathbf{x}, \epsilon) := f(\mathbf{x})$.

Cone-size of monomials. For a monomial \mathbf{x}^a , the cone of \mathbf{x}^a is the set of all sub-monomials of \mathbf{x}^a . The cardinality of this set is called *cone-size* of \mathbf{x}^a . It equals $\prod_{i \in [n]} (a_i + 1)$, where $\mathbf{a} = (a_1, \dots, a_n)$. We will denote $\text{cs}(m)$, as the cone-size of the monomial m .

3 PRELIMINARIES II.

Power series and dlog. In the proof, we will use the ring of formal power series $R[[x_1, \dots, x_n]]$ (in short $R[[\mathbf{x}]]$), for some suitable ring R , see [Niv69, DSS18, Sin19]. One of the key benefits of this ring comes from the inverse identity: $(1 - x)^{-1} = \sum_{i \geq 0} x^i$.

The logarithmic derivative operator $\text{dlog}_z(f) = (\partial_z f)/f$ is another key tool which *linearizes* the product gate, since

$$\text{dlog}_y(f \cdot g) = \partial_y(fg)/(fg) = (f \cdot \partial_y g + g \cdot \partial_y f)/(fg) = \text{dlog}_y(f) + \text{dlog}_y(g). \quad (1)$$

This operator enables us to use power series expansion, and converts the Π -gate to \wedge .

Let $\ell \in \mathbb{F}[\mathbf{x}]$ be a linear polynomial such that the constant term is nonzero. For simplicity, suppose $\ell := 1 + \tilde{\ell}$, where $\tilde{\ell}$ is a homogeneous linear polynomial. Further, let $\Phi : \mathbf{x} \mapsto z\mathbf{x}$. Note that $\Phi(\ell) = 1 + z \cdot \tilde{\ell}$. Therefore, by

573 simple power series expansion as mentioned above, $\text{dlog}_z(\Phi(\ell))$ becomes:

$$574 \quad \text{dlog}_z(\Phi(\ell)) = \frac{\tilde{\ell}}{1+z \cdot \tilde{\ell}} = \sum_{i \geq 0} (-1)^i z^i \cdot \tilde{\ell}^{i+1}. \quad (2)$$

575
576
577 Looking at the above expression, one can also say that $\text{dlog}(\Phi(\ell))$ can be computed by a $\Sigma \wedge \Sigma$ circuit of size
578 $\text{poly}(nd)$ over the ring $\mathcal{R} := \mathbb{F}[z]/\langle z^d \rangle$.

579 One crucial fact that we will use throughout the paper is the following. Let $h \in \mathbb{F}(x, \epsilon)[z]$, and suppose
580 $\text{val}_z(h) = 0$, that is, the maximal power of z dividing h is 0 (for a formal definition, see Section 2). Then $1/h$ is
581 a power series in z , i.e. $1/h \in \mathbb{F}(x, \epsilon)[[z]]$. To give an explicit example, let $h := z + \epsilon$; trivially $\text{val}_z(h) = 0$ and
582 $1/(z + \epsilon) = \sum_{i=0}^{d-1} (-1)^i z^i / \epsilon^{i+1} \pmod{z^d}$.

583
584
585 **Laurent Series.** Let $d \geq 1$ be an integer. Let $\mathbb{F}((z))$ denote the field of formal Laurent series in z , i.e., series of
586 the form $\sum_{k \geq k_0} a_k z^k$ with $k_0 \in \mathbb{Z}$ and $a_k \in \mathbb{F}$. We view $\mathbb{F}(z)$ as a subfield of $\mathbb{F}((z))$ via the Laurent expansion at
587 $z = 0$.

588 For $f(z) = \sum_{k \geq k_0} a_k z^k \in \mathbb{F}((z))$, we define its truncation at order d by

$$589 \quad f \text{ trun } z^d := \sum_{k=k_0}^{d-1} a_k z^k.$$

590
591
592 Equivalently, $f \text{ trun } z^d$ denotes the class of f in the additive quotient $\mathbb{F}((z))/z^d \mathbb{F}[[z]]$, which identifies two
593 Laurent series if they agree in all terms of degree strictly less than d .

594 A Laurent series is allowed to have finitely many negative powers of z . For example,

$$595 \quad \frac{1}{z^2 - 2z^3} = z^{-2} + 2z^{-1} + 4 + 8z + \dots,$$

596 and, more generally,

$$597 \quad \frac{1}{x_1 z - x_2 x_3 z^2} = \frac{1}{x_1} z^{-1} + \frac{x_2 x_3}{x_1^2} + \frac{x_2^2 x_3^2}{x_1^3} z + \frac{x_2^3 x_3^3}{x_1^4} z^2 + \dots.$$

598
599
600 **Looking modulo an ideal generated by linear forms over \mathbb{F} .** Given an n -variate polynomial $f(x)$ and a linear
601 form $L(x)$, let us try to understand what $f(x) \pmod{\langle L \rangle}$ means; since a generalization of this will play a key
602 role in the technical part of our proof.

603 To begin with, let us first understand when L is just a variable. Without loss of generality, suppose $L := x_1$.
604 Then $f \pmod{\langle L \rangle}$, is nothing more than $f(0, x_2, \dots, x_n)$. For an arbitrary linear form L that depends on x_1 ,
605 we apply a suitable isomorphism ϕ in the space of linear forms, sending L to x_1 , and other variables to itself.
606 Since, $\phi(f) = f(Ax)$, for some invertible matrix A , $f \pmod{\langle L \rangle}$ essentially means looking at $f(Ax) \pmod{x_1}$.
607 This is same as $f(\ell_1, \dots, \ell_n)$, for some linear forms ℓ_i , where $\text{rank}_{\mathbb{F}}(\ell_1, \dots, \ell_n) = n - 1$.

608 There is another way of looking at the same expression. Using the same notation as above, $\Phi(f)$ can be
609 uniquely written as $\phi(f) = \sum_{i=0}^d f_i x_1^i$, where $f_i \in \mathbb{F}[x_2, \dots, x_n]$. Note that, ϕ^{-1} sends x_1 to L and x_i to x_i , for
610 $i \in [2, n]$. Therefore, $f = \sum_{i=0}^d \phi^{-1}(f_i) L^i$, and hence $f \pmod{\langle L \rangle} = \phi^{-1}(f_0)$, which depends on x_2, \dots, x_n .

611 As a concrete example, $f \pmod{\langle x_1 + x_2 \rangle}$ equals $f(-x_2, x_2, x_3, \dots, x_n)$. So, in particular, in the case where L
612 has x_1 , we can assume ℓ_2, \dots, ℓ_n to be x_2, \dots, x_n . For a general purpose, L may not contain x_1 , but some other
613 variable x_i , and in that case ℓ_j will be x_j , for $j \neq i$ and ℓ_i will be free of the variable x_i , again making the rank
614 to be exactly $n - 1$. For simplicity, we avoid this and assume that the rank is $n - 1$, which is all we will need in the
615 main proof.

616
617
618
619
620
621
622
623
624 Manuscript submitted to ACM

Similarly, the above can be extended to understanding $f(x) \bmod \langle L_1, \dots, L_t \rangle$, for $t \geq 1$, where each L_i is a linear form. $\langle L_1, \dots, L_t \rangle$ is an ideal generated by the linear forms L_i , i.e., for any $g \in \langle L_1, \dots, L_t \rangle$, we have $g =: \sum_{i \in [t]} a_i \cdot L_i$, for some $a_i \in \mathbb{F}[x]$. Generalizing $t = 1$ in the same way, $f(x) \bmod \langle L_1, \dots, L_t \rangle$ can be thought of as $f(\ell_1, \dots, \ell_n)$, where $\text{rank}_{\mathbb{F}}(\ell_1, \dots, \ell_n) \geq n - t$. The rank becomes exactly $n - t$, when L_i are linearly independent. As a concrete example, $f \bmod \langle x_1 + x_2, x_2 + x_3, x_1 + 2x_2 + x_3 \rangle$ becomes $f(x_3, -x_3, x_3, x_4, \dots, x_n)$. In this case, the rank is $n - 2$, and $t = 3$.

Understanding modulo a linear form over $\mathbb{F}(\epsilon)$. The above formalism works fine when there are no ϵ terms involved. What happens if we look at a polynomial $g(x, \epsilon)$ modulo linear polynomial with ϵ -terms involved? Here is the claim.

Claim 3. Suppose, $g(x, \epsilon) = f(x) + \epsilon \cdot S(x, \epsilon)$, where $S \in \mathbb{F}[\epsilon][x]$. Let $L \in \mathbb{F}[\epsilon][x]$ be a linear polynomial (in x) such that $L|_{\epsilon=0}$ is a nonzero linear form. Then, there exist linear forms ℓ_1, \dots, ℓ_n , with $\text{rank}(\ell_1, \dots, \ell_n) = n - 1$, and a polynomial $S' \in \mathbb{F}[[\epsilon]][x]$ such that the following equality holds over $\mathbb{F}[[\epsilon]]$:

$$g \bmod \langle L \rangle = f(\ell_1, \dots, \ell_n) + \epsilon \cdot S'(x, \epsilon).$$

PROOF. Let $L := \sum_{i=0}^r L_i \epsilon^i$ where $L_i \in \mathbb{F}[x]$ are linear polynomials. Further, by assumption L_0 is a nonzero linear form. Fix a variable ordering. Then, one can write each L_i as $L_i := c_i L_0 + L'_i$, for some $c_i \in \mathbb{F}$, and $L'_i := L_i \bmod \langle L_0 \rangle$ is a unique linear polynomial. Therefore, clubbing all L_0 terms together, one can rewrite L as $(1 + \epsilon \cdot p(\epsilon))L_0 + \epsilon \cdot R$, where $p(\epsilon) = \sum_{i=1}^r c_i \epsilon^{i-1}$, and $R := \sum_{i=1}^r L'_i \epsilon^{i-1} \in \mathbb{F}[\epsilon][x]$. Note that, L'_i as well as R are independent of L_0 . Therefore, doing the operation $\bmod \langle L \rangle$ is same as substituting $L_0 = -\epsilon \cdot R / (1 + \epsilon \cdot p(\epsilon))$. Let $q := -\epsilon \cdot R / (1 + \epsilon \cdot p(\epsilon))$. By definition, $q \in \mathbb{F}[[\epsilon]][x]$ is a linear polynomial in x , and further $q \equiv 0 \bmod \langle \epsilon \rangle$.

Finally, suppose the variable x_1 appears in L_0 . Then, from the above discussion, both f and S can be uniquely written as follows: $f = \sum_{i=0}^d f_i \cdot L_0^i$, where $f_i \in \mathbb{F}[x_2, \dots, x_n]$, and $S = \sum_{i=0}^d S_i \cdot L_0^i$, where $S_i \in \mathbb{F}[\epsilon][x_2, \dots, x_n]$. We assumed the degree upper bound in f and S to be d .

By the discussion above, $\bmod \langle L \rangle$, over the ring $\mathbb{F}[[\epsilon]]$ is nothing but substituting L_0 by q . What we get over the ring $\mathbb{F}[[\epsilon]]$ is:

$$f \bmod \langle L \rangle = \sum_{i=0}^d f_i \cdot q^i, \quad S \bmod \langle L \rangle = \sum_{i=0}^d S_i q^i.$$

Since $q \equiv 0 \bmod \langle \epsilon \rangle$, we get that over $\mathbb{F}[[\epsilon]]$, the following equation holds:

$$\begin{aligned} g \bmod \langle L \rangle &= f \bmod \langle L \rangle + \epsilon \cdot S \bmod \langle L \rangle \\ &= f_0 + \epsilon \cdot S'(x, \epsilon). \end{aligned}$$

By definition, $f_0 = f \bmod \langle L_0 \rangle$. Finally, by the previous discussion, there are linear forms ℓ_1, \dots, ℓ_n with $\text{rank}(\ell_1, \dots, \ell_n) = n - 1$ such that $f_0 = f(\ell_1, \dots, \ell_n)$. This finishes the proof. \square

Remark. Instead of working over $\mathbb{F}[[\epsilon]]$, one can also truncate and work over $\mathbb{F}[\epsilon] / \langle \epsilon^{M+1} \rangle$, for some large positive integer M . To see this formally, one can simply replace $q' \equiv q \bmod \langle \epsilon^{M+1} \rangle$, in the above proof, and work over the desired ring.

By induction, one can easily extend this to working with an ideal generated by multiple linear polynomials over $\mathbb{F}(\epsilon)$ to get the following:

677 **Claim 4.** Suppose, $g(x, \epsilon) = f(x) + \epsilon \cdot S(x, \epsilon)$, where $S \in \mathbb{F}[\epsilon][x]$. Let $L_1, \dots, L_t \in \mathbb{F}[\epsilon][x]$ are t -many linear
 678 polynomials (in x) such that each $L_i|_{\epsilon=0}$ is a nonzero linear form. Then, there exist linear forms ℓ_1, \dots, ℓ_n , with
 679 $n - 1 \geq \text{rank}(\ell_1, \dots, \ell_n) \geq n - t$, a polynomial $S' \in \mathbb{F}[[\epsilon]][x]$, such that the following equality holds over $\mathbb{F}[[\epsilon]]$:
 680

$$681 \quad g \bmod \langle L_1, \dots, L_t \rangle = f(\ell_1, \dots, \ell_n) + \epsilon \cdot S'(x, \epsilon).$$

682
 683 Observe that in the above, we say the rank to be at least $n - t$, and not exactly $= n - t$. This is because, the
 684 linear forms $L_i(\epsilon = 0)$ might not be linearly independent. This above claim will play a key role in the proof
 685 of [Lemma 15](#).
 686

687 Finally, we also add an easy observation that would be useful later.

688
 689 **Observation 1.** Fix a variable ordering. Let $L \in \mathbb{F}[\epsilon][x]$ is a linear polynomial such that $L|_{\epsilon=0}$ is a nonzero linear
 690 polynomial. Let $\ell \in \mathbb{F}[\epsilon][x]$ be another linear polynomial such that $\ell|_{\epsilon=0}$ is a nonzero homogeneous linear polynomial.
 691 Then $L \bmod \langle \ell \rangle$ is again a linear polynomial over $\mathbb{F}[[\epsilon]]$, whose ϵ -free term is a nonzero linear polynomial. Further, if
 692 $L|_{\epsilon=0}$ is non-homogeneous, then the ϵ -free term of $L \bmod \langle \ell \rangle$ is non-homogeneous.
 693

694 The above observation can be extended to a product of linear polynomials, and also we can work with
 695 truncated ϵ -powers. In particular, if $T = \epsilon^{-a} \cdot L_1 \cdots L_s$, where $L_i \in \mathbb{F}[\epsilon][x]$ are linear polynomials (i.e. $T \in$
 696 $\mathbb{F}[\epsilon^{\pm 1}][x]$ is a product of linear polynomials), then $T \bmod \ell$, for some ℓ as above, becomes a product of linear
 697 polynomials, of the form $\epsilon^{-a} \cdot \hat{L}_1 \cdots \hat{L}_s$, where $\hat{L}_i \in \mathbb{F}[[\epsilon]][x]$ are linear polynomials. For border complexity
 698 purposes, we can still continue working over the ring $\mathbb{F}[\epsilon^{\pm 1}]$.
 699
 700

701 4 DETAILED PROOF OF THEOREM 2 FOR $k \leq 3$

702
 703 In this section, we present a self-contained proof of [Theorem 2](#) for the case $k \leq 2$, as this forms the basis for the
 704 proof in the general case of any arbitrary constant k .
 705

706 *Notations for this section.* Let $x := \{x_1, \dots, x_n\}$, fix $k \leq 2$, and $n := (k + 1)d$. Recall, $f := P_{k+1,d} =$
 707 $\sum_{i \in [k+1]} \prod_{j=1}^d x_{(i-1)d+j}$.
 708

709 **Goal:** Suppose that $f \in \overline{\Sigma^{[k]}\Pi\Sigma}$ of size s , i.e. $g = f + \epsilon \cdot S$, where $\text{size}_{\mathbb{F}(\epsilon)}(g) \leq s$ (as a $\Sigma^{[k]}\Pi\Sigma$ -circuit),
 710 $S \in \mathbb{F}[\epsilon, x]$. We want to prove that $s = 2^{\Theta(d)}$.
 711

712 4.1 $k = 1$ case

713
 714 **THEOREM 5 (IMPOSSIBILITY FOR $k = 1$).** $P_{2,d}$ cannot be expressed as a $\overline{\Sigma^{[1]}\Pi\Sigma}$ circuit.
 715

716 **PROOF.** [[BIZ18](#), Prop. A.12] showed that $\overline{\Pi\Sigma} = \Pi\Sigma$, i.e., product of linear polynomials after taking the limit
 717 remains in the same form. Eventually, the proof relies on the fact that the product distributes the ϵ -powers.
 718 Thus, it suffices to show that $P_{2,d}$ does not have any $\Pi\Sigma$ circuit. We prove it below via contradiction.
 719

720 Suppose $P_{2,d} = \ell_1 \cdots \ell_d$, where ℓ_i are linear polynomials (not necessarily distinct). Since $P_{2,d}$ is homogeneous,
 721 we can assume that each ℓ_i is a homogeneous linear form. Assume, without loss of generality, that ℓ_1 has
 722 variable x_1 , i.e. $\ell_1 = ax_1 + \ell_{11}$, where $a \in \mathbb{F}$ and ℓ_{11} is a nonzero x_1 -free linear form; ℓ_{11} must be nonzero,
 723 otherwise x_1 divides $P_{2,d}$, a contradiction. By substituting $x_1 = -\ell_{11}/a$, on both side, we get that
 724

$$725 \quad \ell_{11}x_2 \cdots x_d = ax_{d+1} \cdots x_{2d}.$$

726
 727 This can never hold because x_2 appears on the LHS, but not on the RHS. This finishes the proof. \square
 728

4.2 $k = 2$ case: The basic induction idea

The proof for $k = 2$ is far from obvious and different from that of $k = 1$. Note that $\overline{\Sigma^{[2]}\Pi\Sigma}$ is not closed under taking limit⁷. The idea is to reduce $k = 2$ case to $k = 1$ case carefully. The reduction may make the term *complicated*, however, the wish would be to show that it is still a model where exponential lower bounds can be shown. Formally, we prove the following.

THEOREM 6. $P_{3,d}$ requires $2^{\Theta(d)}$ -size $\overline{\Sigma^{[2]}\Pi\Sigma}$ circuits.

PROOF. Let $\mathbf{x} = \{x_1, \dots, x_n\}$, where $n := 3d$. Recall, $P_{3,d} := x_1 \cdots x_d + x_{d+1} \cdots x_{2d} + x_{2d+1} \cdots x_{3d}$. Suppose,

$$g := T_1 + T_2 = P_{3,d} + \epsilon \cdot S, \quad (3)$$

where the polynomials $T_1, T_2 \in \mathbb{F}[\epsilon^{\pm 1}][\mathbf{x}]$, and each of them is a product of linear polynomials $\Pi\Sigma$ and they have size at most s over $\mathbb{F}[\epsilon^{\pm 1}]$, and $S \in \mathbb{F}[\epsilon][\mathbf{x}]$. Suppose, $T_i = \epsilon^{-a_i} \cdot \ell_{i,1} \cdots \ell_{i,s}$, where $a_i \in \mathbb{Z}_{\geq 0}$, and each $\ell_{i,j} \in \mathbb{F}[\epsilon][\mathbf{x}]$ are linear polynomials (in \mathbf{x}) such that $\ell_{i,j}|_{\epsilon=0} \neq 0$. Now, one of the three things can happen.

- 1 (Easy case). Both T_i have at least one linear factor, say $\ell_{1,1}$ and $\ell_{2,1}$ whose ϵ -free term is a homogeneous linear form over \mathbb{F} ;
- 2 (Intermediate case). Exactly one of T_i , say wlog, T_1 , has at least one factor, say $\ell_{1,1}$ whose ϵ -free term is a homogeneous linear form;
- 3 (Hard case). None of the factors of T_i , has ϵ -free term as a homogeneous linear form.

The first two cases can be treated straightforwardly, as we show that they are impossible, whereas the third requires a more technical analysis and in this case we show an exponential lower bound. We now proceed to examine each case in detail.

Lower bound for Case I: Consider $g \bmod \langle \ell_{1,1}, \ell_{2,1} \rangle$. By **Claim 4** (in section 3), there exists $S' \in \mathbb{F}[[\epsilon]][\mathbf{x}]$ such that

$$0 = g \bmod \langle \ell_{1,1}, \ell_{2,1} \rangle = P_{3,d}(\ell_1, \dots, \ell_n) + \epsilon \cdot S',$$

where $n - 1 \geq \text{rank}(\ell_1, \dots, \ell_n) \geq n - 2$. In particular, this means that $P_{3,d}(\ell_1, \dots, \ell_n) = 0$. Our main claim is that this is not possible.

Claim 7 (Main Claim for Case I). *If $3d - 1 \geq \text{rank}(\ell_1, \dots, \ell_{3d}) \geq 3d - 2$, then $P_{3,d}(\ell_1, \dots, \ell_n) \neq 0$.*

PROOF OF CLAIM 7. We prove the above by contradiction. Suppose, indeed, $P_{3,d}(\ell_1, \dots, \ell_n) = 0$. Then, $\ell_1 \cdots \ell_d + \ell_{d+1} \cdots \ell_{2d} + \ell_{2d+1} \cdots \ell_{3d} = 0$. Two possible cases can happen:

- (i) The rank is $3d - 1$. Without loss of generality, $\ell_1, \dots, \ell_{3d-1}$ are full-rank and ℓ_{3d} is in the span of the $3d - 1$ linear forms. Using an invertible linear transformation, $\ell_i \mapsto x_i$, for $i \in [3d - 1]$, we get the following polynomial: $x_1 \cdots x_d + x_{d+1} \cdots x_{2d} + x_{2d+1} \cdots x_{3d-1} \cdot \ell$, for some linear form $\ell \in \mathbb{F}[x_1, \dots, x_{3d-1}]$. This can never be zero, since, for example, the monomial $x_1 \cdots x_d$ can never be canceled.
- (ii) The rank is $3d - 2$. Then, by pigeonhole principle, at least one of the set of linear forms $\{\ell_1, \dots, \ell_d\}$, $\{\ell_{d+1}, \dots, \ell_{2d}\}$, $\{\ell_{2d+1}, \dots, \ell_{3d}\}$ must be full rank. Without loss of generality, $\text{rank}(\ell_1, \dots, \ell_d) = d$. Now, further one of the three cases is true: (a) $\text{rank}(\ell_{d+1}, \dots, \ell_{2d}) = d - 1$, and $\text{rank}(\ell_{2d+1}, \dots, \ell_{3d}) =$

⁷ $x_1y_1 + x_2y_2 + x_3y_3$ has a small $\overline{\Sigma^{[2]}\Pi\Sigma}$ expression, but cannot be computed by any $\Sigma^{[2]}\Pi\Sigma$ circuit

781 $d - 1$, or (b) $\text{rank}(\ell_{d+1}, \dots, \ell_{2d}) = d$, and $\text{rank}(\ell_{2d+1}, \dots, \ell_{3d}) = d - 2$, or (c) $\text{rank}(\ell_{d+1}, \dots, \ell_{2d}) =$
 782 $d - 2$, and $\text{rank}(\ell_{2d+1}, \dots, \ell_{3d}) = d$. After applying a suitable invertible linear transformation, case (a)
 783 looks like $x_1 \cdots x_d + x_{d+1} \cdots x_{2d-1} \cdot \ell + x_{2d} \cdots x_{3d-2} \cdot \ell'$, for some linear forms $\ell, \ell' \in \mathbb{F}[x_1, \dots, x_{3d-2}]$,
 784 while (b) and (c) look like $x_1 \cdots x_d + x_{d+1} \cdots x_{2d} + x_{2d+1} \cdots x_{3d-2} \cdot \ell \cdot \ell'$, for some linear forms
 785 $\ell, \ell' \in \mathbb{F}[x_1, \dots, x_{3d-2}]$. This can never be zero, since the monomial $x_1 \cdots x_d$ can never be canceled.
 786
 787
 788 □

789 Therefore, we have shown an impossibility in case I.

791 **Lower bound for Case II:** In the second case, consider $g \bmod \langle \ell_{1,1} \rangle$. Then, by [Claim 4](#), we know that

$$792 \quad T_2 \bmod \langle \ell_{1,1} \rangle = g \bmod \langle \ell_{1,1} \rangle = P_{3,d}(\ell_1, \dots, \ell_n) + \epsilon \cdot S',$$

793 where $\text{rank}(\ell_1, \dots, \ell_n) = n - 1$. Let us compare the coefficient of ϵ^0 in both side. By [Observation 1](#), the
 794 coefficient of ϵ^0 in the LHS is a product of non-homogeneous linear polynomials. However, the coefficient of ϵ^0
 795 in the RHS is $P_{3,d}(\ell_1, \dots, \ell_n)$, which is a homogeneous polynomial, a contradiction!
 796
 797
 798
 799 □

800 **Lower bound for Case III:** Before going into the technical details, let us quickly convince the readers why
 801 this is the hardest case to prove. In this case, the idea of looking modulo an ideal becomes unreliable due to
 802 inhomogeneous ideals. The primary counter example being: Let $k = 2$, and $\ell_{1,1}$ appears in T_1 such that $\ell_{1,1} = 1$
 803 and $\ell_{2,1}$ appears in T_2 where $\ell_{2,1} = 1 + \epsilon \cdot \tilde{\ell}$, for some nonzero linear form $\tilde{\ell} \in \mathbb{F}[x]$. Note that $\ell_{i,1}|_{\epsilon=0} = 1$ for
 804 $i \in [2]$; therefore, the idea that we used above for case I-II becomes nonsensical! Strikingly, this is exactly the
 805 format of expressing an arbitrary f as the ϵ^0 -coefficient of $T_1 + T_2$, [[Kum20](#), Section 3.1], where T_1 is simply an
 806 ϵ -power. This implies that the third case (we call it *all-non-homogeneous*) should somehow capture the core and
 807 complicated part of the proof.
 808
 809

810 From now on, we focus on the third case. Before going into the proof, let us revisit the basic notation:

$$811 \quad g := T_1 + T_2 = P_{3,d} + \epsilon \cdot S, \quad (4)$$

812 where the polynomials $T_1, T_2 \in \mathbb{F}[\epsilon^{\pm 1}][x]$, and each of them is a product of linear polynomials $\Pi\Sigma$ and they
 813 have size at most s over $\mathbb{F}[\epsilon^{\pm 1}]$, and $S \in \mathbb{F}[\epsilon][x]$. By assumption, $T_i = \epsilon^{-a_i} \cdot \ell_{i,1} \cdots \ell_{i,s}$, where $a_i \in \mathbb{Z}_{\geq 0}$, and each
 814 $\ell_{i,j} \in \mathbb{F}[\epsilon][x]$ are linear polynomials (in x) such that $\ell_{i,j}|_{\epsilon=0}$ is a nonzero non-homogeneous linear polynomial.
 815
 816
 817

818 **Note:** We say that each T_i satisfies the “all-non-homogeneous” property, if each $\ell_{i,j}|_{\epsilon=0}$ is a nonzero non-
 819 homogeneous linear polynomial. In this case, we will say that $P_{3,d}$ is computed by an all-non-homogeneous
 820 $\overline{\Sigma^{[2]}\Pi\Sigma}$ circuit. In general, we will use this terminology many times.
 821
 822

823 Here are the main two claims that leads to the lower bound for case III.

824 **Claim 8.** *If $P_{3,d}$ is computed by an all-non-homogeneous $\overline{\Sigma^{[2]}\Pi\Sigma}$ circuit of size s , then $P_{3,d}$ can also be computed by a*
 825 *$\overline{\Sigma\wedge\Sigma}$ circuit of size $\text{poly}(s)$.*

826
 827 **Claim 9.** *If a $\overline{\Sigma\wedge\Sigma}$ circuit computes the polynomial $P_{3,d}$, then the size must be at least $2^{\Omega(d)}$.*

828 It is clear that [Claim 8](#) and [Claim 9](#) together imply that $s \geq 2^{\Omega(d)}$, which will finish the proof for $k = 2$.

829 The proof of [Claim 9](#) is not hard. So, first, we give the proof below, before going into the proof of [Claim 8](#).

PROOF OF CLAIM 9. Let \mathcal{C} is a $\Sigma\wedge\Sigma$ circuit of size s' , that computes the polynomial $P_{3,d}$. Then, by analyzing the partial derivative space of \mathcal{C} , one can show that the dimension of its partial derivative space is bounded by $\text{poly}(s')$, see [CKW11, Lemma 10.2]. On the other hand, the partial derivative space of the polynomial $P_{3,d}$ can easily be shown to be $2^{\Omega(d)}$. This shows that $s' \geq 2^{\Omega(d)}$ as desired. \square

We now go into the most technical part of the proof for $k = 2$.

PROOF OF CLAIM 8. Recall Equation 3:

$$g := T_1 + T_2 = P_{3,d} + \epsilon \cdot S. \quad (5)$$

Here, each T_i satisfies the all-non-homogeneous property. Now, apply a simple variable-scaling map $\Phi : x_i \mapsto z \cdot x_i$. This makes z the "degree counter" as it helps track the degree of the polynomial; but more importantly, it allows us univariate derivation.

Note that $\Phi(P_{3,d}) = z^d \cdot P_{3,d}$, and $\Phi(T_i) = \epsilon^{-a_i} \cdot \Phi(\ell_{i,1}) \cdots \Phi(\ell_{i,s})$. Now, for $i \in [2]$, let $\tilde{T}_i := \Phi(\ell_{i,1}) \cdots \Phi(\ell_{i,s})$. Dividing both sides by \tilde{T}_2 , we get

$$\Phi(P_{3,d} + \epsilon \cdot S) / \tilde{T}_2 = \epsilon^{-a_2} + \Phi(T_1) / \tilde{T}_2 \xrightarrow{\text{differentiate with respect to } z} \partial_z (\Phi(P_{3,d} + \epsilon \cdot S) / \tilde{T}_2) = \partial_z (\Phi(T_1) / \tilde{T}_2), \quad (6)$$

where $\partial_z(\cdot) := d(\cdot)/dz$. This has reduced the number of summands on the right hand side to 1, although the surviving summand has become more complicated now. Further, (seemingly) we have no control on the structure on the coefficient of ϵ^0 -term.

Here are two lemmas that gist up the main technical takeaways.

Lemma 10 (Main Lemma 1). $\partial_z (\Phi(P_{3,d} + \epsilon \cdot S) / \tilde{T}_2) \in \mathbb{F}[[\epsilon, z]][x]$, and further

$$\text{coef}_{\epsilon^0 z^{d-1}} \partial_z (\Phi(P_{3,d} + \epsilon \cdot S) / \tilde{T}_2) = c' \cdot P_{3,d}(x),$$

for some nonzero constant $c' \in \mathbb{F}$.

Lemma 11 (Main Lemma 2). $\partial_z (\Phi(T_1) / \tilde{T}_2) \in \mathbb{F}[[\epsilon, z]][x]$, and further $\text{coef}_{\epsilon^0 z^{d-1}} \partial_z (\Phi(T_1) / \tilde{T}_2)$ can be computed by a $\overline{\Sigma\wedge\Sigma}$ circuit of size $O(snd)$.

Looking at Equation 6 and combining Lemma 10-11 gives us Claim 8. So, we focus on proving these two lemmas now.

PROOF OF LEMMA 10. Since we are in the third case (all-non-homogeneous), we know that $\ell_{i,j} = c_{i,j} + \tilde{\ell}_{i,j}$, where each $\tilde{\ell}_{i,j} \in \mathbb{F}[\epsilon][x]$ is a homogeneous linear polynomial, further $c_{i,j}|_{\epsilon=0} \neq 0$. Trivially, $\Phi(\ell_{i,j}) = c_{i,j} + z \cdot \tilde{\ell}_{i,j}$. Hence, $1/\Phi(\ell_{i,j}) \in \mathbb{F}[[\epsilon, z]][x]$, and therefore, $1/\tilde{T}_2 \in \mathbb{F}[[\epsilon, z]][x]$. In fact, $1/\tilde{T}_2 = c + \epsilon \cdot R(x, \epsilon, z)$, where $0 \neq c \in \mathbb{F}$, and $R \in \mathbb{F}[[\epsilon, z]][x]$. Since, $\Phi(P_{3,d} + \epsilon \cdot S) \in \mathbb{F}[\epsilon, z][x]$, clearly $\partial_z (\Phi(P_{3,d} + \epsilon \cdot S) / \tilde{T}_2) \in \mathbb{F}[[\epsilon, z]][x]$.

Now,

$$\begin{aligned}
\text{coef}_{\epsilon^0 z^{d-1}} (\partial_z (\Phi(P_{3,d} + \epsilon \cdot S) / \tilde{T}_2)) &= \text{coef}_{\epsilon^0 z^{d-1}} (\partial_z (\Phi(P_{3,d}) / \tilde{T}_2) + \epsilon \cdot \partial_z (\Phi(S) / \tilde{T}_2)) \\
&= \text{coef}_{\epsilon^0 z^{d-1}} (\partial_z (\Phi(P_{3,d}) / \tilde{T}_2)) \\
&= \text{coef}_{\epsilon^0 z^{d-1}} \left(\partial_z \left(\frac{z^d \cdot P_{3,d}}{c + \epsilon \cdot R} \right) \right) \\
&= \text{coef}_{\epsilon^0 z^{d-1}} \left(\partial_z \left(\left(\frac{z^d \cdot P_{3,d}}{c} \right) \cdot \left(1 - \epsilon \cdot R/c + \epsilon^2 (R/c)^2 - \dots \right) \right) \right) \\
&= P_{3,d} / c .
\end{aligned}$$

This proves the claim.

The above calculation also shows that the **minimum power** of z in the ϵ -free term that appears in the expression $(\partial_z (\Phi(P_{3,d} + \epsilon \cdot S) / \tilde{T}_2))$ is exactly $d - 1$. We will use this fact crucially in the next proof. \square

PROOF OF LEMMA 11. In Lemma 10, we have already shown that $(\partial_z (\Phi(P_{3,d} + \epsilon \cdot S) / \tilde{T}_2)) \in \mathbb{F}[[\epsilon, z]][\mathbf{x}]$, and therefore, from Equation 6, we can conclude that $\partial_z (\Phi(T_1) / \tilde{T}_2) \in \mathbb{F}[[\epsilon, z]][\mathbf{x}]$. To understand the expression better, we use *logarithmic derivative* (aka dlog), which has a bunch of helpful properties; see section 3. Recall the notations: $\Phi(T_i) = \epsilon^{-a_i} \cdot \tilde{T}_i$, where $\tilde{T}_i := \Phi(\ell_{i,1}) \cdots \Phi(\ell_{i,s})$, where $\Phi(\ell_{i,j}) = c_{i,j} + z \cdot \tilde{\ell}_{i,j}$, for some linear form $\ell_{i,j} \in \mathbb{F}[\epsilon][\mathbf{x}]$. Then, the expression $\partial_z (\Phi(T_1) / \tilde{T}_2)$ can be re-written as

$$\partial_z (\Phi(T_1) / \tilde{T}_2) = \epsilon^{-a_1} \cdot \partial_z (\tilde{T}_1 / \tilde{T}_2) \quad (7)$$

$$= \epsilon^{-a_1} \cdot (\tilde{T}_1 / \tilde{T}_2) \cdot \text{dlog} (\tilde{T}_1 / \tilde{T}_2) \quad (8)$$

$$= \epsilon^{-a_1} \cdot (\tilde{T}_1 / \tilde{T}_2) \cdot (\text{dlog}(\tilde{T}_1) - \text{dlog}(\tilde{T}_2)) . \quad (9)$$

Since each $c_{i,j}|_{\epsilon=0} \neq 0$, a nonzero constant, observe that $\text{val}_\epsilon(\tilde{T}_1 / \tilde{T}_2) = 0$, and further it is easy to see that $\tilde{T}_1 / \tilde{T}_2 \in \mathbb{F}[[\epsilon, z]][\mathbf{x}]$. Moreover, from the proof of Lemma 10, we know that $\text{val}_\epsilon(\partial_z (\Phi(P_{3,d} + \epsilon \cdot S) / \tilde{T}_2)) = 0$. Therefore, from Equation 6, one can conclude that $\text{val}_\epsilon(\epsilon^{-a_1} \cdot (\text{dlog}(\tilde{T}_1) - \text{dlog}(\tilde{T}_2))) = 0$.

Since the dlog operator distributes the product terms (see Section 3), by the discussion in Section 3 and Section 2, we get that

$$\text{dlog}(\tilde{T}_i) = \sum_{j=1}^s \text{dlog}(\Phi(\ell_{i,j})) = \sum_{j=1}^s \left(\frac{\tilde{\ell}_{i,j}}{c_{i,j} + z \cdot \tilde{\ell}_{i,j}} \right) = \sum_{j \geq 0} P_{i,j}(\mathbf{x}, \epsilon) \cdot z^j . \quad (10)$$

In the above expression, each $P_{i,j}$ is an $\Sigma \wedge \Sigma$ expression of size $O(snj)$, over $\mathbb{F}[[\epsilon]]$. Therefore, looking at Equation 7, and we get that

$$\text{coef}_{\epsilon^0 z^{d-1}} (\partial_z (\Phi(T_1) / \tilde{T}_2)) = \text{coef}_{z^{d-1}} \left(\text{coef}_{\epsilon^0} \left((\tilde{T}_1 / \tilde{T}_2) \cdot \left(\sum_{j \geq 0} Q_j z^j \right) \right) \right) \quad (11)$$

$$= \text{coef}_{z^{d-1}} \left(\text{coef}_{\epsilon^0} (\tilde{T}_1 / \tilde{T}_2) \right) \cdot \left(\text{coef}_{\epsilon^0} \left(\sum_{j \geq 0} Q_j z^j \right) \right) . \quad (12)$$

In the above, $Q_j := \epsilon^{-a_1} \cdot (P_{1,j} - P_{2,j})$, and each Q_j has a $\Sigma \wedge \Sigma$ expression of size $O(snj)$, over $\mathbb{F}[[\epsilon]]$.

Let $\tilde{\ell}_{i,j}|_{\epsilon=0} =: \tilde{\ell}_{i,j,0} \in \mathbb{F}[\mathbf{x}]$, and $\mathbb{F} \ni c_{i,j,0} := c_{i,j}(\epsilon = 0)$. Then, observe that

$$\text{coef}_{\epsilon^0} \left(\frac{\tilde{T}_1}{\tilde{T}_2} \right) = \frac{\prod_{j=1}^s (c_{1,j,0} + z \cdot \tilde{\ell}_{1,j,0})}{\prod_{j=1}^s (c_{2,j,0} + z \cdot \tilde{\ell}_{2,j,0})} .$$

In the proof of [Lemma 10](#), we have shown that the minimum z power in the term ϵ^0 in $\partial_z (\Phi(T_1)/\tilde{T}_2)$ is $d-1$. Since, $\text{val}_z(\text{coef}_{\epsilon^0}(\tilde{T}_1/\tilde{T}_2)) = 0$, looking at [Equation 11](#), it must happen that $\text{val}_\epsilon(Q_j) \geq 1$, for all $0 \leq j \leq d-2$. Therefore, there exists some constant c'' such that

$$\text{coef}_{\epsilon^0 z^{d-1}}(\partial_z(\Phi(T_1)/\tilde{T}_2)) = \left(\frac{\prod_{j=1}^s c_{1,j,0}}{\prod_{j=1}^s c_{2,j,0}} \right) \cdot \text{coef}_{\epsilon^0}(Q_{d-1}) = c'' \cdot \text{coef}_{\epsilon^0}(Q_{d-1}).$$

Since we already argued that Q_{d-1} has a $\Sigma\wedge\Sigma$ expression of size $O(snd)$ over $\mathbb{F}[[\epsilon]]$, we conclude that $\text{coef}_{\epsilon^0 z^{d-1}}(\partial_z(\Phi(T_1)/\tilde{T}_2))$ can be computed by a $\overline{\Sigma\wedge\Sigma}$ circuit of size $O(snd)$.

This finishes the proof of [Lemma 11](#). \square

We have already argued that proving [Lemma 10-11](#) proves [Claim 8](#), which is the most technical part of $k=2$ proof. And, as we already mentioned before, [Claim 8](#) and [Claim 9](#) finish the proof of the lower bound for case III. Since we have covered case I-III, we have formally proved that $P_{3,d}$ requires exponential-size $\overline{\Sigma^{[2]}\Pi\Sigma}$. \square

4.3 $k=3$ case: The bloated induction idea

For the $k=3$ case, one can similarly show that the hardest case is the setting of *all-non-homogeneous* like in the last section, e.g. [Equation 4](#). So, let us focus on that particular setting and set the following notation:

$$g_0 = f_0(x) + \epsilon \cdot S_0 = \sum_{i=1}^3 T_{i,0}. \quad (13)$$

Here, $n := 4d$, and $f_0 := P_4$. Each $T_{i,0}$ is computable by a $\Pi\Sigma$ -circuit of size at most s over $\mathbb{F}[\epsilon^{\pm 1}]$ with the property: after dividing by the ϵ -power of ϵ -valuation, it is a *nonzero constant* mod $\langle x_1, \dots, x_n \rangle$. Further, let $d_0 := d+1$.

This example leads us to the ‘main reduction’ lemma that reduces the fanin to 1, yielding a ratio of two depth-3 diagonal circuits of small size. The proof of the lemma is an algebraic manipulation that applies the ‘linearizing’ dlog operator, much like the induction step invented in [Claim 8](#). Note: Now we get a slightly more complicated model that is the *ratio* of two border-depth-3-diagonal circuits.

Lemma 12 (Main [Lemma 18](#)’s $k=3$ case). *If f can be approximated by a structured $\Sigma^{[3]}\Pi\Sigma$ -circuit of size s (i.e. as written in [Equation 13](#)), over $\mathbb{F}(\epsilon)$, then there exist two polynomials $F_1, F_2 \in \mathbb{F}[x]$ such that each F_i can be computed by a $\overline{\Sigma\wedge\Sigma}$ circuit of size $\text{poly}(s)$, and $f = F_1/F_2$.*

In the above, when we say *structured*, we mean the all-non-homogeneous setting, as mentioned before.

PROOF. We will prove it by reducing the top-fanin to 1 in two steps ($j=0 \rightarrow 1 \rightarrow 2$).

From the previous subsection, recall the variable-scaling map $\Phi : x_i \mapsto z \cdot x_i$. To save space, assume that in [Equation 13](#), Φ has been applied. Thus, our initial model becomes: $\sum_{i \in [3]} T_{i,0} =: g_0$, over $\mathcal{R}_0(x, \epsilon)$, such that it approximates f_0 correctly, where $f_0 \in \mathcal{R}_0(x)$, where $\mathcal{R}_0 := \mathbb{F}[z]/\langle z^{d_0} \rangle$, for any $d_0 > \deg(f)$.

For the purpose of induction, we now define a more general, or *bloated*, class as,

$$\text{Gen}(3, s) := \Sigma^{[3]}(\Pi\Sigma/\Pi\Sigma)(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma).$$

This model computes functions of the form $\sum_{i=1}^3 (U_i/V_i) \cdot (P_i/Q_i)$, where $U_i, V_i \in \Pi\Sigma$, invertible in the underlying ring, and $P_i, Q_i \in \Sigma\wedge\Sigma$, and the circuit (with division allowed) has size s ; see [Definition 2](#). Informally speaking, when we “divide and derive” twice, we will land in this model (unlike for $k=2$, where single

“division and derivative” was enough). From g_0 we will derive g_1 , followed by g_2 , using algebraic manipulations as described next ($j = 0 \rightarrow 1 \rightarrow 2$).

So, we think of $g_0 = \sum_{i \in [3]} T_{i,0}$ as the $\text{Gen}(3, s)$ -type expression $\sum_{i=1}^3 (U_{i,0}/V_{i,0}) \cdot (P_{i,0}/Q_{i,0})$.

Let $v_{i,j} := \text{val}_z(T_{i,j}) \geq 0$, for $i \in [3-j]$. Moreover, $U_{i,j}|_{z=0} \in \mathbb{F}(\epsilon) \setminus \{0\}$ (similarly for $V_{i,j}$), i.e. they are *units* in $\mathbb{F}(\epsilon)$ due to the hypothesis of all-non-homogeneous.

Divide and Derive. Let $T_{3-j,j} =: \epsilon^{a_{3-j,j}} \cdot \tilde{T}_{3-j,j}$, where $\tilde{T}_{3-j,j} =: (t_{3-j,j} + \epsilon \cdot \tilde{t}_{3-j,j})$ is not divisible by ϵ . Divide $g_j =: f_j + \epsilon \cdot S_j$, by $\tilde{T}_{3-j,j}$, to get:

$$\begin{aligned} f_j / \tilde{T}_{3-j,j} + \epsilon \cdot S_j / \tilde{T}_{3-j,j} &= \epsilon^{a_{3-j,j}} + \sum_{i \in [2-j]} T_{i,j} / \tilde{T}_{3-j,j} \\ \implies \partial_z (f_j / \tilde{T}_{3-j,j}) + \epsilon \cdot \partial_z (S_j / \tilde{T}_{3-j,j}) &= \sum_{i \in [2-j]} \partial_z (T_{i,j} / \tilde{T}_{3-j,j}) \\ &= \sum_{i \in [2-j]} (T_{i,j} / \tilde{T}_{3-j,j}) \cdot \text{dlog} (T_{i,j} / \tilde{T}_{3-j,j}) \\ &=: g_{j+1}. \end{aligned} \quad (14)$$

The base ring. Let $\mathcal{R} := \mathbb{F}(z) \subseteq \mathbb{F}(\widehat{\mathcal{R}}) =: \widehat{\mathcal{R}}$. It helps to keep a standard fact in mind while going through our calculations: each rational function has a *Laurent series*⁸. Also, the above equation motivates, for $i \in [2-j]$, the definition of a new model that is needed to move to the next level of induction,

$$T_{i,j+1} := (T_{i,j} / \tilde{T}_{3-j,j}) \cdot \text{dlog} (T_{i,j} / \tilde{T}_{3-j,j}), \text{ and } f_{j+1} := \partial_z (f_j / \tilde{T}_{3-j,j}).$$

Claim 13 (Induction hypotheses over the new base ring). *For each $0 \leq j \leq 1$, the circuit g_{j+1} approximates f_{j+1} correctly, i.e. $\lim_{\epsilon \rightarrow 0} g_{j+1} = f_{j+1}$, where g_{j+1} (respec. f_{j+1}) is well-defined in the ring $\mathcal{R}(x, \epsilon) \subset \widehat{\mathcal{R}}(x, \epsilon)$ (respec. $\widehat{\mathcal{R}}(x)$).*

PROOF. The key observation is: f_j and $T_{i,j}$'s are elements in $\mathbb{F}(x, z, \epsilon)$ which also belong to $\mathbb{F}(x, \epsilon)((z))$, and f_j is ϵ -free, $\forall i \in [2-j]$.

Equation 14 holds over $\mathcal{R}(x, \epsilon)$; this is because of the division by z -valuation of $v_{3-j,j}$ and then differentiation. It is easy to see that the least ϵ -power in the sum $f_j + \epsilon \cdot S_j$ is contributed by f_j and not by $\epsilon \cdot S_j$. Further, since $\lim_{\epsilon \rightarrow 0} \tilde{T}_{3-j,j} = t_{3-j,j}$, we must have that $f_j / t_{3-j,j} \in \mathcal{R}(x)$ and thus the ‘limit’ f_{j+1} exists over the base ring \mathcal{R} .

In other words, $\lim_{\epsilon \rightarrow 0} \partial_z (g_j / \tilde{T}_{3-j,j}) = \lim_{\epsilon \rightarrow 0} g_{j+1} = f_{j+1} \in \mathcal{R}(x)$. This finishes Claim 13. \square

The primary structural claim is as follows.

Claim 14. *For $0 \leq j \leq 2$, $g_j \in \text{Gen}(3-j, s_j)$, where $s_j = \text{poly}(sd)$.*

PROOF. One can prove this inductively on j . It is trivially true for $j = 0$, with $s_0 = s$. Assume this holds for g_j , and we will show this for $j + 1$.

Note that $\text{dlog}(\Sigma \wedge \Sigma) \in \Sigma \wedge \Sigma / \Sigma \wedge \Sigma$, where each $\Sigma \wedge \Sigma$ circuits are of poly-size (an easy proof is in Lemma 28).

Since dlog distributes the product *additively*, so it suffices to work with $\text{dlog}(\Pi \Sigma)$; and we showed that $\text{dlog}(\Pi \Sigma) \in \Sigma \wedge \Sigma$, is of poly-size if truncated (see section 3, and explicitly in Equation 10); we will also reprove it in the size analysis (see subsection 6.3). Let us assume that at j -th step, we will truncate till $d_j - 1$, i.e. consider

⁸Since the target f is a polynomial, our proof requires only a *finite* truncation of this Laurent series; this then is in the ring $\mathbb{F}(x)[z^{\pm 1}]$.

the equation $\text{trun } z^{d_j}$, where d_j is polynomially bounded. One can simplify to get the following:

$$T_{i,j}/\tilde{T}_{3-j,j}(\text{trun } z^{d_j}) = \epsilon^{-a_{3-j,j}} \cdot \left((U_{i,j} \cdot V_{3-j,j}) / (V_{i,j} \cdot U_{3-j,j}) \right) \cdot (P_{i,j} \cdot Q_{3-j,j}) / (Q_{i,j} \cdot P_{3-j,j}) (\text{trun } z^{d_j}),$$

and its dlog. Let $U_{i,j+1} := U_{i,j} \cdot V_{3-j,j}$; similarly $V_{i,j+1} := V_{i,j} \cdot U_{3-j,j}$. Essentially, dlog computation will produce $(\Sigma \wedge \Sigma / \Sigma \wedge \Sigma)$ -circuits, which will further multiply with P 's and Q 's, and we multiply $\epsilon^{-a_{3-j,j}}$ there; for details see [Claim 21](#). This is done to ensure the invariant: $U_{i,j+1}|_{z=0}, V_{i,j+1}|_{z=0} \in \mathbb{F}(\epsilon) \setminus \{0\}$, so, U 's and V 's remain units in the z -power-series ring.

Therefore, $T_{i,j+1}(\text{trun } z^{d_j}) = (T_{i,j}/\tilde{T}_{3-j,j}) \cdot \text{dlog} (T_{i,j}/\tilde{T}_{3-j,j}) (\text{trun } z^{d_j})$ is of the form $(\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$, with polynomial blow up in the size. Therefore, $g_{j+1} \in \text{Gen}(m-j-1, s_{j+1})$, where $s_j = \text{poly}(sd)$ as desired. \square

As we mentioned before, at each step, we will do truncation $\text{trun } z^{d_j}$, i.e. each $T_{i,j}$ is computed as an element in $\mathbb{F}(x, z, \epsilon)$. This viewpoint is important because it allows us to extract the coefficient of a power of z , (or ϵ^0), if required. Also, $v_{i,j}$ can become negative, but they still remain polynomially bounded, since we have the following relation $v_{i,j+1} := v_{i,j} - v_{i,3-j}$, and they are all polynomially bounded.

Extracting the right z -power to get f . As we move from $j = 0 \rightarrow 1 \rightarrow 2$ (with the given f free of z, ϵ), we show that f_j evolves rather predictably and keeps computing a 'multiple' of f , namely: $\text{coef}_{z^{b_j}}(f_j) = \lim_{\epsilon \rightarrow 0} f \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$, for appropriate degree bounds b_j (which may be positive or negative⁹, but polynomially bounded in the magnitude).

Since $\tilde{T}_{m-j,j} \in (\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$, the term $t_{m-j,j}$, which is the ϵ -free part of is of $\tilde{T}_{m-j,j}$, must be of the form $\lim_{\epsilon \rightarrow 0} (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$. Here, we used that $\lim_{\epsilon \rightarrow 0} (\Pi\Sigma/\Pi\Sigma)$ is a nonzero constant in \mathbb{F} , since $\Pi\Sigma|_{z=0}$ part in the equation is a unit, in $\mathbb{F}(\epsilon) \setminus \{0\}$.

Let $\text{val}_z(t_{3-j,j}) = d'_j$. Therefore, $\text{coef}_{z^{d'_j}}(t_{3-j,j}) = \lim_{\epsilon \rightarrow 0} \Sigma\wedge\Sigma/\Sigma\wedge\Sigma$, where these $\Sigma\wedge\Sigma$ circuits are z -free. Here, we use the fact that $\Sigma\wedge\Sigma$ is closed under interpolation and substitution. We will prove that everything remains polynomially bounded in the size-analysis section ([subsection 6.3](#)).

Also, by definition, $\text{val}_z(f_{j+1}) = b_{j+1}$, implying $\text{val}_z(f_j/t_{3-j,j}) = b_{j+1} + 1$. Therefore, $b_{j+1} + 1 = \text{val}_z(f_j) - \text{val}_z(t_{3-j,j}) = b_j - d'_j$. Hence,

$$\text{coef}_{z^{b_j-d'_j}}(f_j/t_{3-j,j}) = \text{coef}_{z^{b_j}}(f_j) / \text{coef}_{z^{d'_j}}(t_{3-j,j}) = \lim_{\epsilon \rightarrow 0} f \cdot \Sigma\wedge\Sigma/\Sigma\wedge\Sigma.$$

It is very important to note that the above goes through since we are only looking at the coefficient of the minimum- z power, and also $\Sigma\wedge\Sigma$ is closed under multiplication (with polynomial blowup in size; see [Lemma 25](#)). Since the RHS is z -free, and $b_{j+1} = b_j - d'_j - 1$, it directly follows that

$$\text{coef}_{z^{b_{j+1}}}(f_{j+1}) = \text{coef}_{z^{b_{j+1}}}(\partial_z (f_j/t_{3-j,j})) = \lim_{\epsilon \rightarrow 0} f \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma).$$

Since $f_{j+1} = \lim_{\epsilon \rightarrow 0} g_{j+1}$, we get:

$$\text{coef}_{z^{b_{j+1}}}(\lim_{\epsilon \rightarrow 0} g_{j+1}) = \text{coef}_{z^{b_{j+1}}}(f_{j+1}) = \lim_{\epsilon \rightarrow 0} f \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma) \quad (15)$$

This completes the inductive proof of the connection between g_2 and f .

Wrapping up. We have, as just shown above, $\text{coef}_{z^{b_2}}(\lim_{\epsilon \rightarrow 0} g_2) = f \cdot (\overline{\Sigma\wedge\Sigma/\Sigma\wedge\Sigma})$. On the other hand, doing 'divide and derive' twice (see [Equation 14](#)) has given us $g_2 \in \text{Gen}(1, s_2)$ where $s_2 = \text{poly}(sd)$ (see [Claim 14](#)): $\lim_{\epsilon \rightarrow 0} g_2 = \lim_{\epsilon \rightarrow 0} (\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$ with the $(\Pi\Sigma)$ -part being all-non-homogeneous. Again, since

⁹E.g. $\lim_{\epsilon \rightarrow 0} \frac{z+\epsilon}{z^2+\epsilon} = \frac{1}{z}$.

1093 we are interested in the *minimum* z -power, and $\Pi\Sigma|_{z=0}$ are nonzero elements of $\mathbb{F}(\epsilon)$, the contribution comes
 1094 only from the $\Sigma\wedge\Sigma$ terms. In particular, we deduce:

$$1096 \quad \lim_{\epsilon \rightarrow 0} g_2 = \lim_{\epsilon \rightarrow 0} (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma) \implies \text{coef}_{z^{b_2}}(\lim_{\epsilon \rightarrow 0} g_2) = \text{coef}_{z^{b_2}}(\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma}) = \overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma}$$

1097
 1098 Again, we use the fact that $\Sigma\wedge\Sigma$ is closed under interpolation (with polynomial blow up in size). Comparing
 1099 this with Equation 15, one can easily conclude that $f = \overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma}$. The size of this representation can be,
 1100 quite straightforwardly, checked to be $\text{poly}(s)$. \square
 1101

1102
 1103 Finally, we prove a simple analog of Claim 9 (for ratio of two $\overline{\Sigma\wedge\Sigma}$ circuits) to finish the exponential
 1104 lower-bound proof for $P_{4,d}$ (see Lemma 23).
 1105

1106 4.4 Brief pathway to general k

1107
 1108 The proof for the general k is similar to $k = 3$, and it invokes the general bloated class over *any* ring R :

$$1109 \quad \text{Gen}(k, s) := \Sigma^{[k]}(\Pi\Sigma/\Pi\Sigma)(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma) .$$

1110
 1111 This model computes functions of the form $\Sigma_{i=1}^k (U_i/V_i) \cdot (P_i/Q_i)$, where $U_i, V_i \in \Pi\Sigma$, invertible in the
 1112 underlying ring, and $P_i, Q_i \in \Sigma\wedge\Sigma$, and the circuit (with division allowed) has size s (over R); see Definition 2.
 1113 Informally speaking, when we divide an derive for more than one time, we will land up in these kind of models,
 1114 and for $k = 2$, only one division and derivative was enough. Also, we will work with $\widehat{\mathcal{R}}$ as the base ring.
 1115

1116 For general $k > 3$, the proof consists of four claims.
 1117

1118 (1) if $P_{k+1,d}$ was computed by a $\overline{\Sigma^{[k]}\Pi\Sigma}$ -circuit of size s , then there are linear forms ℓ_1, \dots, ℓ_n , with
 1119 $n := (k+1)d$, such that $\text{rank}(\ell_1, \dots, \ell_n) \geq n - k$, and $P_{k+1,d}(\ell_1, \dots, \ell_n)$ can be computed by an
 1120 ‘all-non-homogeneous’ $\overline{\Sigma^{[m]}\Pi\Sigma}$ -circuit of size $O(sn)$, for some $m \leq k$. Formally,
 1121

$$1122 \quad P_{k+1,d}(\ell_1, \dots, \ell_n) + \epsilon \cdot S = T_1 + \dots + T_m , \tag{16}$$

1123
 1124 where $S \in \mathbb{F}[\epsilon][x]$, and each T_i satisfies the *all-non-homogeneous* property, i.e. if $T_i := \epsilon^{-a_i} \cdot \ell_{i,1} \cdots \ell_{i,s}$,
 1125 where $\ell_{i,j} \in \mathbb{F}[\epsilon][x]$, such that $\ell_{i,j}|_{\epsilon=0}$ is a nonzero non-homogeneous linear polynomial.
 1126

1127 This step is collecting homogeneous linear forms from each product (if exists) and taking modulo all
 1128 those linear forms, as already done for the $k = 2$ case in the previous section. For a formal proof, see
 1129 Lemma 15. After this point onward, the proof is essentially the one described in subsection 4.3.
 1130

1131 (2) We start with applying $x_i \mapsto zx_i$, and then do division and derivative multiple times. Initially we start
 1132 with $\text{Gen}(m, s)$, as seen above in Equation 16, with the $\Sigma\wedge\Sigma$ circuits being just 1. After i -many divisions
 1133 and derivatives, we will still try to argue that the model becomes $\text{Gen}(m - i, s_i)$, where $s_i = s^{\exp(i)}$,
 1134 and more interestingly, the main claim would be:
 1135

$$1136 \quad g_i + \epsilon \cdot S_i \in \text{Gen}(m - i, s_i) ,$$

1137
 1138 where $\text{coef}_{\epsilon^0 z^b}(g_i) = P_{k+1,d}(\ell_1, \dots, \ell_n) \cdot (W_{i,1}/W_{i,2})$, for some b (and in fact that b would be the
 1139 minimum z -power occurring in the RHS), and $W_{i,j}$ has $\overline{\Sigma\wedge\Sigma}$ -size circuits of size at most s_i .
 1140

1141 In particular, this means that when $i = m - 1$, is the coefficient of the *minimum* z -power in the limit
 1142 of a small $\text{Gen}(1, s^{\exp(m)})$ circuit, see induction hypothesis 4 in Part II discussion of Section 6.1 and
 1143 Section 6.2.
 1144

- (3) the coefficient of the minimum z -power in $\text{coef}_{\epsilon^0}(\text{Gen}(1, s^{\exp(m)}))$ is a ratio of two $s^{\exp(m)}$ -size $\overline{\Sigma\wedge\Sigma}$ circuits (Lemma 18). By combining this claim with the previous claim (for $i = m - 1$), and using the fact that product of two $\overline{\Sigma\wedge\Sigma}$ circuits is again a small $\overline{\Sigma\wedge\Sigma}$ circuit (with polynomial blowup in size), it means that $P_{k+1,d}(\ell_1, \dots, \ell_d)$ must be a ratio of two $\overline{\Sigma\wedge\Sigma}$ circuits, each of size $s^{\exp(m)}$. Both Points (2) and (3) require the modified DiDIL framework introduced in [DDS21]. However, structurally, we prove something stronger here, by carefully analyzing the coefficient of the minimum z -power in the ϵ^0 -term.
- (4) If $P_{k+1,d}(\ell_1, \dots, \ell_d)$ can be written as a ratio of two $\overline{\Sigma\wedge\Sigma}$ circuits, i.e. $P_{k+1,d}(\ell_1, \dots, \ell_d) = F_1/F_2$ where each F_i can be computed by $\overline{\Sigma\wedge\Sigma}$ circuit of size at most $s^{\exp(m)}$, then the size of the circuits must be $2^{\Omega(d)}$ (Lemma 23). This would prove that s must be exponential (since $m \leq k$ is a constant). Here, we use the cone-based argument (for definition, see section 2). Note that for $k = 2$, it was not a ratio, but simply a $\overline{\Sigma\wedge\Sigma}$ circuit; however, a similar lower bound technique can be lifted for a ratio as well. This would finish the proof.

4.5 (im) Possibility of a ‘simpler’ proof, and comparison with Dutta, Dwivedi and Saxena [DDS21]

The crux of this lower bound approach is to — (1) ‘convert’ the general problem into all-non-homogeneous setting (to reap the advantages of $\Pi\Sigma|_{z=0=\epsilon}$ being a nonzero constant), (2) use the non-homogeneity to express the input polynomial as a ratio of two border depth-3 diagonal circuits, and finally (3) use partial derivative techniques to show that such a robust polynomial like $P_{k,d}$ cannot be written as a ratio of two polynomial-size $\overline{\Sigma\wedge\Sigma}$ circuits.

So, one would wonder whether we can just *shift* the x -variables randomly at first, achieving the desired non-homogeneity, and proceed without doing the case-analysis! However, if we work with $P_{k+1,d}(x + \mathbf{a})$, for some $\mathbf{a} \in \mathbb{F}^n$, and variable-scale to get $P_{k+1,d}(z\mathbf{x} + \mathbf{a})$, this becomes hard to handle even after one division and derivation (Section 6.2). Since $\partial_z(P_{k+1,d}(z\mathbf{x} + \mathbf{a}))$ ‘spreads’ the coefficients of $P_{k+1,d}$ across different z -powers; we cannot anymore use the second step of the argument described in the previous section. Basically, the power of z comes from the polynomial multiplication *convolution*, and this makes it hard to do any further analysis. For slightly more detailed understanding of the complications, see below.

Shift complicates for $k = 2$. Suppose that for $k = 2$, there exists a size- s depth-3 fanin-2 circuit approximating $P_{3,d}$. After shifting and scaling, dividing, deriving, using dlog and finally taking the limit, we will get the following:

$$\partial_z (P_{3,d}(z\mathbf{x} + \mathbf{a})/\Pi\Sigma) \equiv \overline{\Sigma\wedge\Sigma} \cdot (\Pi\Sigma) \pmod{z^d},$$

where $\overline{\Sigma\wedge\Sigma}$ and $\Pi\Sigma$ circuits are of size $\text{poly}(s)$. From this expression, showing an exponential lower bound on s is not clear at all. In our proof sketch in the previous two sections, we claimed that the minimum z -degree carries the full ‘hardness information’ of $P_{3,d}$ (see proof sketch of Claim 8, and more details in Section 6.2, induction-hypothesis-(IV)). Unfortunately, the shift by \mathbf{a} kills this property.

Difference and similarities with [DDS21] when $k \geq 3$. The homogeneity of the polynomials is crucial for our proof to work, it requires us to reduce the border circuit to all-non-homogeneous setting (namely, the ‘hard case 3’ above). We do this carefully *without* any variable-shift (Lemma 15), so that the homogeneity of f is maintained and yet ‘all’ the linear functions are invertible! For a ‘coarser’ upper bound of ABP, [DDS21] did not require these innovations.

We also remark that without the shift, it is not clear whether we can ‘lift’ and de-border f . The lifting in [DDS21] required interpolation, which is why they moved to the model ABP/ABP, instead of our weaker model $\overline{\Sigma \wedge \Sigma} / \overline{\Sigma \wedge \Sigma}$.

After reducing to the all-non-homogeneous case, we do use the DiDIL technique to analyze $\overline{\Sigma^{[k]} \Pi \Sigma}$, introduced in [DDS21]. DiDIL is an acronym for the steps: **D**ivide, **D**erive, **I**nduct, with **L**imit. In this paper, DiDIL process is applied in a new, somewhat simplified, setting; namely, to the all-non-homogeneous case spelled in Lemma 15. Since, we do not use any shifting unlike in [DDS21], this really gives us the advantage to closely study the target polynomial $P_{k+1,d}$, and yields certain ‘bloated’ structures, which need an intricate analysis; for details see Section 6.1-6.2. In other words, our method of first reducing to all-non-homogeneous and then using the DiDIL analysis and finally looking at the minimum z -degree is tailor-made to work for the lower bound, and not the upper bound.

5 HARDNESS LIES IN ALL-NON-HOMOGENEITY

We want to prove lower bounds on border depth-3 circuits $\overline{\Sigma^{[k]} \Pi^{[d]} \Sigma}$. By definition, $f(x) + \epsilon \cdot S(x, \epsilon) = \sum_{i \in [k]} T_i$, such that each $T_i = \prod_{j \in [d]} \ell_{ij}$, where $\ell_{ij} \in \mathbb{F}[\epsilon^{\pm 1}][x]$, are linear polynomials. In the homogeneous setting, an exponential lower bound is easy-to-show (see Corollary 16).

Since Kumar’s expression [Kum20] also involves the polynomials T_i with each linear factor being strictly non-homogeneous, the all-non-homogeneous case should be the hardest. We also briefly mentioned this being the hardest to analyze, in Section 4. Here is the formal reduction lemma, which shows that proving a lower bound for $\overline{\Sigma^{[k]} \Pi^{[d]} \Sigma}$, reduces to proving for the ‘all-non-homogeneous’ case, i.e. every linear factor in T_i has the minimum ϵ -degree term to be a non-homogeneous linear polynomial (or a constant). The following lemma holds for any general n -variate homogeneous polynomial $f(x)$.

Lemma 15 (Reduction to all-non-homogeneous setting). *If an n -variate polynomial $f(x)$ is computed by a $\overline{\Sigma^{[k]} \Pi \Sigma}$ -circuit of size s , then there exists $0 \leq m \leq k$, and homogeneous linear forms ℓ_1, \dots, ℓ_n such that*

- 1 (large rank). $\text{Rank}(\ell_1, \dots, \ell_n) = n - t$, for some $0 \leq t \leq k - m$,
- 2 (all-non-homogeneous). There exists $H(x, \epsilon) \in \mathbb{F}[x, \epsilon]$ such that

$$f(\ell_1, \dots, \ell_n) + \epsilon \cdot H(x, \epsilon) = T_1 + \dots + T_m,$$

where each $T_i \in \mathbb{F}[\epsilon^{\pm 1}][x]$ is a product of linear polynomials, and $\forall i \in [m]$,

$$\left(\frac{T_i}{\epsilon^{\text{val}_\epsilon(T_i)}} \right) \Big|_{x=\epsilon=0} \in \mathbb{F} - \{0\} \text{ and } \sum_i \text{size}_{\mathbb{F}(\epsilon)}(T_i) \leq O(s \cdot n).$$

This makes sure that after the reduction, each T_i is of the form $\epsilon^{-a_i} \cdot \ell_{i,1} \cdots \ell_{i,s}$, where $\ell_{i,j} \in \mathbb{F}[\epsilon][x]$ are linear polynomials such that $\ell_{i,j}|_{\epsilon=0}$ is a non-homogeneous nonzero linear polynomial.

PROOF OF LEMMA 15. The proof almost immediately follows from Claim 4. Suppose, $f + \epsilon \cdot S(x, \epsilon) = \hat{T}_1 + \dots + \hat{T}_k$, where $\hat{T}_i \in \mathbb{F}[\epsilon^{\pm 1}][x]$, product of linear polynomials. Now if already each T_i is such that after taking out the maximum ϵ -power, ϵ -free term $\text{mod } \langle x_1, \dots, x_n \rangle \in \mathbb{F} - \{0\}$, then we are already in the required setup and we do not need to reduce further. Otherwise, at least one \hat{T}_i must have a linear factor whose ϵ -free term is a homogeneous linear form. Collect such linear polynomials L_1, \dots, L_t , each from one \hat{T}_i , if exists. Consider the given equation $\text{mod } \mathcal{I}$, where $\mathcal{I} := \langle L_1, \dots, L_t \rangle$.

Without loss of generality, assume that $\hat{T}_i = 0 \pmod{\mathcal{I}}$, for $i \in [m+1, k]$, and first m -many \hat{T}_i survive. Please note that while doing modulo \mathcal{I} , more than r many \hat{T}_j could vanish. Let $T_i := \hat{T}_i \pmod{\mathcal{I}}$. Note that by [Observation 1](#) and the remark followed by it, each T_i must be of the form $T_i = \hat{T}_i \pmod{\mathcal{I}} = \epsilon^{-a_i} \cdot \ell_{i,1} \cdots \ell_{i,s}$, where $\ell_{i,r} \in \mathbb{F}[[\epsilon]][\mathbf{x}]$, and further $\ell_{i,r}|_{\epsilon=0}$ is a nonzero non-homogeneous linear polynomial. As mentioned earlier, for the border complexity purposes, it suffices to work with finite precisions, i.e. over $\mathbb{F}[\epsilon^{\pm 1}]$. Therefore, using [Claim 4](#), one can easily conclude that there are linear forms ℓ_1, \dots, ℓ_n with $\text{rank}(\ell_1, \dots, \ell_n) = n - t$, for some $t \leq k - m$, such that over $\mathbb{F}[\epsilon^{\pm 1}]$, the following equation holds:

$$f(\ell_1, \dots, \ell_n) + \epsilon \cdot H(\mathbf{x}, \epsilon) = T_1 + \dots + T_m,$$

This finishes the proof. □

As a warmup application, we prove a ‘folklore’ impossibility result in *homogeneous* models.

Corollary 16 (Impossibility result). *For $1 \leq k < d$, the polynomial $P_{k+1,d}$ is uncomputable by homogeneous $\overline{\Sigma^{[k]}}\Pi\Sigma$ circuits.*

PROOF. The proof goes through a contradiction. Suppose there is a homogeneous $\overline{\Sigma^{[k]}}\Pi\Sigma$ circuit computing $P_{k+1,d}$, i.e. $P_{k+1,d} + \epsilon \cdot S(\mathbf{x}, \epsilon) = \hat{T}_1 + \dots + \hat{T}_k$. Suppose $T_i = \epsilon^{-a_i} \cdot \ell_{i,1} \cdots \ell_{i,s}$, where $a_i \geq 0$, and $\ell_{i,j} \in \mathbb{F}[\epsilon][\mathbf{x}]$ are linear forms (in \mathbf{x}). Without loss of generality, we can pick one linear form, say $\ell_{i,1}$ from each product gate, and consider the given equation modulo the ideal $\mathcal{I} := \langle \ell_{1,1}, \dots, \ell_{k,1} \rangle$. This would make the RHS vanish, and from the above reduction ([Lemma 15](#)) the ϵ -free term on the LHS would be $P_{k+1,d}(\ell_1, \dots, \ell_n)$, where $n := (k+1)d$, and $\{\ell_1, \dots, \ell_n\}$, are the linear forms with $r := \text{rank}(\ell_1, \dots, \ell_n) = n - t \geq n - k$. This in particular implies that $P_{k+1,d}(\ell_1, \dots, \ell_n)$ must be a zero polynomial, and we show that this cannot happen.

To argue $P_{k+1,d}(\ell_1, \dots, \ell_n) \neq 0$, note that by the pigeonhole principle, there exists $i \in \{0, \dots, k\}$, such that the set of linear forms $\{\ell_{id+1}, \dots, \ell_{(i+1)d}\}$, must be of full rank. Without loss of generality, $\text{rank}(\ell_1, \dots, \ell_d) = d$. Furthermore, from each set of linear forms $\{\ell_{id+1}, \dots, \ell_{(i+1)d}\}$, for each $i \in [k]$, one can pick at least one from each set such that all the picked-up linear forms along with ℓ_1, \dots, ℓ_d are linearly independent. Define an invertible linear transformation that sends these linear forms to different variables, and further, for notational ease, suppose: $\phi : \ell_i \mapsto x_i$, for $i \in [d]$. Note that, after applying the transformation, the monomial $x_1 \cdots x_d$ appears in the new polynomial $\phi(P_{k+1,d}(\ell_1, \dots, \ell_n))$ which is produced from the first product. We claim that this monomial can never get canceled which would imply the desired nonzeroness. To show this, observe that in each product $\ell_{id+1} \cdots \ell_{(i+1)d}$, for $i \geq 1$, at least one linear form has been mapped to a new variable (except x_1, \dots, x_d), and hence all the monomials produced from each such product must be different from $x_1 \cdots x_d$. This finishes the proof. □

5.1 Reduced problem

In the next section, we will mainly prove the following theorem.

THEOREM 17 (REDUCED LOWER BOUND THEOREM). *For constants $m < k$, let*

$$P_{k+1,d}(\ell_1, \dots, \ell_n) + \epsilon \cdot H(\mathbf{x}, \epsilon) = T_1 + \dots + T_m$$

where –

- 1301 (1) each ℓ_i is a linear form such that $\text{rank}(\ell_1, \dots, \ell_n) \geq n - k$,
 1302 (2) (all-non-hom.). Each $T_i \in \mathbb{F}[\epsilon^{\pm 1}][\mathbf{x}]$ is a product of linear polynomials such that $(T_i/\epsilon^{v_i})|_{x=\epsilon=0} \in \mathbb{F} - \{0\}$
 1303 (where $v_i := \text{val}_\epsilon(T_i)$).
 1304

1305 Then, $\sum_{i=1}^m \text{size}(T_i) \geq 2^{\Omega(d)}$.
 1306

1307 Using Lemma 15, one can easily deduce that Theorem 2 is a simple corollary of Theorem 17., Therefore, for
 1308 the rest of the paper, all our efforts will be put to prove that Theorem 17 is true.
 1309

1310 The proof of the above theorem is in two parts:

- 1311 (1) Showing that $P_{k+1,d}(\ell_1, \dots, \ell_n)$ can be expressed as a ratio of two small border depth-3 diagonal
 1312 circuits (Lemma 18),
 1313 (2) $P_{k+1,d}(\ell_1, \dots, \ell_n)$ cannot be written as a ratio of two small border depth-3 diagonal circuits via analyz-
 1314 ing cone-size of the leading monomials Lemma 23
 1315

1316 Remark. Interestingly, exponential lower bounds can be shown for polynomials such as $\text{IMM}_{k+1,d}$, det_d , perm_d ,
 1317 since one can show that Lemma 15, Corollary 16, as well as the technical proof of Theorem 17 continue to hold
 1318 for these polynomials. But we will only focus on $P_{k+1,d}$.
 1319
 1320

1321 6 PROVING THEOREM 17: THE MAIN LOWER BOUND

1322 Assume that $P_{k+1,d}(\ell_1, \dots, \ell_n)$ can be computed by a structured $\overline{\Sigma^{[m]}\Pi\Sigma}$ circuit, as in the statement of Theo-
 1323 rem 17, of size s . We aim to prove an exponential lower bound on s . The proof now proceeds in a different
 1324 direction: we induct over a more general circuit class. For $P_{k+1,d}$, we set $n := (k+1)d$, to be the number of
 1325 variables x , and we use this notation throughout. We recall the bloated model, which we defined already in
 1326 Section 4.3.
 1327
 1328
 1329

1330 **Definition 2** (Bloated model, [DDS21]). We say that a circuit \mathcal{C} is in the class $\text{Gen}(m, s)$, over the fraction field
 1331 $R(\mathbf{x})$, with parameter m and size s , if it computes polynomials $f \in R(\mathbf{x})$, of the form $f = \sum_{i \in [k]} T_i$, where $T_i =$
 1332 $(U_i/V_i) \cdot (P_i/Q_i)$, with $U_i, V_i, P_i, Q_i \in R[\mathbf{x}]$ such that
 1333

- 1334 (1) $U_i, V_i \in \Pi\Sigma$; with $U_i \bmod \langle x_1, \dots, x_n \rangle$ and $V_i \bmod \langle x_1, \dots, x_n \rangle$ are invertible in R , and
 1335 (2) $P_i, Q_i \in \Sigma\wedge\Sigma$.
 1336

1337 Further, $\text{size}(\mathcal{C}) := \sum_{i \in [m]} \text{size}(T_i)$, and $\text{size}(T_i) := \text{size}(U_i) + \text{size}(V_i) + \text{size}(P_i) + \text{size}(Q_i)$.
 1338

1339 Our model for the induction will be $\overline{\text{Gen}(m, s)}$, the approximative closure of the above bloated model. The
 1340 lower bound proof (of Theorem 17) can now be divided into two parts:
 1341

- 1342 (1) Reducing Theorem 17 to proving lower bound on $\text{Gen}(1, s')$, for some parameter s' , in some appropriate
 1343 ring; see Lemma 18.
 1344 (2) Proving the corresponding lower bound for $\overline{\text{Gen}(1, s')}$, see Lemma 23.
 1345
 1346

1347 6.1 Main fanin reduction lemma and formulating the induction hypothesis

1348 In this subsection, we prove the first step, namely, reducing the top fan-in to 1. Now we use the DiDIL-technique
 1349 developed in [DDS21]. By hypothesis, each $T_i \bmod \langle x \rangle$ has a nonzero constant term (after extracting the
 1350 appropriate ϵ -power). So, for the existence of $1/T_i$, DiDIL technique *does not* require any additional shift: a
 1351

1353 mere scaling by a new variable z suffices. This scaling is used both for taking derivatives and for keeping track
 1354 of the x -degree.

1355 Before we begin, let us first recall the notation and make a slight modification for ease of use. We have

$$1356 \quad g_0 = f_0(\mathbf{x}) + \epsilon \cdot S_0 = \sum_{i=1}^m T_{i,0}. \quad (17)$$

1357 Here, $n := (k+1)d$, and $f_0 := P_{k+1}(\ell_1, \dots, \ell_n)$, and we rename H as S_0 (see the statement of [Theorem 17](#)).
 1358 Each $T_{i,0}$ is computable by a $\Pi\Sigma$ -circuit of size at most s over $\mathbb{F}[\epsilon^{\pm 1}]$ and has the following property: after
 1359 dividing by the highest power of ϵ dividing $T_{i,0}$, the result is a nonzero constant mod $\langle x_1, \dots, x_n \rangle$. Further, let
 1360 $d_0 := d + 1$.

1361 We now state the main reduction lemma, which reduces to fan-in 1 by expressing the polynomial as a ratio
 1362 of two depth-3 diagonal circuits of small size. The proof of the lemma appears at the end of this subsection
 1363 and assumes the induction (Part II), which is stated at the end of this page. (It might help the reader to first
 1364 see an easier case of this proof, as provided under [Lemma 12](#).)

1365 **Lemma 18 (Main Lemma).** *If f can be approximated by a structured $\Sigma^{[m]}\Pi\Sigma$ -circuit of size s (as in the statement of
 1366 [Theorem 17](#), or written in [Equation 17](#)), over $\mathbb{F}(\epsilon)$, then there exist two polynomials $F_1, F_2 \in \mathbb{F}[\mathbf{x}]$ such that each F_i can
 1367 be computed by a $\overline{\Sigma\wedge\Sigma}$ circuit of size $s^{O(7^m)}$, and $f = F_1/F_2$.*

1368 The proof of the lemma is lengthy, inductive, and somewhat technical. We therefore prove a sequence of
 1369 smaller claims along the way, which together imply [Lemma 18](#). The induction will be on the more general
 1370 bloated model over a ring (and not \mathbb{F}), which is our **Part II**, the main proof of the above lemma. We will discuss
 1371 the detailed proof of Part II in the next subsections (Subsection [6.2-6.3](#)). But before going into the induction,
 1372 the first part (**Part I**) is to apply a scaling map and see a circuit in $\overline{\Sigma^{[m]}\Pi\Sigma}$ as a circuit in $\overline{\text{Gen}(m, s)}$ as follows.

1373 **Part I: Reducing $\overline{\Sigma^{[m]}\Pi\Sigma}$ to $\overline{\text{Gen}(m, s)}$ via scaling.** To ensure the invertibility and facilitate differentiation, we
 1374 define a homomorphism (essentially a variable-scaling):

$$1375 \quad \Phi : \mathbb{F}[\epsilon^{\pm 1}][\mathbf{x}] \rightarrow \mathbb{F}[\epsilon^{\pm 1}][\mathbf{x}, z], \quad \text{such that } x_i \mapsto z \cdot x_i.$$

1376 Recall [Equation 17](#). After applying ϕ , we have

$$1377 \quad \Phi(g_0) = \Phi(f_0) + \epsilon \cdot \Phi(S_0) = \sum_{i=1}^m \Phi(T_{i,0}).$$

1378 One can do the following trivial observation.

1379 **Observation 2.** $\Phi(f_0) \in \overline{\text{Gen}(m, s)}$.

1380 To see the above, define $U_{i,0} := \Phi(T_{i,0})$ and $V_{i,0} := P_{i,0} := Q_{i,0} = 1$. Note that the $\Pi\Sigma$ circuits (i.e. factors of
 1381 $U_{i,0}$) are invertible mod z^{d_0} . Let $\mathcal{R}_0 := \mathbb{F}[z] / \langle z^{d_0} \rangle$, and think of the z -variables as ‘cost-free’. Clearly, $\Phi(f_0) \in$
 1382 $\overline{\text{Gen}(m, s)}$.

1383 Now we formulate the induction hypothesis as follows. After the hypothesis, we will argue why this induction
 1384 suffices to prove [Lemma 18](#).

1385 **Part II: Top fan-in reduction.** We will do induction on j , where $0 \leq j \leq m - 1$. Suppose, we are at the j -th step
 1386 where $j \geq 0$. We call a circuit in $\overline{\text{Gen}(m - j, s_j)}$, for some parameter s_j , *admissible* if it satisfies the following four
 1387 properties:

- 1405 (I) $\sum_{i \in [m-j]} T_{i,j} =: g_j$, over $\mathcal{R}(x, \epsilon)$, where $\mathcal{R} := \mathbb{F}(z) \subseteq \mathbb{F}((z))$. Moreover, g_j approximates f_j correctly,
 1406 i.e. $\lim_{\epsilon \rightarrow 0} g_j = f_j$.
 1407 Moreover, each $T_{i,j}$ is an element of $\mathcal{R}_j := \mathbb{F}((z))/z^{d_j}\mathbb{F}[[z]]$, for some d_j . Further, let $v_{i,j} := \text{val}_z(T_{i,j})$,
 1408 for $i \in [m-j]$.
 1409
 1410 (II) Here, $T_{i,j} =: (U_{i,j}/V_{i,j}) \cdot (P_{i,j}/Q_{i,j})$, where $U_{i,j}, V_{i,j} \in \Pi\Sigma$ and $P_{i,j}, Q_{i,j} \in \Sigma\wedge\Sigma$, each in $\mathbb{F}(\epsilon)[x, z]$.
 1411 Therefore, each $T_{i,j}$ can be thought as a function in $\mathbb{F}(x, z, \epsilon)$, of size at most s_j .
 1412 Assume that the syntactic degree of each denominator and numerator of $T_{i,j}$ is bounded by D_j ; this
 1413 will be required in the size analysis.
 1414 Moreover, if $\text{val}_z(f_j) = b_j$, then $b_j + D_j - 1 \geq b_{j+1} \geq b_j - D_j - 1$.
 1415 (III) Each $U_{i,j}|_{z=0} \in \mathbb{F}(\epsilon) \setminus \{0\}$ (similarly for $V_{i,j}$), is a *unit* in $\mathbb{F}(\epsilon)$.
 1416 (IV) $\text{coef}_{z^{b_j}}(\lim_{\epsilon \rightarrow 0} g_j) = f_0 \cdot (\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma})$, where $\overline{\Sigma\wedge\Sigma}$ are z -free. Equivalently, the coefficient of the
 1417 minimum z -power in f_j , is of the form $f_0 \cdot (\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma})$, where both $\overline{\Sigma\wedge\Sigma}$ circuits are nonzero.
 1418
 1419

1420 **Remark.** One could also work with the more general ring $\mathbb{F}[\epsilon^{\pm 1}](x)$, i.e. with finite precision (or power)
 1421 of ϵ and ϵ^{-1} . But for the ease of notation, we do not truncate ϵ -powers and work with the underlying ring of
 1422 $\mathbb{F}(x, \epsilon)$.
 1423

1424 Here is the main induction lemma.
 1425

1426 **Lemma 19** (Main Induction Lemma). *For $0 \leq j \leq m-1$, if a circuit C in $\overline{\text{Gen}(m-j, s_j)}$ is admissible, then there
 1427 exists another admissible circuit in $\overline{\text{Gen}(m-j-1, s_{j+1})}$, such that $s_{j+1} = s_j^7 d^{O(2^j)}$.*
 1428
 1429

1430 **Lemma 19 implies Lemma 18.** Once Lemma 19 is proved, setting $j = m-1$ yields Lemma 18. Indeed, we have
 1431 the following. Recall, we have shown that the minimum z -power in g_{m-1} gives $f(\ell_1, \dots, \ell_n) \cdot \overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma}$, where
 1432 $\Sigma\wedge\Sigma$ are of size $s_{m-1} = s^{O(7^m)}$. Moreover, since $g_{m-1} \in \overline{\text{Gen}(1, s^{O(7^m)})}$, over $\mathbb{F}(\epsilon)$, the limit of g_{m-1} itself is of the
 1433 form $\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma}$, because the other factors satisfy $\Pi\Sigma|_{x=\epsilon=0} \in \mathbb{F}^*$. Therefore, comparing the coefficients of the
 1434 minimum z -power both sides, we have eventually shown: $f_0 \cdot \overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma} = \overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma} \implies f_0 = \overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma}$.
 1435 Here $\overline{\Sigma\wedge\Sigma}$ circuits have size $s^{O(7^m)}$, because of Lemma 25.
 1436
 1437
 1438
 1439

1440 So from now on, the proof will focus only on proving the above by assuming the induction hypotheses.
 1441
 1442

1443 6.2 Proof of the main induction lemma

1444 In this subsection, we will see how to reduce the fanin by induction to 1 using dlog. The size analysis is deferred
 1445 to the next subsection (Subsection 6.3).
 1446
 1447

1448 **PROOF.** We will prove the reduction by reducing the top-fanin by 1 at each step. The first step is to **Divide**
 1449 and **Derive**, to get a circuit in $\overline{\text{Gen}(m-j-1, s_{j+1})}$, from a circuit in $\overline{\text{Gen}(m-j, s_j)}$ such that $s_{j+1} = s_j^7 d^{O(2^j)}$.
 1450
 1451

1452 *Deriving circuit in $\overline{\text{Gen}(m-j-1, s_{j+1})}$.* Let $T_{m-j,j} =: \epsilon^{a_{m-j,j}} \cdot \tilde{T}_{m-j,j}$, where $\tilde{T}_{m-j,j} =: (t_{m-j,j} + \epsilon \cdot \tilde{t}_{m-j,j})$ is not
 1453 divisible by ϵ . Let C_j be the circuit (over $\mathbb{F}(\epsilon)$) computing $g_j =: f_j + \epsilon \cdot S_j$. Divide it by $\tilde{T}_{m-j,j}$ and derive, to
 1454 get the circuit C_{j+1} computing g_{j+1} as follows:
 1455
 1456

$$\begin{aligned}
& f_j / \tilde{T}_{m-j,j} + \epsilon \cdot S_j / \tilde{T}_{m-j,j} = \epsilon^{a_{m-j,j}} + \sum_{i \in [m-j-1]} T_{i,j} / \tilde{T}_{m-j,j} \\
\implies \partial_z \left(f_j / \tilde{T}_{m-j,j} \right) + \epsilon \cdot \partial_z \left(S_j / \tilde{T}_{m-j,j} \right) &= \sum_{i \in [m-j-1]} \partial_z \left(T_{i,j} / \tilde{T}_{m-j,j} \right) \\
&= \sum_{i \in [m-j-1]} \left(T_{i,j} / \tilde{T}_{m-j,j} \right) \cdot \text{dlog} \left(T_{i,j} / \tilde{T}_{m-j,j} \right) \quad (18) \\
&=: g_{j+1}.
\end{aligned}$$

Definability. Let $\mathcal{R} := \mathbb{F}(z) \subseteq \mathbb{F}((z)) =: \widehat{\mathcal{R}}$. It helps to keep a standard fact in mind, as explicitly stated in [subsection 4.3](#) while proving $k = 3$ case, that each rational function has a *Laurent series*. Further, $\mathcal{R}_{j+1} := \mathbb{F}((z)) / z^{d_{j+1}} \mathbb{F}[[z]]$, where $d_{j+1} := d_j - v_{m-j,j} - 1$. For $i \in [m-j-1]$, Define

$$T_{i,j+1} := \left(T_{i,j} / \tilde{T}_{m-j,j} \right) \cdot \text{dlog} \left(T_{i,j} / \tilde{T}_{m-j,j} \right), \text{ and } f_{j+1} := \partial_z(f_j / t_{m-j,j}).$$

We will use this notation for the rest of the proof.

We will now prove properties (I) to (IV) in the definition of admissibility, and then analyze the size bound on s_{j+1} .

Claim 20 (Proof of Property I). *Circuit C_{j+1} that computes g_{j+1} approximates f_{j+1} correctly, i.e. $\lim_{\epsilon \rightarrow 0} g_{j+1} = f_{j+1}$, where g_{j+1} (respec. f_{j+1}) are well-defined in the ring $\mathcal{R}(x, \epsilon) \subset \widehat{\mathcal{R}}(x, \epsilon)$ (respec. $\mathcal{R}(x)$). Moreover, each $T_{i,j+1}$ is well-defined in the ring $\mathcal{R}_{j+1}(x, \epsilon)$, and $\text{val}_z(f_j) + D_j - 1 \geq \text{val}_z(f_{j+1}) \geq \text{val}_z(f_j) - D_j - 1$.*

PROOF. Remember, f_j and $T_{i,j}$ are elements in $\mathcal{R}(x, \epsilon)$. Further f_j is ϵ -free. Since we divide by $T_{m-j,j}$, it follows that

$$T_{i,j+1} \in \mathcal{R}(x, \epsilon), \forall i \in [m-j-1], \text{ and } f_{j+1} \in \mathcal{R}(x),$$

proving the second part of induction-hypothesis-(II).

Further, $\text{val}_\epsilon(\tilde{T}_{m-j,j}) = 0$. Therefore, using [Fact 2](#), we get

$$\lim_{\epsilon \rightarrow 0} g_{j+1} = \lim_{\epsilon \rightarrow 0} \left(\partial_z \left(f_j / \tilde{T}_{m-j,j} \right) + \epsilon \cdot \partial_z \left(S_j / \tilde{T}_{m-j,j} \right) \right) = \lim_{\epsilon \rightarrow 0} \partial_z \left(f_j / \tilde{T}_{m-j,j} \right) = \partial_z \left(f_j / t_{m-j,j} \right).$$

Clearly $f_j / t_{m-j,j} \in \mathcal{R}(x)$, and hence f_{j+1} exists in the same ring.

By the induction hypothesis, $T_{i,j} \in \mathcal{R}_j(x, \epsilon)$. Since we divide by $\tilde{T}_{m-j,j}$, whose z -valuation is $v_{m-j,j}$, we have

$$\text{val}_z \left(\frac{T_{i,j}}{\tilde{T}_{m-j,j}} \right) = v_{i,j} - v_{m-j,j},$$

and hence

$$\text{val}_z \left(\partial_z \left(\frac{T_{i,j}}{\tilde{T}_{m-j,j}} \right) \right) = v_{i,j} - v_{m-j,j} - 1.$$

Since the equation initially holds in $\mathcal{R}_j(x, \epsilon)$, it continues to hold after dividing by $T_{m-j,j}$ and applying ∂_z , and therefore Eqn. (18) holds $\text{trun } z^{d_j - v_{m-j,j} - 1}$. Equivalently, Eqn. (18) holds in $\mathcal{R}_{j+1}(x, \epsilon)$. This is precisely Induction Hypothesis (I), and it justifies that Eqn. (18) may indeed be truncated up to $z^{d_{j+1} - 1}$ (equivalently $\text{trun } z^{d_{j+1}}$).

Let $\text{val}_z(t_{m-j,j}) = d'_j$. Since $\text{val}_z(f_j) = b_j$, and we assumed the syntactic degree of the numerator and denominator being bounded by D_j , trivially $-D_j \leq d_j \leq D_j$, by [Fact 1](#), $b_j + D_j \geq \text{val}_z(f_j / t_{m-j,j}) = b_j - d'_j \geq b_j - D_j$, which implies $b_j + D_j - 1 \geq \text{val}_z(f_{j+1}) \geq b_j - D_j - 1$.

This finishes [Claim 20](#). □

1509 **Proving property III.** Note that $\text{dlog}(\Sigma\wedge\Sigma) \in \Sigma\wedge\Sigma/\Sigma\wedge\Sigma$, where each $\Sigma\wedge\Sigma$ circuits are of poly-size (Lemma 28).

1510 Since dlog distributes over products additively, it suffices to work with $\text{dlog}(\Pi\Sigma)$. We have already seen that
 1511 $\text{dlog}(\Pi\Sigma) \in \Sigma\wedge\Sigma$ and that, after truncation, it has polynomial size (see section 3); we will also reprove this in
 1512 the size analysis (see Subsection 6.3). We simplify

$$1514 \frac{T_{i,j}}{\tilde{T}_{m-j,j}} = \epsilon^{-a_{m-j,j}} \cdot \frac{U_{i,j}V_{m-j,j}}{V_{i,j}U_{m-j,j}} \cdot \frac{P_{i,j}Q_{m-j,j}}{Q_{i,j}P_{m-j,j}},$$

1517 and apply dlog to this expression. Let $U_{i,j+1} := U_{i,j} \cdot V_{m-j,j}$ and $V_{i,j+1} := V_{i,j} \cdot U_{m-j,j}$. The dlog computation
 1518 produces $(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$ -circuits, which are then multiplied with the P 's and Q 's, while the factor $\epsilon^{-a_{m-j,j}}$ is
 1519 carried along; see Claim 21 for details. In particular, $U_{i,j+1}|_{z=0}, V_{i,j+1}|_{z=0} \in \mathbb{F}(\epsilon) \setminus \{0\}$.

1521 By induction, the syntactic degree of each numerator and denominator of $T_{i,j}$ is bounded by D_j . Hence
 1522 $\text{val}_z(t_{m-j,j}) \leq D_j$, and the same bound propagates to the numerators and denominators at stage $j+1$. This
 1523 establishes the second part of induction-hypothesis (III).
 1524

1525 **Proving Property-(IV).** We want to show that f_j evolves in a predictable way and always computes a ‘multiple’
 1526 of f , namely

$$1527 \text{coef}_{z^{b_j}}(f_j) = f_0 \cdot (\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma}).^{10}$$

1529 To see this, we first give names to the corresponding $\Sigma\wedge\Sigma$ -circuits for future reference.

1531 In particular, by induction, assume that

$$1532 \text{coef}_{z^{b_j}}(f_j) = f_0 \cdot (E_{j,1}/E_{j,2}),$$

1534 where $E_{j,1}, E_{j,2} \in \overline{\Sigma\wedge\Sigma}$ and each is z -free. Since $\tilde{T}_{m-j,j} \in (\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$, its ϵ -free part $t_{m-j,j}$ must
 1535 be of the form

$$1537 t_{m-j,j} = \lim_{\epsilon \rightarrow 0} (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma).$$

1538 Here we crucially use Property (III): $\lim_{\epsilon \rightarrow 0} (\Pi\Sigma/\Pi\Sigma)$ is a nonzero constant in \mathbb{F} , since the $\Pi\Sigma|_{z=0}$ part of the
 1539 expression is a unit in $\mathbb{F}(\epsilon) \setminus \{0\}$.

1541 Recall that $\text{val}_z(t_{m-j,j}) = d'_j$. We denote

$$1542 \text{coef}_{z^{d'_j}}(t_{m-j,j}) = E_{j,3}/E_{j,4},$$

1544 where $E_{j,3}, E_{j,4}$ are computed by $\overline{\Sigma\wedge\Sigma}$ -circuits and are z -free. Here we use that $\overline{\Sigma\wedge\Sigma}$ is closed under interpolation
 1545 and substitution. We will prove in the size-analysis section (subsection 6.3) that all such circuits remain
 1546 polynomially bounded in size.

1548 By definition, $\text{val}_z(f_{j+1}) = b_{j+1}$, which implies $\text{val}_z(f_j/t_{m-j,j}) = b_{j+1} + 1$. Hence

$$1550 b_{j+1} + 1 = \text{val}_z(f_j) - \text{val}_z(t_{m-j,j}) = b_j - d'_j.$$

1552 Therefore, by the j -th induction hypothesis,

$$1553 \text{coef}_{z^{b_j-d'_j}}(f_j/t_{m-j,j}) = \text{coef}_{z^{b_j}}(f_j)/\text{coef}_{z^{d'_j}}(t_{m-j,j}) = f_0 \cdot (E_{j,1}/E_{j,2}) \cdot (E_{j,4}/E_{j,3}).$$

1558 ¹⁰Note that $f_j \in \mathcal{R}(x)$, hence b_j can be negative.

It is important to note that this works because we only look at the coefficient of the minimum z -power. Since the right-hand side is z -free and $b_{j+1} = b_j - d'_j - 1$, it follows that

$$\text{coef}_{z^{b_{j+1}}}(f_{j+1}) = \text{coef}_{z^{b_{j+1}}}(\partial_z(f_j/t_{m-j,j})) = f_0 \cdot (E_{j,1}/E_{j,2}) \cdot (E_{j,4}/E_{j,3}).$$

Since $(\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma})$ is closed under multiplication (Lemma 25), the right-hand side is of the form $f_0 \cdot (\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma})$. Using $f_{j+1} = \lim_{\epsilon \rightarrow 0} g_{j+1}$, we obtain

$$\text{coef}_{z^{b_{j+1}}}\left(\lim_{\epsilon \rightarrow 0} g_{j+1}\right) = \text{coef}_{z^{b_{j+1}}}(f_{j+1}) = f_0 \cdot (\overline{\Sigma\wedge\Sigma}/\overline{\Sigma\wedge\Sigma}).$$

This completes the inductive proof of the connection between g_j and f , establishing Property (IV). \square

6.3 The effect of fan-in reduction on the circuit size

Finally, we show the bound on s_{j+1} in terms of s_j that is stated in the main induction lemma. The analysis depends on dlog which is the crux of our reduction and the blowup in the size in each reduction. This is also the part where we prove Property II. We can assume that at the j -th step, $\text{size}(T_{i,j}) \leq s_j$ and by assumption $s_0 \leq s$.

Claim 21 (Size blowup from DiDIL, Main Claim). $T_{1,m-1} \in (\Pi\Sigma/\Pi\Sigma)(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$ over $\mathcal{R}_{m-1}(x, \epsilon)$ of size $s^{O(7^m)}$. It is computed as an element in $\mathbb{F}(\epsilon, x, z)$, with syntactic degree (in x, z) $d^{O(2^m)}$.

Towards the above claim, we prove the following claim (stated below). The main idea of using dlog and expand it as a power-series is the same, which eventually shows that $\text{dlog}(\Pi\Sigma) \in \Sigma\wedge\Sigma$ with a controlled blowup.

Claim 22 (Size analysis of $\text{dlog}(T_{i,j}/\tilde{T}_{m-j,j})$, Subclaim). $\text{dlog}(T_{i,j}/\tilde{T}_{m-j,j})$ can be written as a ratio of two $\Sigma\wedge\Sigma$ circuits, each of size $O(s_j^5 D_j^{16} d)$.

PROOF OF CLAIM 22. We have already discussed the basic idea of dlog on a linear polynomial (under the scaling map Φ in section 3). To elaborate (which helps us for the size analysis), ℓ of the form $\ell = A - zB$, where $A \in \mathbb{F}(\epsilon) \setminus \{0\}$ and $B \in \mathbb{F}(\epsilon)[x]$. Let us first see what happens when going from $j = 0$ to 1. Using the power series expansion, we have the following, over $\mathcal{R}_1(x, \epsilon)$:

$$\text{dlog}(\ell) = -\frac{\partial_z(z \cdot B)}{A(1 - z \cdot B/A)} = -\frac{B}{A} \cdot \sum_{j=0}^{d_1-1} \left(\frac{z \cdot B}{A}\right)^j. \quad (19)$$

Note, (B/A) and $(-z \cdot B/A)^j$ have trivial powering circuits ($\wedge\Sigma$ over \mathcal{R}_1), each of size $O(dn)$. By Lemma 25, we get the final $\Sigma\wedge\Sigma$ circuit for $\text{dlog}(\Pi\Sigma)$ of size $O(d^2 \cdot s)$. We use the fact that $d_1 < d_0 = d + 1$. Here the syntactic degree blowsup to $O(d)$. This settles $j = 0$ case.

For $j > 0$, the above equation holds over $\mathcal{R}_j(x)$. However, the degree could be D_j (possibly $> d_j$) of the corresponding ℓ (nonlinear), and after exponentiation further increase to $d_j \cdot D_j$. This is exactly why we need to keep track of D_j , the syntactic degree as mentioned in induction-hypotheses-(II). Also, each $v_{i,j}$ is bounded by D_j , and therefore $d_{j+1} \leq d_j + D_j - 1$. We will crucially use this.

Since, $\Sigma\wedge\Sigma$ is closed under differentiation (Lemma 28), effect of dlog on $\Sigma\wedge\Sigma$ is straightforward. Using Lemma 28, we obtain a $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ circuit, computing $\text{dlog}(P_{i,j})$ (similarly $\text{dlog}(Q_{i,j})$) of size $O(D_j^2 \cdot s_j)$. Also,

1613 by the distributive property of dlog on a product, and the action of dlog on a linear polynomial as discussed
 1614 above, $\text{dlog}(U_{i,j} \cdot V_{m-j,j}) \in \Sigma \wedge \Sigma$, which could be computed using the above Equation. Thus,
 1615

$$1616 \quad \text{dlog}(T_{i,j}/\tilde{T}_{m-j,j}) \in \text{dlog}(\Pi\Sigma/\Pi\Sigma) \pm \Sigma^{[4]}\text{dlog}(\Sigma \wedge \Sigma)$$

$$1617 \quad \subseteq \Sigma \wedge \Sigma + \Sigma^{[4]}(\Sigma \wedge \Sigma/\Sigma \wedge \Sigma) = (\Sigma \wedge \Sigma/\Sigma \wedge \Sigma) .$$

1619 Here, $\Sigma^{[4]}(\cdot)$ means sum of 4-many expressions of the form (\cdot) . The first containment is by linearization. We
 1620 can express $\text{dlog}(\Pi\Sigma/\Pi\Sigma)$ as a single $\Sigma \wedge \Sigma$ -expression since from the above discussion, $\text{dlog}(\ell) \in \Sigma \wedge \Sigma$, for a
 1621 linear polynomial ℓ . Similarly, 4-many $\text{dlog}(\Sigma \wedge \Sigma)$ expressions give 4-many $(\Sigma \wedge \Sigma/\Sigma \wedge \Sigma)$ expressions.
 1622

1623 The $\Sigma \wedge \Sigma$ expression, obtained from $\text{dlog}(\Pi\Sigma/\Pi\Sigma)$ is of size $O(D_j^2 d_j s_j)$. Next, there are 4-many $\Sigma \wedge \Sigma/\Sigma \wedge \Sigma$
 1624 expressions of size $O(D_j^2 s_j)$ as there are 4-many P 's and Q 's. Additionally, the syntactic degree of each denomi-
 1625 nator and numerator of $\Sigma \wedge \Sigma/\Sigma \wedge \Sigma$ grows up to $O(D_j)$. Finally, we club $\Sigma \wedge \Sigma/\Sigma \wedge \Sigma$ expressions (4 of them) to
 1626 express it as a single $\Sigma \wedge \Sigma/\Sigma \wedge \Sigma$ expression using [Lemma 28](#), with size blowup of $O(D_j^{12} s_j^4)$. Finally, add the
 1627 single $\Sigma \wedge \Sigma$ expression of size $O(D_j^3 s_j)$, and degree $O(d_j D_j)$, to get $O(s_j^5 D_j^{16} d_j)$ size representation. \square
 1628
 1629
 1630

1631 **PROOF OF CLAIM 21.** With the dlog expression, we need to multiply with $T_{i,j}/\tilde{T}_{m-j,j}$ which is of the form
 1632 $(\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma \wedge \Sigma/\Sigma \wedge \Sigma)$, where each $\Sigma \wedge \Sigma$ is basically product of two $\Sigma \wedge \Sigma$ expressions of size s_j and syntactic
 1633 degree D_j and clubbed together, owing a blowup of $O(D_j s_j^2)$. Hence, multiplying this $(\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma \wedge \Sigma/\Sigma \wedge \Sigma)$ -
 1634 expression with the $\Sigma \wedge \Sigma/\Sigma \wedge \Sigma$ expression obtained from dlog -computation, gives a size blowup of $s_{j+1} :=$
 1635 $s_j^7 D_j^{O(1)} d_j$.
 1636
 1637

1638 As mentioned before, the main blowup of syntactic degree in the dlog computation could be $O(d_j D_j)$ and
 1639 clearing expressions and multiplying the without- dlog expression increases the syntactic degree only by a
 1640 constant multiple. Therefore, $D_{j+1} := O(d_j D_j) \implies D_j = d^{O(2i)}$. Hence, $s_{j+1} = s_j^7 \cdot d^{O(2i)} \implies s_j \leq (sd)^{O(7i)}$.
 1641 In particular, $s_{m-1} \leq s^{O(7m)}$; here we used that $d \leq s$. It is also easy to see that $b_j \geq -d^{O(2i)}$. This calculation
 1642 quantitatively establishes induction-hypothesis-(2). This finishes [Claim 21](#). \square
 1643
 1644

1645 6.4 Proving lower bound for ratio of depth-3 diagonal circuits

1646 Now we show that for constant $1 \leq m \leq k$, s in [Lemma 18](#) must be exponentially large.

1648 **Lemma 23** (Lower bound for ratio of depth-3 diagonals). *If $f_0 = F_1/F_2$, such that*

- 1649 (1) *each F_i is computed by $\overline{\Sigma \wedge \Sigma}$ circuit, of size at most $s^{O(7^k)}$, and*
 1650 (2) *$\text{rank}(\ell_1, \dots, \ell_n) \geq n - k$,*

1651 *then $s \geq 2^{\Omega(d/7^k)}$.*

1654 **PROOF.** The proof is based on the cone-size measure. Note that, $f_0 = F_1/F_2 \implies f_0 \cdot F_2 = F_1$. Let G_i be
 1655 $\Sigma \wedge \Sigma$ circuits over $\mathbb{F}(\epsilon)$ such that $G_i|_{\epsilon=0} = F_i$. We will work with the partial derivatives spaces (defined below):
 1656 Denote

$$1657 \quad V_{\epsilon,i} := \left\langle \frac{\partial G_i}{\partial \mathbf{x}^e} \mid e < \infty \right\rangle_{\mathbb{F}(\epsilon)}, \quad \text{and} \quad V_i := \left\langle \frac{\partial F_i}{\partial \mathbf{x}^e} \mid e < \infty \right\rangle_{\mathbb{F}} .$$

1660 Since, $\text{size}(G_i) \leq s^{O(7^k)}$, the partial derivative space of G_i , over $\mathbb{F}(\epsilon)$, is also bounded by $s^{O(7^k)}$, i.e. $\dim(V_{\epsilon,i}) \leq$
 1661 $s^{O(7^k)}$ (see [[CKW11](#), Lemma 10.2]). Consider the partial-derivative *matrix* $M_{\epsilon,i}$, where we index the rows by
 1662 $\partial_{\mathbf{x}^e}$, while columns are indexed by monomials supporting G_i ; and each row expresses the operator-values
 1663

1665 $\partial_{x^e} G_i$. We have, $q_i := \dim(V_{\epsilon,i}) \leq s^{O(7^k)}$ (because of G_i). So, any $(q_i + 1)$ -many polynomials $\frac{\partial G_i}{\partial x^e}$ are \mathbb{F} -linearly
 1666 dependent. In other words, determinant of any $(q_i + 1) \times (q_i + 1)$ minor of $M_{\epsilon,i}$ is 0. Note that $\lim_{\epsilon \rightarrow 0} M_{\epsilon,i} = M_i$,
 1667 the corresponding partial-derivative matrix for F_i .
 1668

1669 Since, \det is a *continuous* function, the zeroness of the determinant of any $(q_i + 1) \times (q_i + 1)$ minor of $M_{\epsilon,i}$
 1670 translates to the corresponding $(q_i + 1) \times (q_i + 1)$ submatrix of M_i as well. In particular,
 1671

$$1672 \dim(V_i) \leq q_i \leq s^{O(7^k)}. \quad (20)$$

1673 From this, it follows that leading monomial in F_i (denoted $\text{LM}(F_i)$) has cone-size at most $s^{O(7^k)}$; see [Lemma 30](#).
 1674

1675 Since, $\text{cone-size}(\text{LM}(f_0 \cdot F_2)) \geq \text{cone-size}(\text{LM}(f_0))$, it suffices to show that leading monomial of f_0 has
 1676 cone-size 2^d .
 1677

1678 **Exploiting ‘large rank’.** As argued in [Corollary 16](#), $f_0 = P_{k+1,d}(\ell_1, \dots, \ell_n)$ is nonzero, and satisfies the property
 1679 that after suitable isomorphism (which does not affect our proof), there is a pure multilinear monomial,
 1680 i.e. product of distinct variables, that survives. This implies that with respect to a suitable monomial ordering,
 1681 the leading term has cone-size $= 2^d$. Implying: $\text{size}(F_1) \geq \text{cone-size}(\text{LM}(F_1)) = \text{cone-size}(\text{LM}(f_0 \cdot F_2)) \geq 2^d$
 1682 (also, see the remark below). Consequently, $s^{7^k} \geq 2^{\Omega(d)}$. This finishes the lower bound proof.
 1683 □
 1684
 1685

1686 **Remark.** In the above, we only work with $\overline{\Sigma \wedge \Sigma}$. Since, $\overline{\Sigma \wedge \Sigma}$ has *low* partial derivative space, the cone-size-based
 1687 proof goes through ([Lemma 30](#)). This is exactly why our proofs naively cannot give interesting lower bounds
 1688 for $\overline{\Sigma^{[k]} \Pi \Sigma \wedge}$ and $\overline{\Sigma^{[k]} \Pi \Sigma \Pi}$, since the cone-size-based argument no longer works.
 1689
 1690

1691 6.5 Tying the pieces together: Proof of [Theorem 2](#)

1692 *Proof of [Theorem 2](#).* We have shown and proved all the necessary steps for [Theorem 2](#). To summarize, we start
 1693 with $P_{k+1,d}$ and assume that it can be computed by a $\overline{\Sigma^{[k]} \Pi \Sigma}$ -circuit of size s . We now reduce top-fanin in two
 1694 totally different ways.
 1695

- 1696 (1) (*Setting linear forms zero*). If needed, use the reduction [Lemma 15](#) to reduce to the setting of [Theorem 17](#).
 1697 If we end up with $m = 0$, then a proof similar to [Corollary 16](#) shows: $P_{k+1,d}(\ell_1, \dots, \ell_n) \neq 0$ for
 1698 $\text{rank}(\ell_1, \dots, \ell_n) \geq n - k$. Thus $m = 0$ is an impossibility.
 1699
- 1700 (2) (*DiDIL process*). If $m \geq 1$, then it suffices to solve [Theorem 17](#) as mentioned before. Use the reduction
 1701 from [subsection 6.1](#) which reduces the top-fanin m to 1 (at the cost of moving to the bloated model).
 1702 Finally, use the lower bound for fanin-1 ([Lemma 23](#)) to conclude that $s \geq 2^{\Omega(d)}$ (for constant k). This
 1703 finishes [Theorem 2](#).
 1704 □
 1705
 1706

1707 7 CONCLUSION

1708 In this work, we show a strong top-fanin-hierarchy theorem for depth-3 class in the border setting. The methods
 1709 used here, open a wide avenue of plausible questions, some of which may not be very hard to answer. We list a
 1710 few of them below.
 1711

- 1712 (1) Can we show exponential lower bound for $\overline{\Sigma^{[o(n)]} \Pi \Sigma}$ -circuits? The current method gives subexponential
 1713 lower bound only as long as $k = o(\log n)$.
 1714
- 1715 (2) Can we show exponential lower bound for $\overline{\Sigma^{[k]} \Pi \Sigma \wedge}$ -circuits (i.e. rather special depth-4)?
 1716

- 1717 (3) Can we extend the hierarchy theorem to bounded (top & bottom fanin) depth-4 circuits? i.e., for a fixed
 1718 constant δ , is $\Sigma^{[1]}\Pi\Sigma\Pi^{[\delta]} \subseteq \Sigma^{[2]}\Pi\Sigma\Pi^{[\delta]} \subseteq \Sigma^{[3]}\Pi\Sigma\Pi^{[\delta]} \dots$, where the respective gaps are exponential?
 1719 Clearly, $\delta = 1$ holds, from this work.
 1720

1721 *Acknowledgement.* P.D. is currently supported by the project "Foundation of Lattice-based Cryptography",
 1722 funded by NUS-NCS Joint Laboratory for Cyber Security, Singapore. N.S. thanks the funding support from
 1723 DST-SERB (CRG/2020/000045) and N. Rama Rao Chair. The majority of the work was done when the first
 1724 author was a PhD student at CMI (& IIT Kanpur). We thank Prateek Dwivedi and Amit Sinhababu for helpful
 1725 comments on the draft.
 1726
 1727
 1728

1729 REFERENCES

- 1730 [AF22] Robert Andrews and Michael A. Forbes. *Ideals, Determinants, and Straightening: Proving and Using Lower Bounds for*
 1731 *Polynomial Ideals*. *The 54th ACM Symposium on Theory of Computing (STOC 2022)*, 2022. 7
- 1732 [AGS19] Manindra Agrawal, Sumanta Ghosh, and Nitin Saxena. Bootstrapping variables in algebraic circuits. *Proceedings of the National*
 1733 *Academy of Sciences*, 116(17):8107–8118, 2019. 8
- 1734 [ASSS16] Manindra Agrawal, Chandan Saha, Ramprasad Satharishi, and Nitin Saxena. *Jacobian hits circuits: Hitting sets, lower bounds*
 1735 *for depth-D occur-k formulas and depth-3 transcendence degree-k circuits*. *SIAM Journal on Computing*, 45(4):1533–1562, 2016. 9
- 1736 [AW16] Eric Allender and Fengming Wang. *On the power of algebraic branching programs of width two*. *Computational Complexity*,
 1737 25(1):217–253, 2016. 10
- 1738 [AZGL⁺18] Zeyuan Allen-Zhu, Ankit Garg, Yuanzhi Li, Rafael Oliveira, and Avi Wigderson. *Operator scaling via geodesically convex*
 1739 *optimization, invariant theory and polynomial identity testing*. In *Proceedings of the 50th Annual ACM SIGACT Symposium on*
 1740 *Theory of Computing*, pages 172–181, 2018. 4
- 1741 [BCRL79] Dario Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti. $O(n^{2.7799})$ *Complexity for $n \times n$ Approximate Matrix*
 1742 *Multiplication*. *Inf. Process. Lett.*, 8(5):234–235, 1979. 4
- 1743 [BCS13] Peter Bürgisser, Michael Clausen, and Mohammad A Shokrollahi. *Algebraic complexity theory*, volume 315. Springer Science &
 1744 *Business Media*, 2013. 4
- 1745 [BFG⁺19] Peter Bürgisser, Cole Franks, Ankit Garg, Rafael Oliveira, Michael Walter, and Avi Wigderson. *Towards a theory of non-*
 1746 *commutative optimization: Geodesic 1st and 2nd order methods for moment maps and polytopes*. In *2019 IEEE 60th Annual*
 1747 *Symposium on Foundations of Computer Science (FOCS)*, pages 845–861. IEEE, 2019. 4
- 1748 [BGO⁺18] Peter Bürgisser, Ankit Garg, Rafael Oliveira, Michael Walter, and Avi Wigderson. *Alternating minimization, scaling algorithms,*
 1749 *and the null-cone problem from invariant theory*. *Proceedings of Innovations in Theoretical Computer Science (ITCS 2018)*, 2018. 4
- 1750 [BI13] Peter Bürgisser and Christian Ikenmeyer. *Explicit lower bounds via geometric complexity theory*. In *Proceedings of the 45th*
 1751 *Annual ACM symposium on Theory of Computing*, pages 141–150, 2013. 4
- 1752 [BIL⁺21] Markus Bläser, Christian Ikenmeyer, Vladimir Lysikov, Anurag Pandey, and Frank-Olaf Schreyer. *On the Orbit Closure Con-*
 1753 *tainment Problem and Slice Rank of Tensors*. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete*
 1754 *Algorithms, SODA 2021*, pages 2565–2584. SIAM, 2021. 4
- 1755 [Bin80] Dario Bini. *Relations between exact and approximate bilinear algorithms. Applications*. *Calcolo*, 17(1):87–97, 1980. 4, 5
- 1756 [BIP19] Peter Bürgisser, Christian Ikenmeyer, and Greta Panova. *No occurrence obstructions in geometric complexity theory*. *Journal of*
 1757 *the American Mathematical Society*, 32(1):163–193, 2019. 7
- 1758 [BIZ18] Karl Bringmann, Christian Ikenmeyer, and Jeroen Zuiddam. *On Algebraic Branching Programs of Small Width*. *J. ACM*,
 1759 65(5):32:1–32:29, 2018. 9, 11, 14
- 1760 [BLMW11] Peter Bürgisser, J. M. Landsberg, Laurent Manivel, and Jerzy Weyman. *An Overview of Mathematical Issues Arising in the*
 1761 *Geometric Complexity Theory Approach to VP \neq VNP*. *SIAM J. Comput.*, 40(4):1179–1209, 2011. 4
- 1762 [BOC92] Michael Ben-Or and Richard Cleve. *Computing algebraic formulas using a constant number of registers*. *SIAM Journal on*
 1763 *Computing*, 21(1):54–58, 1992. 10
- 1764 [BS83] Walter Baur and Volker Strassen. *The complexity of partial derivatives*. *Theoretical computer science*, 22(3):317–330, 1983. 6
- 1765 [Bür00] Peter Bürgisser. *Completeness and Reduction in Algebraic Complexity Theory*. Algorithms and Computation in Mathematics. Springer,
 1766 2000. Online version: <http://math-www.uni-paderborn.de/agpb/work/habil.ps>. 4
- 1767 [Bür04] Peter Bürgisser. *The Complexity of Factors of Multivariate Polynomials*. *Found. Comput. Math.*, 4(4):369–396, 2004. 4, 5
- 1768 [Bür20] Peter Bürgisser. *Correction To: The Complexity of Factors of Multivariate Polynomials*. *Found. Comput. Math.*, 20(6):1667–1668,
 2020. 4
- [Bür24] Peter Bürgisser. *Completeness classes in algebraic complexity theory*. *arXiv preprint arXiv:2406.06217*, 2024. 7

- 1769 [CGJ⁺18] Mahdi Cheraghchi, Elena Grigorescu, Brendan Juba, Karl Wimmer, and Ning Xie. [AC⁰ ◦ MOD₂ lower bounds for the Boolean](#)
1770 [Inner Product](#). *Journal of Computer and System Sciences*, 97:45–59, 2018. 3
- 1771 [CHI⁺18] L. Chiantini, J. D. Hauenstein, C. Ikenmeyer, J. M. Landsberg, and G. Ottaviani. Polynomials and the exponent of matrix
1772 multiplication. *Bull. Lond. Math. Soc.*, 50(3):369–389, 2018. 5
- 1773 [CKR⁺20] Prerona Chatterjee, Mrinal Kumar, C Ramya, Ramprasad Saptharishi, and Anamay Tengse. [On the Existence of Algebraically](#)
1774 [Natural Proofs](#). 61st Annual IEEE Symposium on the Foundations of Computer Science (FOCS 2020), 2020. 4
- 1775 [CKW11] Xi Chen, Neeraj Kayal, and Avi Wigderson. [Partial derivatives in arithmetic complexity and beyond](#). Now Publishers Inc, 2011. 17, 32
- 1776 [CW90] Don Coppersmith and Shmuel Winograd. [Matrix Multiplication via Arithmetic Progressions](#). *J. Symb. Comput.*, 9(3):251–280,
1990. 4
- 1777 [DDS21] Pranjal Dutta, Prateek Dwivedi, and Nitin Saxena. [Demystifying the border of depth-3 algebraic circuits](#). 62nd Annual IEEE
1778 [Symposium on the Foundations of Computer Science \(FOCS 2021\)](#), 2021. 2, 3, 4, 5, 6, 8, 23, 24, 26, 37, 38
- 1779 [DGI⁺22] Pranjal Dutta, Fulvio Gesmundo, Christian Ikenmeyer, Gorav Jindal, and Vladimir Lysikov. De-bordering and geometric
1780 complexity theory for waring rank and related models. *arXiv preprint arXiv:2211.07055*, 2022. 7
- 1781 [DGI⁺24] Pranjal Dutta, Fulvio Gesmundo, Christian Ikenmeyer, Gorav Jindal, and Vladimir Lysikov. Homogeneous algebraic complexity
1782 theory and algebraic formulas. In Venkatesan Guruswami, editor, *15th Innovations in Theoretical Computer Science Conference, ITCS*
1783 *2024, January 30 to February 2, 2024, Berkeley, CA, USA*, volume 287 of *LIPICs*, pages 43:1–43:23. Schloss Dagstuhl - Leibniz-Zentrum
1784 für Informatik, 2024. 7
- 1785 [DIP20] J. Dörfler, C. Ikenmeyer, and G. Panova. On geometric complexity theory: Multiplicity obstructions are stronger than occurrence
1786 obstructions. *SIAM Journal on Applied Algebra and Geometry*, 4(2):354–376, 2020. 7
- 1787 [DL25] Pranjal Dutta and Vladimir Lysikov. Recent advances in debordering methods. *arXiv preprint arXiv:2510.13049*, 2025. 5
- 1788 [DSS18] Pranjal Dutta, Nitin Saxena, and Amit Sinhababu. [Discovering the roots: uniform closure results for algebraic classes under](#)
1789 [factoring](#). In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *STOC 2018*, pages 1152–1165. ACM, 2018. 11
- 1790 [For16] Michael Forbes. [Some concrete questions on the border complexity of polynomials](#). Presentation given at the Workshop on
1791 Algebraic Complexity Theory WACT 2016 in Tel Aviv, 2016. 38
- 1792 [FS13] Michael A Forbes and Amir Shpilka. [Explicit noether normalization for simultaneous conjugation via polynomial identity](#)
1793 [testing](#). In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 527–542. Springer, 2013.
1794 4
- 1795 [GGOW16] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. [A deterministic polynomial time algorithm for non-commutative](#)
1796 [rational identity testing](#). In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 109–117. IEEE,
1797 2016. 4
- 1798 [Gho19] Sumanta Ghosh. [Low Variate Polynomials: Hitting Set and Bootstrapping](#). PhD thesis, PhD thesis. Indian Institute of Technology
1799 Kanpur, 2019. 38
- 1800 [GK98] Dima Grigoriev and Marek Karpinski. [An Exponential Lower Bound for Depth 3 Arithmetic Circuits](#). In *STOC 1998*, pages
1801 577–582, 1998. 9
- 1802 [GKKS16] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. [Arithmetic Circuits: A Chasm at Depth 3](#). *SIAM J.*
1803 *Comput.*, 45(3):1064–1079, 2016. FOCS 2013. 9
- 1804 [GKS16] Rohit Gurjar, Arpita Korwar, and Nitin Saxena. [Identity Testing for Constant-Width, and Commutative, Read-Once Oblivious](#)
1805 [ABPs](#). In Ran Raz, editor, *31st Conference on Computational Complexity, CCC 2016*, volume 50 of *LIPICs*, pages 29:1–29:16. Schloss
1806 Dagstuhl - Leibniz-Zentrum für Informatik, 2016. 38
- 1807 [GKSS17] Joshua A Grochow, Mrinal Kumar, Michael Saks, and Shubhangi Saraf. [Towards an algebraic natural proofs barrier via polynomial](#)
1808 [identity testing](#). *arXiv preprint arXiv:1701.01717*, 2017. 4
- 1809 [GR00] Dima Grigoriev and Alexander A. Razborov. [Exponential Lower Bounds for Depth 3 Arithmetic Circuits in Algebras of Functions](#)
1810 [over Finite Fields](#). *Appl. Algebra Eng. Commun. Comput.*, 10(6):465–487, 2000. FOCS 1998. 9
- 1811 [Gro12] Joshua Abraham Grochow. [Symmetry and equivalence relations in classical and geometric complexity theory](#). The University of Chicago,
1812 2012. 7
- 1813 [Gro15] Joshua A Grochow. [Unifying known lower bounds via geometric complexity theory](#). *Computational Complexity*, 24(2):393–475,
1814 2015. 4, 7
- 1815 [HS65] Juris Hartmanis and Richard E Stearns. [On the computational complexity of algorithms](#). *Transactions of the American Mathematical*
1816 *Society*, 117:285–306, 1965. 2
- 1817 [HS80] Joos Heintz and Claus-Peter Schnorr. Testing polynomials which are easy to compute (extended abstract). In *STOC 1980*, pages
1818 262–272, 1980. 8
- 1819 [IK20] C. Ikenmeyer and U. Kandasamy. Implementing geometric complexity theory: On the separation of orbit closures via symmetries.
1820 In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, page 713–726, 2020. 7
- [IP17] C. Ikenmeyer and G. Panova. Rectangular Kronecker coefficients and plethysms in geometric complexity theory. *Adv. Math.*, 319:40–66, 2017. 7

- 1821 [IQS18] Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. [Non-commutative Edmonds' problem and matrix semi-invariants](#). *Comput. Complex.*, 26(3):717–763, 2018. 4
- 1822
- 1823 [Kal85] Kyriakos Kalorkoti. [A Lower Bound for the Formula Size of Rational Functions](#). *SIAM Journal of Computing*, 14(3):678–687, 1985. 6
- 1824
- 1825 [KI03] Valentine Kabanets and Russell Impagliazzo. [Derandomizing polynomial identity tests means proving circuit lower bounds](#). In Lawrence L. Larmore and Michel X. Goemans, editors, *Proceedings of the 35th Annual ACM Symposium on Theory of Computing, June 9-11, 2003, San Diego, CA, USA*, pages 355–364. ACM, 2003. 8
- 1826
- 1827 [KLSS17] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. [An exponential lower bound for homogeneous depth four arithmetic formulas](#). *SIAM Journal on Computing*, 46(1):307–335, 2017. 9
- 1828
- 1829 [Kra84] H. Kraft. *Geometrische Methoden in der Invariantentheorie*. Aspects of Mathematics, D1. Friedr. Vieweg & Sohn, Braunschweig, 1984. 5
- 1830
- 1831 [KRST20] Mrinal Kumar, C Ramya, Ramprasad Saptharishi, and Anamay Tengse. [If VNP is hard, then so are equations for it](#). *arXiv preprint arXiv:2012.07056*, 2020. 4
- 1832
- 1833 [KS16] Neeraj Kayal and Chandan Saha. [Lower bounds for depth-three arithmetic circuits with small bottom fanin](#). *computational complexity*, 25(2):419–454, 2016. 9
- 1834
- 1835 [KS17] Mrinal Kumar and Shubhangi Saraf. [On the power of homogeneous depth 4 arithmetic circuits](#). *SIAM Journal on Computing*, 46(1):336–387, 2017. 9
- 1836
- 1837 [KS19] Mrinal Kumar and Ramprasad Saptharishi. [Hardness-randomness tradeoffs for algebraic computation](#). *Bulletin of EATCS*, 3(129), 2019. 8
- 1838
- 1839 [Kum20] Mrinal Kumar. [On the Power of Border of Depth-3 Arithmetic Circuits](#). *ACM Trans. Comput. Theory*, 12(1):5:1–5:8, 2020. 3, 5, 6, 7, 8, 16, 24, 39
- 1840
- 1841 [Lan15] Joseph M Landsberg. [Geometric complexity theory: an introduction for geometers](#). *Annali dell'universita'di Ferrara*, 61(1):65–117, 2015. 6, 8
- 1842
- 1843 [Lan17] J. M. Landsberg. *Geometry and Complexity Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2017. 6, 7
- 1844
- 1845 [LO15] Joseph M. Landsberg and Giorgio Ottaviani. [New Lower Bounds for the Border Rank of Matrix Multiplication](#). *Theory Comput.*, 11:285–298, 2015. 4
- 1846
- 1847 [LST21] Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. [Superpolynomial Lower Bounds Against Low-Depth Algebraic Circuits](#). *62nd Annual IEEE Symposium on the Foundations of Computer Science (FOCS 2021)*, page 81, 2021. 7, 8, 9, 10
- 1848
- 1849 [Mah13] Meena Mahajan. [Algebraic Complexity Classes](#). *CoRR*, abs/1307.3863, 2013. 3
- 1850
- 1851 [MS01] Ketan Mulmuley and Milind A. Sohoni. [Geometric Complexity Theory I: An Approach to the P vs. NP and Related Problems](#). *SIAM J. Comput.*, 31(2):496–526, 2001. 4, 7
- 1852
- 1853 [MS08] K. D. Mulmuley and M. Sohoni. [Geometric Complexity Theory II: Towards explicit obstructions for embeddings among class varieties](#). *SIAM J. Computing*, 38(3):1175–1206, 2008. 4
- 1854
- 1855 [Muk16] Partha Mukhopadhyay. [Depth-4 Identity Testing and Noether's Normalization Lemma](#). In *CSR 2016*, 2016. 4
- 1856
- 1857 [Mul12a] Ketan Mulmuley. [Geometric Complexity Theory V: Equivalence between Blackbox Derandomization of Polynomial Identity Testing and Derandomization of Noether's Normalization Lemma](#). In *53rd Annual IEEE Symposium on the Foundations of Computer Science (FOCS 2012)*, pages 629–638, 2012. 4
- 1858
- 1859 [Mul12b] Ketan D Mulmuley. [The GCT program toward the P vs. NP problem](#). *Communications of the ACM*, 55(6):98–107, 2012. 4
- 1860
- 1861 [Mum95] David Mumford. *Algebraic geometry I: complex projective varieties*. Springer Science & Business Media, 1995. 5
- 1862
- 1863 [Niv69] Ivan Niven. [Formal Power Series](#). *The American Mathematical Monthly*, 76(8):871–889, 1969. 11
- 1864
- 1865 [NW96] Noam Nisan and Avi Wigderson. [Lower bounds on arithmetic circuits via partial derivatives](#). *Computational complexity*, 6(3):217–234, 1996. 9
- 1866
- 1867 [Sap19] Ramprasad Saptharishi. [A survey of lower bounds in arithmetic circuit complexity](#). *GitHub Survey*, 2019. 6
- 1868
- 1869 [Sax08] Nitin Saxena. [Diagonal Circuit Identity Testing and Lower Bounds](#). In *ICALP 2008*, pages 60–71, 2008. 38
- 1870
- 1871 [SHL65] Richard Edwin Stearns, Juris Hartmanis, and Philip M Lewis. [Hierarchies of memory limited computations](#). In *6th annual symposium on switching circuit theory and logical design (SWCT 1965)*, pages 179–190. IEEE, 1965. 2
- 1872
- 1873 [Shp02] Amir Shpilka. [Affine projections of symmetric polynomials](#). *Journal of Computer and System Sciences*, 65(4):639–659, 2002. 7
- 1874
- 1875 [Sin19] Amit Kumar Sinhababu. [Power series in complexity: Algebraic Dependence, Factor Conjecture and Hitting Set for Closure of VP](#). PhD thesis, Indian Institute of Technology Kanpur, 2019. 11
- 1876
- 1877 [Str73] Volker Strassen. [Die Berechnungskomplexität von elementarsymmetrischen Funktionen und von Interpolationskoeffizienten](#). *Numerische Mathematik*, 20(3):238–251, 1973. 6
- 1878
- 1879 [Str74] Volker Strassen. [Polynomials with Rational Coefficients Which are Hard to Compute](#). *SIAM J. Comput.*, 3(2):128–149, 1974. 4
- 1880
- 1881 [SW01] Amir Shpilka and Avi Wigderson. [Depth-3 arithmetic circuits over fields of characteristic zero](#). *Computational Complexity*, 10(1):1–27, 2001. CCC 1999. 9

- [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. 3
- [Val79] Leslie G. Valiant. [Completeness Classes in Algebra](#). In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing*, pages 249–261, 1979. 3

A BASICS OF ALGEBRAIC COMPLEXITY

Definition 3 (Algebraic Branching Program (ABP)). *ABP is a computational model which is described using a layered graph with a source vertex s and a sink vertex t . All edges connect vertices from layer i to $i + 1$. Further, edges are labelled by univariate polynomials. The polynomial computed by the ABP is defined as*

$$f = \sum_{\text{path } \gamma: s \rightsquigarrow t} \text{wt}(\gamma)$$

where $\text{wt}(\gamma)$ is product of labels over the edges in path γ . Number of layers (Δ) defines the *depth* and the maximum number of vertices in any layer (w) defines the *width* of an ABP. The *size* (s) of an ABP is the sum of the graph-size and the degree of the univariate polynomials that label. If d is the maximum degree of univariates then $s \leq dw^2\Delta$; in fact, we will take the latter as the ABP-size bound in our calculations.

Our interest primarily is in the following two ABP-variants: ROABP and ARO.

Definition 4 (Read-once Oblivious Algebraic Branching Program (ROABP)). *An ABP is defined as Read-Once Oblivious Algebraic Branching Program (ROABP) in a variable order $(x_{\sigma(1)}, \dots, x_{\sigma(n)})$ for some permutation $\sigma: [n] \rightarrow [n]$, if edges of i -th layer of ABP are univariate polynomials in $x_{\sigma(i)}$.*

Definition 5 (Any-order ROABP (ARO)). *A polynomial $f \in \mathbb{F}[x]$ is computable by ARO of size s if for all possible permutation of variables there exists a ROABP of size at most s in that variable order.*

Remark. We can de-border $\overline{\Sigma \wedge \Sigma}$. Since $\overline{\Sigma \wedge \Sigma} \subseteq \overline{\text{ARO}}$, using duality trick (Lemma 31) and $\overline{\text{ARO}} = \text{ARO}$, from Nisan’ characterization (Lemma 32), it follows that $\overline{\Sigma \wedge \Sigma} \subseteq \text{ARO}$. Note that, $\overline{\Sigma \wedge \Sigma}$ is a *strict* subset of ARO since $\prod_{i=1}^n x_i$ has a small ARO, but it requires $\exp(n)$ -size $\overline{\Sigma \wedge \Sigma}$ -circuits.

B TECHNICAL LEMMAS

Here is an important lemma to show that positive valuation with respect to y , lets us express a function as a power-series of y .

Lemma 24 (Valuation lemma, [DDS21, Lemma A.17]). *Let $f \in \mathbb{F}(x, y)$ such that $\text{val}_y(f) \geq 0$. Then, $f \in \mathbb{F}(x)[[y]] \mathbb{F}(x, y)$.*

In this section we will also discuss various properties of $\Sigma \wedge \Sigma$ circuits and basic waring-rank. The corresponding bloated model is $\Sigma \wedge \Sigma / \Sigma \wedge \Sigma$, that computes elements of the form f/g , where $f, g \in \Sigma \wedge \Sigma$. For the detailed proofs, we refer to [DDS21].

Firstly, it is known that $\Sigma \wedge \Sigma$ is *closed* under *constant-fold* multiplication.

Lemma 25 ($\Sigma \wedge \Sigma$ closed under multiplication, [DDS21, Lemma A.10]). *Let $f_i \in \mathbb{F}[x]$, of syntactic degree $\leq d_i$, be computed by a $\Sigma \wedge \Sigma$ circuit of size s_i , for $i \in [k]$. Then, $f_1 \cdots f_k$ has $\Sigma \wedge \Sigma$ circuit of size $O((d_2 + 1) \cdots (d_k + 1) \cdot s_1 \cdots s_k)$.*

Using the additive and multiplicative closure of $\Sigma \wedge \Sigma$, one can show that $\Sigma \wedge \Sigma / \Sigma \wedge \Sigma$ is closed under constant-fold addition.

1925 **Lemma 26** ($\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ closed under addition, [DDS21, Lemma A.11]). Let $f_i \in \mathbb{F}[x]$, of syntactic degree
 1926 $< d_i$, be computable by $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ of size s_i , for $i \in [k]$. Then, $\sum_{i \in [k]} f_i$ has a $(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$ representation of size
 1927 $O((\prod_i d_i) \cdot \prod_i s_i)$.
 1928

1929 Using a simple interpolation, the coefficient of y^ϵ can be extracted from $f(x, y) \in \Sigma\wedge\Sigma$ again as a small $\Sigma\wedge\Sigma$
 1930 representation.
 1931

1932 **Lemma 27** ($\Sigma\wedge\Sigma$ coefficient extraction, [DDS21, Lemma A.12]). Let $f(x, y) \in \mathbb{F}[x][y]$ be computed by a $\Sigma\wedge\Sigma$
 1933 circuit of size s and degree d . Then, $\text{coef}_{y^\epsilon}(f) \in \mathbb{F}[x]$ is a $\Sigma\wedge\Sigma$ circuit of size $O(sd)$, over $\mathbb{F}[x]$.
 1934

1935 Like coefficient extraction, differentiation of $\Sigma\wedge\Sigma$ circuit is easy too.
 1936

1937 **Lemma 28** ($\Sigma\wedge\Sigma$ differentiation, [DDS21, Lemma A.13]). Let $f(x, y) \in \mathbb{F}[x][y]$ be computed by a $\Sigma\wedge\Sigma$ circuit of
 1938 size s and degree d . Then, $\partial_y(f)$ is a $\Sigma\wedge\Sigma$ circuit of size $O(sd^2)$, over $\mathbb{F}[x][y]$.
 1939

1940 Let \mathcal{C} and \mathcal{D} be two classes over $\mathbb{F}[x]$. Consider the bloated-class $(\mathcal{C}/\mathcal{C}) \cdot (\mathcal{D}/\mathcal{D})$, which has elements of the
 1941 form $(g_1/g_2) \cdot (h_1/h_2)$, where $g_i \in \mathcal{C}$ and $h_i \in \mathcal{D}$ ($g_2 h_2 \neq 0$). One can also similarly define its border (which
 1942 will be an element in $\mathbb{F}(x)$). Here is an important observation.
 1943

1944 **Lemma 29** ([DDS21, Lemma A.19]). $\overline{(\mathcal{C}/\mathcal{C}) \cdot (\mathcal{D}/\mathcal{D})} \subseteq \overline{(\mathcal{C}/\mathcal{C})} \cdot \overline{(\mathcal{D}/\mathcal{D})}$.
 1945

1946 **PROOF.** Suppose $(g_1/g_2) \cdot h_1/h_2 = f + \epsilon \cdot Q$, where $Q \in \mathbb{F}(x, \epsilon)$ and $f \in \mathbb{F}(x)$. Let $\text{val}_\epsilon(g_i) =: a_i$ and
 1947 $\text{val}_\epsilon(h_i) =: b_i$. Denote, $\tilde{g}_i =: \epsilon^{a_i} \cdot \tilde{g}_i$, similarly \tilde{h}_i . Further, assume $\tilde{g}_i =: \hat{g}_i + \epsilon \cdot \hat{g}'_i$; similarly for \tilde{h}_i , we define
 1948 $\hat{h}_i \in \mathbb{F}[x]$. Note that $\hat{g}_i \in \overline{\mathcal{C}}$, similarly $\hat{h}_i \in \overline{\mathcal{D}}$.
 1949

1950 So, LHS = $\epsilon^{a_1 - a_2 + b_1 - b_2} \cdot (\tilde{g}_1/\tilde{g}_2) \cdot (\tilde{h}_1/\tilde{h}_2)$. This has a limit $\lim_{\epsilon \rightarrow 0}$, so $a_1 + b_1 - a_2 - b_2 \geq 0$. If it is ≥ 1 , the
 1951 limit in RHS is 0 and so $f = 0$. If $a_1 + b_1 - a_2 - b_2 = 0$, then
 1952

$$f = (\hat{g}_1/\hat{g}_2) \cdot (\hat{h}_1/\hat{h}_2) \in \overline{(\mathcal{C}/\mathcal{C})} \cdot \overline{(\mathcal{D}/\mathcal{D})}.$$

1953
 1954
 1955 □
 1956

1957 **Lemma 30** ([Gho19, Lemma 2.3.15]). Let \mathbb{F} be a field of characteristic 0 or greater than d . Let \mathcal{P} be a set of n -variate
 1958 degree d polynomials over \mathbb{F} such that for all $P \in \mathcal{P}$, the dimension of the partial derivative space of P is at most k . Then,
 1959 every nonzero $P \in \mathcal{P}$ has a $(\leq k)$ -cone-size leading-monomial.
 1960

1961 **Lemma 31** (Duality trick [Sax08]). The polynomial $f = (x_1 + \dots + x_n)^d$ can be written as
 1962

$$f = \sum_{i \in [t]} f_{i1}(x_1) \cdots f_{in}(x_n),$$

1963 where $t = O(nd)$, and f_{ij} is a univariate polynomial of degree at most d .
 1964

1965 We remark that the above proof works for fields of characteristic = 0, or $> d$. This lemma eventually shows
 1966 that $\Sigma\wedge\Sigma \subseteq \text{ARO}$.
 1967

1968 Next we state that polynomials approximated by ARO can be easily de-bordered. To the best of our knowledge
 1969 the following lemma was sketched in [For16]; also implicitly in [GKS16]. For a detailed proof, see [DDS21,
 1970 Lemma A.21]
 1971

1972 **Lemma 32** (De-bordering ARO). Consider a polynomial $f \in \mathbb{F}[x]$ which is approximated by ARO of size s over
 1973 $\mathbb{F}(\epsilon)[x]$. Then, there exists an ARO of size s which exactly computes $f(x)$.
 1974

1975 Manuscript submitted to ACM
 1976

C LIMITATIONS OF BOUNDED FAN-IN DEPTH-3 CIRCUITS

The following theorem is a folklore (see [Kum20]), which shows that bounded depth-3 circuits are *not* universal. We present a proof below, for the completeness.

THEOREM 33. *The inner product polynomial*

$$\text{IP}_n = \sum_{i=1}^n x_i \cdot y_i,$$

cannot be computed by any $\Sigma^{[n-1]}\Pi\Sigma$ circuit, no matter how large the product fan-in is.

PROOF. Assume, for the sake of contradiction, IP_n can be computed by a $\Sigma\Pi\Sigma$ circuit of top fan-in $n - 1$:

$$\text{IP}_n = \sum_{i=1}^{n-1} \prod_{j=1}^{d_i} \ell_{ij}. \quad (21)$$

The proof idea is as follows: we reduce Equation 21 modulo $n - 1$ many linear polynomials suitably picked from each summand of the RHS. We get a contradiction, as the RHS becomes zero modulo those linear polynomials, but the LHS remains nonzero. Formally, we implement the proof below.

Assume, wlog, $\prod_{j=1}^{d_1} \ell_{1j}$ contains the variable x_1 in one of its factors. For some j , wlog we have $\ell_{1j} = x_1 - r_1(x_2, \dots, x_n, \mathbf{y})$. If $\ell_{1j} = a_1 x_1 - r_1(x_2, \dots, x_n, \mathbf{y})$, we can take out a_1 , and work with $(x_1 - r_1/a_1)$. Taking mod by $(x_1 - r_1/a_1)$ essentially means substituting $x_1 = r_1/a_1$, in Equation 21.

Now, we go modulo $x_1 - r_1$ in both sides of Equation 21. This changes the IP_n polynomial in the LHS, which becomes $r_1 y_1 + x_2 y_2 + \dots + x_n y_n$. In the RHS, the first summand vanishes! So, we get

$$r_1 y_1 + x_2 y_2 + \dots + x_n y_n = \sum_{i=2}^{n-1} \prod_{j=1}^{d_i} \ell_{ij}(r_1, x_2, \dots, x_n, \mathbf{y}). \quad (22)$$

Now, note that $r_1 y_1 + x_2 y_2 + \dots + x_n y_n$ is *not free* of x_2 , as $r_1 y_1$ cannot cancel the term $x_2 y_2$. Thus, x_2 must be present also in the RHS of Equation 22. Wlog, assume that x_2 is present in $\ell_{2j}(r_1, x_2, \dots, x_n, \mathbf{y})$, for some j . Assume it is of the form $x_2 - r_2(x_3, \dots, x_n, \mathbf{y})$.

Now, we reduce Equation 22 modulo $x_2 - r_2$. This changes the LHS to

$$r_1(r_2, \dots, x_n, \mathbf{y})y_1 + r_2 y_2 + x_3 y_3 + x_n y_n.$$

At least one term in the RHS gets vanished. The resulting polynomial in the LHS is not free of x_3 . Thus, x_3 must be present in the RHS too. We assume there is a linear term in RHS of the form $(x_3 - r_3)$. Next, we go modulo $(x_3 - r_3)$.

We can continue this, $n - 1$ times, and in the end RHS would completely vanish. In the LHS, we would have $\tilde{r}_1 y_1 + \tilde{r}_2 y_2 + \dots + \tilde{r}_{n-1} y_{n-1} + x_n y_n$, for some linear polynomials $\tilde{r}_1, \dots, \tilde{r}_{n-1}$. This polynomial would have a nonzero term $x_n y_n$, as $x_n y_n$ cannot be cancelled by $\tilde{r}_1 y_1 + r_2 y_2 + \dots + \tilde{r}_{n-1} y_{n-1}$. This leads to a contradiction. \square

Remark. The above proof fails in the approximative setting, mainly because it could happen that each $\ell_{ij} = 1 \pmod{\langle x \rangle}$, (e.g. $1 + \epsilon x_1$). In this case, $x_1 = -1/\epsilon$ is an *illegal* substitution, which makes the proof fail dismally.