DEMYSTIFYING THE BORDER OF DEPTH-3 ALGEBRAIC CIRCUITS*

PRANJAL DUTTA†, PRATEEK DWIVEDI‡, AND NITIN SAXENA§

Abstract. Border complexity of polynomials plays an integral role in the GCT (Geometric complexity theory) approach to $\mathsf{P} \neq \mathsf{NP}$. It tries to formalize the notion of 'approximating a polynomial' via limits (Bürgisser FOCS'01). This raises the open question $\overline{\mathsf{VP}} \stackrel{?}{=} \mathsf{VP}$, as the approximation involves *exponential precision* which may not be efficiently simulable. Recently (Kumar TOCT'20) proved the universal power of the border of top fan-in two depth-3 circuits ($\overline{\Sigma^{[2]}\Pi\Sigma}$). Here we answer some of the related open questions. We show that the border of bounded top fan-in depth-3 circuits ($\overline{\Sigma^{[k]}\Pi\Sigma}$ for constant k) is relatively easy—it can be computed by a polynomial size algebraic branching program (ABP). There were hardly any *de-bordering* results known for prominent models before our result.

Moreover, we give the *first* quasipolynomial-time black-box identity test for the same. Prior best construction was in PSPACE (Forbes,Shpilka STOC'18). Also, with more technical work, we extend our results to restricted depth-4 circuits. Our de-bordering paradigm is a multi-step process; in short we call it DiDIL –divide, derive, induct, with limit. It 'almost' reduces $\overline{\Sigma^{[k]}\Pi\Sigma}$ to special cases of read-once oblivious algebraic branching programs (ROABPs) in any-order.

Key words. approximative, border, depth-3, depth-4, circuits, de-border, derandomize, black-box, PIT, GCT, any-order ROABP, ABP, VBP, VP, VNP.

AMS subject classifications. 32A05, 32A40, 68W30, 68Q15, 68Q25

1. Introduction: Border complexity, GCT and beyond.

Algebraic circuits are a natural and a non-uniform model of polynomial computation, which forms the basis for the vast study of algebraic complexity. We say that a polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$, over a field $\mathbb{F}$ is computable by a circuit of size s and depth d if there exists a directed acyclic graphs of size s (nodes + edges) and depth d such that its leaf nodes are labelled by variables or field constants, internal nodes are labelled with $+$ and $\times$, and the polynomial computed at the root is f. Further, if the output of a gate is never re-used then it is a *Formula*. Any formula can be converted into a layered graph called *Algebraic Branching Program* (ABP). Various complexity measures can be defined on the computational model to classify polynomials in different complexity classes. For example VP (respectively VBP, respectively VF) is the class of polynomials of polynomial degree, computable by polynomial-sized circuits (respectively ABPs, respectively formulas). Finally, VNP is the class of polynomials which can be expressed as an exponential-sum of projection of a VP circuit family. For more details, refer to subsection 2.1 and [119, 113, 86].

The problem of separating algebraic complexity classes has been a central theme of this study. As an algebraic analog of P vs. NP problem, Valiant [119] conjectured that $\mathsf{VBP} \neq \mathsf{VNP}$ and further strengthened it by conjecturing $\mathsf{VP} \neq \mathsf{VNP}$. Over the years, impressive progress has been made towards resolving this, however, the existing tools have not been able to resolve this conclusively. Towards settling these conjectures

*A preliminary version appeared in 62nd Symposium on Foundations of Computer Science (FOCS 2021).

Funding: Pranjal Dutta: "Foundation of Lattice-based Cryptography", by NUS-NCS Joint Laboratory for Cyber Security; Nitin Saxena: DST (SJF/MSA-01/2013-14), DST-SERB (CRG/2020/000045), and N. Rama Rao Chair.

†National University of Singapore, Singapore. (pranjal@nus.edu.sg).

‡Dept. of Computer Science & Engineering, IIT Kanpur (pdwivedi@cse.iitk.ac.in).

§Dept. of Computer Science & Engineering, IIT Kanpur (nitin@cse.iitk.ac.in).

Mulmuley and Sohoni [92] introduced *Geometric Complexity Theory* (GCT) program. In this program, they studied the border (or approximative) complexity, with the aim of approaching Valiant's conjecture and strengthening it to: $\mathsf{VNP} \not\subseteq \overline{\mathsf{VBP}}$, or equivalently, the padded permanent does not lie in the orbit closure of 'small' determinants. This notion was already studied in the context of designing matrix multiplication algorithms [116, 17, 18, 36, 82]. The hope, in the GCT program, was to use tools from algebraic geometry and representation theory, and possibly settle the question once and for all. This also gave a natural reason to understand the relationship between VP and $\overline{\mathsf{VP}}$ (or VBP and $\overline{\mathsf{VBP}}$).

In addition to the VP vs. VNP implication, GCT has deep connections with computational invariant theory [50, 90, 53, 29, 69], algebraic natural proofs [57, 21, 34, 79], lower bounds [30, 56, 82], optimization [8, 28] and many more. We refer to [31, Sec. 9] and [90, 91] for expository references.

The simplest notion of the approximative closure comes from the following definition [25, 26]: a polynomial $f(\boldsymbol{x}) \in \mathbb{F}[x_1, \ldots, x_n]$ is approximated by $g(\boldsymbol{x}, \varepsilon) \in \mathbb{F}(\varepsilon)[\boldsymbol{x}]$ if there exists a $Q(\boldsymbol{x}, \varepsilon) \in \mathbb{F}[\varepsilon][\boldsymbol{x}]$ such that $g = f + \varepsilon\, Q$. When $\mathbb{F} = \mathbb{R}$, and under Euclidean topology, we can analytically think of approximation as $\lim_{\varepsilon \to 0} g = f$. If g belongs to a circuit class $\mathcal{C}$ (over $\mathbb{F}(\varepsilon)$, i.e. any *arbitrary* ε-power is allowed as 'cost-free' constants), then we say that $f \in \overline{\mathcal{C}}$, the approximative closure of $\mathcal{C}$. Further, one could draw parallels with algebraic definition of *Zariski* closure that works over every field, i.e. taking the closure of the set of polynomials (considered as points) of $\mathcal{C}$: Let $\mathcal{I}$ be the smallest (annihilating) ideal whose zeros cover $\{\text{coefficient-vector of } g \mid g \in \mathcal{C}\}$; then put in $\overline{\mathcal{C}}$ each polynomial f with coefficient-vector being a zero of $\mathcal{I}$. Interestingly, all these notions are equivalent over the algebraically closed field (refer [25, Theorem 2.4] and [94, §2.C]).

The size of the circuit computing g defines the *approximative* (or *border*) complexity of f, denoted $\overline{\mathsf{size}}(f)$; evidently, $\overline{\mathsf{size}}(f) \leq \mathsf{size}(f)$. Due to the possible $1/\varepsilon^M$ terms in the circuit computing g, evaluating it at $\varepsilon = 0$ may not be necessarily valid (though the limit exists). Hence, given $f \in \overline{\mathcal{C}}$, does not immediately reveal anything about the *exact* complexity of f. Since $g(\boldsymbol{x}, \varepsilon) = f(\boldsymbol{x}) + \varepsilon \cdot Q(\boldsymbol{x}, \varepsilon)$, we could extract the coefficient of ε^0 from g using the standard interpolation trick, by setting random ε-values from $\mathbb{F}$. However, the trivial bound on the circuit size of f would depend on the degree M of ε, which could provably be *exponential* in the size of the circuit computing g, i.e. $\overline{\mathsf{size}}(f) \leq \mathsf{size}(f) \leq \exp(\overline{\mathsf{size}}(f))$ [25, Thm. 5.7].

1.1. De-bordering: The upper bound results. The major focus of this paper is to address the power of approximation in the restricted circuit classes. Given a polynomial $f \in \overline{\mathcal{C}}$, for an interesting class $\mathcal{C}$, we want to upper bound the exact complexity of f (we call it 'de-bordering'). If $\mathcal{C} = \overline{\mathcal{C}}$, then $\mathcal{C}$ is said to be closed under approximation: For example 1) $\Sigma\Pi$, sparse polynomials (with complexity measure being sparsity), 2) Monotone ABPs [22], and 3) ROABP (read-once ABP) and ARO (*any-order* ROABP), with measure being the width. ARO is an ABP with a natural restriction on the use of variables per layer; for definition and a formal proof, see Definition 2.8 and Lemma 2.22.

Why care about upper bounds? One of the fundamental questions in the GCT paradigm is whether $\overline{\mathsf{VP}} \stackrel{?}{=} \mathsf{VP}$ [91, 58]. Confirmation or refutation of this question has multiple consequences, both in the algebraic complexity and at the frontier of algebraic geometry. If $\mathsf{VP} = \overline{\mathsf{VP}}$, then any proof of $\mathsf{VP} \neq \mathsf{VNP}$ will in fact also show that $\mathsf{VNP} \not\subseteq \overline{\mathsf{VP}}$, as conjectured in [90]; however a refutation would imply that any realistic approach to the VP vs. VNP conjecture would even have to separate the permanent from the families in $\overline{\mathsf{VP}}\backslash\mathsf{VP}$ (and for this, one needs a far better understanding than the current state of the art).

The other significance of the upper bound result arises from the *flip* [89, 90] whose basic idea in a nutshell is to understand the theory of upper bounds first, and then use this theory to prove lower bounds later. Taking this further to the realm of algorithms: showing de-bordering results, for even restricted classes (for example depth-3, small-width ABPs), could have potential identity testing implications. For details, see subsection 1.2.

De-bordering results in GCT are in a very nascent stage; for example, the boundary of 3×3 determinants was only recently understood [68]. Note that here both the number of variables n and the degree d are constant. In this work, however, we target polynomial families with both n and d unbounded. So getting exact results about such border models is highly nontrivial considering the current state of the art.

De-bordering small-width ABPs. The exponential degree dependence of ε [25, 26] suggests us to look for separation of restricted complexity classes or try to upper bound them by some other means. In [24], the authors showed that $\mathsf{VBP}_2 \subseteq \overline{\mathsf{VBP}_2} = \overline{\mathsf{VF}}$; here VBP_2 denotes the class of polynomials computed by width-2 ABP. Surprisingly, we also know that $\mathsf{VBP}_2 \subsetneq \mathsf{VF} = \mathsf{VBP}_3$ [13, 9]. Very recently, [22] showed polynomial gap between ABPs and border-ABPs, in the trace model, for noncommutative and also for commutative monotone settings (along with $\mathsf{VQP} \neq \overline{\mathsf{VNP}}$).

Quest for de-bordering depth-3 circuits. Outside such ABP results and depth-2 circuits, we understand very little about the border of other important models. Thus, it is natural to ask the same for depth-3 circuits, plausibly starting with depth-3 diagonal circuits ($\Sigma\wedge\Sigma$), i.e. polynomials of the form $\sum_{i\in[s]} c_i \cdot \ell_i^d$, where ℓ_i are linear polynomials. Interestingly, the relation between Waring rank (minimum s to compute f) and *border* Waring rank (minimum s, to approximate f) has been studied in mathematics for ages [117, 23, 15, 54], yet it is not clear whether the measures are polynomially related or not. However, we point out that $\overline{\Sigma\wedge\Sigma}$ has a small ARO; this follows from the fact that $\Sigma\wedge\Sigma$ has small ARO by the *duality trick* [106], and ARO is closed under approximation [95, 46]; for details see Lemma 2.23.

This pushes us further to study depth-3 circuits $\Sigma^{[k]}\Pi^{[d]}\Sigma$; these circuits compute polynomials of the form $f = \sum_{i\in[k]} \prod_{j\in[d]} \ell_{ij}$ where ℓ_{ij} are linear polynomials. This model with bounded fan-in has been a source of great interest for derandomization [43, 74, 71, 109, 6]. In a recent twist, Kumar [78] showed that border depth-3 fan-in two circuits are 'universally' expressive; i.e. $\overline{\Sigma^{[2]}\Pi^{[D]}\Sigma}$ over $\mathbb{C}$ can approximate *any* homogeneous d-degree, n-variate polynomial; though his expression requires an exceedingly large $D = \exp(n, d)$.

Our upper bound results. The universality result of border depth-3 fan-in three circuits makes it imperative to study $\overline{\Sigma^{[2]}\Pi^{[d]}\Sigma}$, for $d = \mathsf{poly}(n)$ and understand its computational power. To start with, are polynomials in this class even 'explicit' (i.e. the coefficients are efficiently computable)? If yes, is $\overline{\Sigma^{[2]}\Pi^{[d]}\Sigma} \subseteq \mathsf{VNP}$? (See [58, 98] for more general questions in the same spirit.) To our surprise, we show that the class is very explicit; in fact every polynomial in this class has a small ABP. The statement and its proof is first of its kind which eventually uses analytic approach and 'reduces' the Π-gate to $\wedge$-gate. We remark that it does not reveal the polynomial dependence on the ε-degree. However, this positive result could be thought as a baby step towards $\overline{\mathsf{VP}} = \mathsf{VP}$. We assume the field $\mathbb{F}$ characteristic to be $= 0$, or large enough. For a detailed statement, see Theorem 3.2.

THEOREM 1.1 (De-bordering depth-3 circuits). *For any constant k, $\overline{\Sigma^{[k]}\Pi\Sigma} \subseteq$ VBP, i.e. any polynomial in the border of constant top fan-in size-s depth-3 circuits, can also be computed by a $\mathsf{poly}(s)$-size algebraic branching program (ABP).*

Remarks. 1. When $k = 1$, it is easy to show that $\overline{\Pi\Sigma} = \Pi\Sigma$ [24, Prop. A.12] (see Lemma 2.21).

2. The size of the ABP turns out to be $s^{\exp(k)}$. It is an interesting open question whether $f \in \overline{\Sigma^{[k]}\Pi\Sigma}$ has a subexponential ABP when $k = \Theta(\log s)$.

3. $\overline{\Sigma^{[k]}\Pi\Sigma}$ is the *orbit closure* of k-sparse polynomials [87, Thm. 1.31]. Understanding the orbit and its closure of certain classes is at the core of the GCT program. Theorem 1.1 is one of the first results that deborder orbit closures, in particular closure of constant-sparse polynomials.

Extending to depth-4. Once we have dealt with depth-3 circuits, it is natural to ask the same for constant top fan-in depth-4 circuits. Polynomials computed by $\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}$ circuits are of the form $f = \sum_{i\in[k]}\prod_j g_{ij}$ where $\deg(g_{ij}) \le \delta$. Unfortunately, our technique cannot be generalised to this model, primarily due to the inability to de-border $\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$. However, when the bottom Π is replaced by $\wedge$, we can show $\overline{\Sigma^{[k]}\Pi\Sigma\wedge} \subseteq \mathsf{VBP}$; we sketch the proof in Theorem 5.1.

1.2. Derandomizing the border: The black-box PITs.

Polynomial Identity Testing (PIT) is one of the fundamental decision problems in complexity theory. The Polynomial Identity Lemma [99, 38, 121, 111] gives an efficient randomized algorithm to test the zeroness of a given polynomial, even in the black-box settings (known as Black-box PIT), where we are not allowed to see the internal structure of the model (unlike the 'whitebox' setting), but evaluations at points are allowed. It is still an open problem to derandomize black-box PIT. Designing a *deterministic* black-box PIT algorithm for a circuit class is equivalent to finding a set of points such that for every nonzero circuit, the set contains a point where it evaluates to a nonzero value [47, Sec. 3.2]. Such a set is called *hitting* set.

A trivial explicit hitting set for a class of degree d polynomial of size $O(d^n)$ can be obtained using the Polynomial Identity Lemma. Heintz and Schnorr [67] showed that $\mathsf{poly}(s, n, d)$ size hitting set *exists* for d-degree, n-variate polynomials computed (as well as approximated) by circuits of size s. However, the real challenge is to efficiently obtain such an *explicit* set.

Constructing small size explicit hitting set for VP is a long standing open problem in algebraic complexity theory, with numerous algorithmic applications in graph theory [85, 93, 45], factoring [77, 41], cryptography [5], and hardness vs randomness results [67, 96, 1, 70, 44, 42]. Moreover, a long line of depth reduction results [120, 7, 76, 118, 64] and the bootstrapping phenomenon [3, 81, 61, 10] has justified the interest in hitting set construction for restricted classes; e.g. depth 3 [43, 74, 109, 6], depth 4 [51, 12, 48, 112, 100, 101, 39], ROABPs [4, 66, 51, 60, 19] and log-variate depth-3 diagonal circuits [49]. We refer to [113, 107, 80] for expositions.

PIT in the border. In this paper we address the question of constructing hitting set for restrictive border circuits. $\mathcal{H}$ is a hitting set for a class $\overline{\mathcal{C}}$, if $g(\boldsymbol{x}, \varepsilon) \in \mathcal{C}_{\mathbb{F}(\varepsilon)}$, approximates a *non-zero* polynomial $f(\boldsymbol{x}) \in \overline{\mathcal{C}}$, then $\exists \boldsymbol{a} \in \mathcal{H}$ such that $g(\boldsymbol{a}, \varepsilon) \notin \varepsilon\cdot\mathbb{F}[\varepsilon]$, i.e. $f(\boldsymbol{a}) \neq 0$. Note that, as $\mathcal{H}$ will also 'hit' polynomials of class $\mathcal{C}$, construction of hitting set for the border classes (we call it 'border PIT') is a natural and possibly a different avenue to derandomize PIT. Here, we emphasize that $\boldsymbol{a} \in \mathbb{F}^n$ such that $g(\boldsymbol{a}, \varepsilon) \neq 0$, *may not* hit the limit polynomial f since $g(\boldsymbol{a}, \varepsilon)$ might still lie in $\varepsilon \cdot \mathbb{F}[\varepsilon]$; because f could have really high complexity compared to g. Intrinsically, this property makes it harder to construct an explicit hitting set for $\overline{\mathsf{VP}}$.

We also remark that there is no 'whitebox' setting in the border and thus we cannot really talk about 't-time algorithm'; rather we would only be using the term 't-time hitting set', since the given circuit after evaluating on $\boldsymbol{a} \in \mathbb{F}^n$, may require *arbitrarily* high-precision in $\mathbb{F}(\varepsilon)$.

Prior known border PITs. Mulmuley [91] asked the question of constructing an efficient hitting set for $\overline{\mathsf{VP}}$. Forbes and Shpilka [52] gave a PSPACE algorithm over the field $\mathbb{C}$. In [62], the authors extended this result to *any* field. Very few better hitting set constructions are known for the restricted border classes, for example poly-time hitting set for $\overline{\Pi\Sigma} = \Pi\Sigma$ [14, 75], quasi-poly hitting set for $\overline{\Sigma\wedge\Sigma} \subseteq \overline{\text{ARO}} \subseteq \overline{\text{ROABP}}$ [51, 4, 66] and poly-time hitting set for the border of a restricted sum of log-variate ROABPs [19].

Why care about border PIT? PIT for $\overline{\mathsf{VP}}$ has a lot of applications in the context of algebraic geometry and computational complexity, as observed by Mulmuley [91]. For example Noether's Normalization Lemma (NNL); it is a fundamental result in algebraic geometry where the computational problem of constructing explicit *normalization map* reduces to constructing small size hitting set of $\overline{\mathsf{VP}}$ [91, 50]. Close connection between certain formulation of derandomization of NNL, and the problem of showing explicit circuit lower bounds is also known [91, 88].

The second motivation comes from the hope to find an explicit 'robust' hitting set for VP [52]; this is a hitting set $\mathcal{H}$ such that after an adequate normalization, there will be a point in $\mathcal{H}$ on which f evaluates to (say) 1. This notion overcomes the discrepancy between a hitting set for VP and a hitting set for $\overline{\mathsf{VP}}$ [52, 87]. We know that small robust hitting set exists [32], but an explicit PSPACE construction was given in [52]. It is not at all clear whether the efficient hitting sets known for restricted depth-3 circuits are robust or not.

Our border PIT results. We continue our study on $\overline{\Sigma^{[k]}\Pi^{[d]}\Sigma}$ and ask for a better than PSPACE constructible hitting set. A polynomial-time hitting set is known for $\Sigma^{[k]}\Pi^{[d]}\Sigma$ [108, 109, 6]. But, the border class seems to be more powerful, and the known hitting sets seem to fail. However, using our structural understanding and the analytic technique, we are able to quasi-derandomize the class completely. For the detailed statement, see Theorem 4.1.

THEOREM 1.2 (Quasi-derandomizing depth-3). *There exists an explicit quasi-polynomial time ($s^{O(\log\log s)}$) hitting set for $\overline{\Sigma^{[k]}\Pi\Sigma}$-circuits of size s and constant k.*

Remarks. 1. For $k = 1$, as $\overline{\Pi\Sigma} = \Pi\Sigma$, there is an explicit polynomial-time hitting set.

2. Our technique *necessarily* blows up the size to $s^{\exp(k)\cdot\log\log s}$. Therefore, it would be interesting to design a *subexponential* time algorithm when $k = \Theta(\log s)$; or poly-time for $k = O(1)$.

3. We can not directly use the de-bordering result of Theorem 1.1 and try to find efficient hitting set, as we do not know explicit good hitting set for general ABPs.

4. One can extend this technique to construct quasi-polynomial time hitting set for depth-4 classes: $\overline{\Sigma^{[k]}\Pi\Sigma\wedge}$ and $\overline{\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}}$, when k and δ are constants. For details, see section 6.

The log*-variate regime.* In recent developments [3, 81, 61, 42] low-variate polynomials, even in highly restricted models, have gained a lot of interest and attention for their general implications in the context of derandomization and hardness results. A slightly *non-trivial* hitting set for trivariate $\Sigma\Pi\Sigma\wedge$-circuits [3, Theorem 4] would in fact give a PIT algorithm for general circuits that runs in quasipolynomial time. With a hardness hypothesis [61, Theorem 1.6] optimizes the algorithm to polynomial time. This motivation has pushed researchers to work on log-variate regime and design efficient PITs. In [49], the authors showed a $\mathsf{poly}(s)$-time black-box identity test for $n = O(\log s)$ variate size-s circuits that have $\mathsf{poly}(s)$-dimensional partial derivative space; for example log-variate depth-3 diagonal circuits. Very recently, Bisht and Saxena [19] gave the first $\mathsf{poly}(s)$-time black-box PIT for sum of constant-many, size-s, $O(\log s)$-variate constant-width ROABPs (and its border).

We remark that non-trivial border-PIT in the low-variate bootstraps to non-trivial PIT for $\overline{\mathsf{VP}}$ as well [3, 61]. That motivates us to derandomize log-variate $\overline{\Sigma^{[k]}\Pi\Sigma}$-circuits. Unfortunately, direct application of Theorem 1.2 fails to give a polynomial-time PIT. Surprisingly, adapting techniques from [49] to extend the existing result (Theorem 4.3), combined with our DiDIL technique, we prove the following. For details, see Theorem 4.4.

Theorem 1.3 (Derandomizing log-variate depth-3). *There exists an explicit* $\mathsf{poly}(s)$*-time hitting set for* $n = O(\log s)$ *variate, size-s,* $\overline{\Sigma^{[k]}\Pi\Sigma}$ *circuits, for constant* k*.*

1.3. Limitation of standard techniques. In this section, we briefly discuss about the standard techniques for both the upper bounds and PITs, in the border sense, and point out why they fail to yield our results.

Why known upper bound techniques fail? One of the most obvious way to de-border restricted classes is to essentially show a polynomial ε-degree bound and interpolate. In general, the bound is known to be exponential [26, Thm. 5.7] which crucially uses [83, Prop. 1]. This proposition essentially shows the existence of an irreducible curve C whose degree is bounded in terms of the degree of the affine variety that we are interested in. The degree is in general exponentially upper bounded by the size [27, Thm. 8.48]. Unless and until one improves these bounds for varieties induced by specific models (which seems hard), one should not expect to improve the ε-degree bound, and thus the interpolation trick seems useless.

As mentioned before, $\overline{\Sigma\wedge\Sigma}$-circuits could be de-bordered using the duality trick [106] (see Lemma 2.16) to make it an $\overline{\text{ARO}}$ and finally using Nisan's characterization giving $\overline{\text{ARO}} = \text{ARO}$ [95, 46, 66] (Lemma 2.22). The trick is directly inapplicable to our model of interest, primarily due to the expected exponential blow in the top fan-in to convert the Π-gate to $\wedge$-gate. We also remark that the duality trick was made *field independent* in [47, Lemma 8.6.4]. In fact, very recently, [20, Theorem 4.3] gave an *improved* duality trick with no size blowup, independent of degree and number of variables.

Due to possibly heavy cancellation of ε-powers, all the non-trivial upper bound methods currently known for border complexity classes seems to not work for $\overline{\Sigma^{[2]}\Pi\Sigma}$ (refer [46, 24]). To elaborate, one of the major bottleneck is that individually limit of T_i as $\varepsilon \to 0$, for $i \in [2]$ *may not* exist, however, $\lim_{\varepsilon\to 0}(T_1 + T_2)$ does exist, where $T_i \in \Pi\Sigma$ (over $\mathbb{F}(\varepsilon)[\boldsymbol{x}]$). For example $T_1 := \varepsilon^{-1}(x + \varepsilon^2 y)y$ and $T_2 := -\varepsilon^{-1}(y + \varepsilon x)x$. No generic tool is available to 'capture' such cancellations, and may even suggest a non-linear algebraic approach to tackle the problem.

Furthermore, [102] explicitly classified certain factor polynomials to solve non-border $\Sigma^{[2]}\Pi\Sigma\wedge$ PIT. This factoring-based idea seems to fail miserably when we study factoring mod $\langle \varepsilon^M \rangle$; in that case, we get non-unique, usually exponentially-many, factorizations. For example $x^2 \equiv (x - a\cdot\varepsilon^{M/2})\cdot(x + a\cdot\varepsilon^{M/2}) \mod \langle\varepsilon^M\rangle$; for all

$a \in \mathbb{F}$. In this case, there are, in fact, infinitely many factorizations. Moreover, $\lim_{\varepsilon\to 0} 1/\varepsilon^M \cdot \left(x^2 - (x - a\cdot\varepsilon^{M/2})\cdot(x + a\cdot\varepsilon^{M/2})\right) = a^2$. Therefore, infinitely many factorizations may give infinitely many limits. To top it all, Kumar's result [78] hinted a possible hardness of border-depth-3 (top fan-in two). In that sense, ours is a very non-linear algebraic proof for restricted models which successfully opens up a possibility of finding non-representation-theoretic, and elementary, upper bounds.

Why known PIT techniques fail? Once we understand $\overline{\Sigma^{[k]}\Pi\Sigma}$, it is natural to look for efficient derandomization. However, as we do not know efficient PIT for ABPs, known techniques would not yield an efficient PIT for the same. Further, in a nutshell—1) limited (almost non-existent) understanding of linear/algebraic dependence under limit, 2) exponential upper bound on ε, and 3) not-good-enough understanding of restricted border classes make it really hard to come up with an efficient hitting set. We elaborate these points below.

Dvir and Shpilka [43] gave a rank-based approach to design the first quasipolynomial time algorithm for $\Sigma^{[k]}\Pi\Sigma$. A series of works [73, 108, 109, 110] finally gave a $s^{O(k)}$-time algorithm for the same. Their techniques depend on either generalizing Chinese remaindering (CR) via ideal-matching or certifying paths, or via efficient variable-reduction, to obtain a good enough rank-bound on the multiplication ($\Pi\Sigma$) terms. Most of these approaches required a linear space, but possibility of exponential ε-powers and non-trivial cancellations make these methods fail miserably in the limit. Similar obstructions also hold for [87, 103, 16] which give efficient hitting sets for the orbit of sparse polynomials (which is in fact *dense* in $\Sigma\Pi\Sigma$). In particular, Medini and Shpilka [87] gave PIT for the orbits of variable disjoint monomials (see [87, Defn. 1.29]), under the affine group, but not the closure of it. Thus, they do not even give a subexponential PIT for $\overline{\Sigma^{[2]}\Pi\Sigma}$.

Recently, Guo [59] gave a s^{δ^k}-time PIT, for non-SG (Sylvester-Gallai) $\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}$ circuits, by constructing explicit variety evasive subspace families; but to apply this idea to border PIT, one has to devise a radical-ideal based PIT idea. Currently, this does not work in the border, as $\varepsilon \mod \langle\varepsilon^M\rangle$ has an exponentially high nilpotency. Since radical$\langle\varepsilon^M\rangle = \langle\varepsilon\rangle$, it 'kills' the necessary information unless we can show a polynomial upper bound on M.

Finally, [6] came up with *faithful* map by using Jacobian + certifying path technique, which is more about algebraic rank rather than linear-rank. However, it is not at all clear how it behaves in the limit as ε goes to zero. For example $f_1 = x_1 + \varepsilon^M \cdot x_2$, and $f_2 = x_1$, where M is arbitrary large. Note that the underlying Jacobian $J(f_1, f_2) = \varepsilon^M$ is nonzero; but it flips to zero in the limit. This makes the whole Jacobian machinery collapse in the border setting; as it cannot possibly give a variable reduction for the border model. (for example one needs to keep both x_1 and x_2 above.)

Very recently, [39] gave a quasipolynomial time hitting set for exact $\Sigma^{[k]}\Pi\Sigma\wedge$ and $\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}$ circuits, when k and δ are constant. This result is dependent on the Jacobian technique which fails under taking limit, as mentioned above. However, a polynomial-time whitebox PIT for $\Sigma^{[k]}\Pi\Sigma\wedge$ circuits was shown using DiDI-technique (Divide, Derive and Induct). This cannot be directly used because there was no ε (i.e. without limit) and $\overline{\Sigma^{[k]}\Pi\Sigma\wedge}$ has only black-box access. Further, Theorem 1.1 gives an ABP, where DiDI-technique cannot be directly applied. Therefore, our DiDIL-technique can be thought of as a *strict* generalization of the DiDI-technique, first introduced in [39], which now applies to uncharted borders.

In a recent breakthrought result, Limaye, Srinivasan and Tavenas [84] showed the first *super*polynomial lower bound for constant-depth circuits. Their lower bound result, together with the 'hardness vs randomness' tradeoff result of [35] gives the first deterministic *subexponential*-time black-box PIT algorithm for general constant-depth circuits. Interestingly, these methods can be adapted in the border setting as well [11]. However, compared to their algorithms, our hitting sets are significantly faster!

1.4. Main tools and a brief road-map. In this section, we sketch the proof of Theorems 1.1-1.3. The proofs are analytic, based on induction on the top fan-in and rely on a common high level picture. They use *logarithmic derivative*, and its power-series expansion; we call the unifying technique as DiDIL (**Di** = Divide, **D**=Derive, **I** = Induct, **L** = Limit). We *essentially* reduce to the well-known 'wedge' models (as fractions, with unbounded top fan-in) and then 'interpolate' it (for Theorem 1.1) or deduce directly about its nonzeroness (Theorem 1.2-1.3).

Basic tools and notations. The analytic tool that we use, appears in algebra (and complexity theory) through the ring of *formal power series* $R[[x_1, \ldots, x_n]]$ (in short $R[[\boldsymbol{x}]]$), see [97, 41, 114]. One of the advantages of the ring $R[[\boldsymbol{x}]]$ emerges from the following *inverse* identity: $(1 - x_1)^{-1} = \sum_{i \geq 0} x_1^i$, which *does not* make sense in $R[x]$, but is available now. Lastly, the logarithmic derivative operator $\mathsf{dlog}_y(f) = (\partial_y f)/f$ plays a very crucial role in 'linearizing' the product gate, since $\mathsf{dlog}_y(f \cdot g) = \partial_y(fg)/(fg) = (f \cdot \partial_y g + g \cdot \partial_y f)/(fg) = \mathsf{dlog}_y(f) + \mathsf{dlog}_y(g)$. Essentially, this operator enables us to use power-series expansion and converts the $\prod$-gate to $\wedge$.

The road-map. The base case when the top fan-in $k = 1$, i.e., we have a single product of affine linear forms, and we are interested in its border. It is not hard to see that the polynomial in the border is also just a product of appropriate affine forms; for details refer to section 3). Now, suppose we have a depth-3 circuit of top fan-in 2, $g(\boldsymbol{x}, \varepsilon) = T_1 + T_2$, where each T_i is a product of affine linear forms. The goal is to somehow reduce this to the case of single summand. Before moving forward, we remark that some ideas described below, directly, can even be formally incorrect! Nonetheless, this sketch is "morally'" correct and, the eventual road-map insinuates the strength of the DiDIL-technique.

For simplicity, let us assume that each linear form has a non-zero constant term (for instance by a random translation of the variables). Moreover, every variable x_i is replaced by $x_i \cdot z$ for a new variable z; this variable z is the 'degree counter' that helps to keep track of the degree of the polynomials involved. Now, dividing both sides by T_1, we get $g/T_1 = 1 + T_2/T_1$, and taking derivatives with respect to the variable z, we get $\partial_z(g/T_1) = \partial_z(T_2/T_1)$. This has reduced the number of summands on the right hand side to 1, although each summand has become more complicated now, and we have no control on what happens as $\varepsilon \to 0$.

Since T_1 is invertible in the power series ring in z, T_2/T_1 is well defined as well. Moreover, $\lim_{\varepsilon \to 0} T_1$ exists (well *not really*, but formally a proper ε-scaling of it does, which suffices since derivarive with respect to z does not affect the ε-scaling!) and is non-zero. From this it follows that after some truncation with respect to high degree z monomials, $\lim_{\varepsilon \to 0} \partial_z(T_2/T_1)$ exists and has a nice relation to the original limit of g; see Claim 3.4!

Lastly, and crucially, $\partial_z(T_2/T_1) \mod z^d = (T_2/T_1) \cdot \mathsf{dlog}(T_2/T_1) \mod z^d$ can be computed by a not-too-complicated circuit structure. Interestingly, the circuit form is *closed* under this operation of dividing, taking derivatives and taking limits! Note that the dlog operator distributes the product gate into summation giving $\mathsf{dlog}(T_2/T_1) = \sum \mathsf{dlog}(\Sigma)$, where Σ denotes linear polynomials, and we observe that $\mathsf{dlog}(\Sigma) = \Sigma/\Sigma \in$

$\Sigma\wedge\Sigma$, the depth-3 powering circuits, over some 'nice' ring. The idea is to expand $1/\ell$, where ℓ is a linear polynomial, as sum of powers of linear terms using the inverse identity:

$$1/(1-a\cdot z) \equiv 1 + a\cdot z + \cdots + a^{d-1}\cdot z^{d-1} \mod z^d .$$

When there is a single remaining summand, the border of the more general structure is easy-to-compute, and can be shown to have an algebraic branching program of not too large size. For details, we refer to Claim 3.6. For a constant k (& even general bounded depth-4 circuits), the above idea can be extended with some additional clever division and computation.

The PIT results also have a similar high level strategy, although there are additional technical difficulties which need some care at every stage. At the core, the idea is really "primal" and depends on the following: If a bivariate polynomial $G(X,Z) \neq 0$, then either its derivative $\partial_Z G(X,Z) \neq 0$, or its constant-term $G(X,0) \neq 0$ (note: $G(X,0) = G \mod Z$). So, if $G(a,0) \neq 0$ or $\partial_Z G(b,Z) \neq 0$, then the union-set $\{a,b\}$ hits $G(X,Z)$, i.e. either $G(a,Z) \neq 0$ or $G(b,Z) \neq 0$.

2. Preliminaries.

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

Notation. We use $[n]$ to denote the set $\{1,\ldots,n\}$, and $\boldsymbol{x} = (x_1,\ldots,x_n)$. For, $\boldsymbol{a} = (a_1,\ldots,a_n), \boldsymbol{b} = (b_1,\ldots,b_n) \in \mathbb{F}^n$, and a variable t, we denote $\boldsymbol{a} + t\cdot\boldsymbol{b} := (a_1 + tb_1, \ldots, a_n + tb_n)$.

We also use $\mathbb{F}[[x]]$, to denote the ring of formal power 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, $\mathbb{F}(\boldsymbol{x})$ denotes the function field, where the elements are of the form f/g, where $f,g \in \mathbb{F}[\boldsymbol{x}]$ $(g \neq 0)$.

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

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

Field. We denote the underlying field as $\mathbb{F}$ and assume that it is of characteristic 0 (for example $\mathbb{Q}, \mathbb{Q}_p$). All our results hold for other fields (for example $\mathbb{F}_{p^e}$) of *large* characteristic p.

Approximative closure. For an algebraic complexity class $\mathcal{C}$, the approximation is defined as follows [24, Def. 2.1].

DEFINITION 2.1 (Approximative closure of a class). *Let $\mathcal{C}_{\mathbb{F}}$ be a class of polynomials defined over a field $\mathbb{F}$. Then, $f(\boldsymbol{x}) \in \mathbb{F}[x_1,\ldots,x_n]$ is said to be in Approximative Closure $\overline{\mathcal{C}}$ if and only if there exists polynomial $Q \in \mathbb{F}[\varepsilon,\boldsymbol{x}]$ such that $g(\boldsymbol{x},\varepsilon) := f(\boldsymbol{x}) + \varepsilon\cdot Q(\boldsymbol{x},\varepsilon)$ is in $\mathcal{C}_{\mathbb{F}(\varepsilon)}$.*

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

Partial Derivative Space of a polynomial f is a vector space formed by considering all possible linear combinations of partial derivatives of f, of all orders. The definition naturally extends to a set of polynomials. Here is an important lemma, originally from [47, Corollary 4.14], which shows that small partial derivative space implies existence of small cone-size monomial. For a detailed proof, we refer [55, Lemma 2.3.15]

THEOREM 2.2 (Cone-size concentration). *Let $\mathbb{F}$ be a field of characteristic 0 or greater than d. Let $\mathcal{P}$ be a set of n-variate d-degree 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 every nonzero $P \in \mathcal{P}$ has a cone-size-k monomial with nonzero coefficient.*

The next lemma shows that there are only few low-cone monomials in a non-zero n-variate polynomial.

LEMMA 2.3 (Counting low-cones, [49, Lemma 5]). *The number of n-variate monomials with cone-size at most k is $O(rk^2)$, where $r := (3n/\log k)^{\log k}$.*

The following lemma can be proved using multi-variate interpolation.

LEMMA 2.4 (Coefficient extraction, [49, Lemma 4]). *Given a circuit C, over the underlying field $\mathbb{F}(\varepsilon)$, and a monomial m, there is a $\mathsf{poly}(\mathsf{size}(C), cs(m), d)$ time algorithm to compute the coefficient of m in C, where $cs(m)$ denotes the cone-size of m.*

2.1. Basics of algebraic complexity. We will give a brief definition of various computational models and tools used in our results. Interested readers can refer [113, 47, 105] for more refined versions.

Algebraic Circuits, defined over a field $\mathbb{F}$, are directed acyclic graphs with a unique root node. The leaf nodes of the graph are labelled by variables or field constants and internal nodes are either labelled with $+$ or $\times$. Further the edges can be labelled by field constants to denote scaler multiplication. The circuit naturally computes the polynomial at the root node from bottom to top. The *size* and *depth* of circuit is the size and depth of the underlying graph.

Circuit size. Some of the complexity parameters of a circuit are *depth* (number of layers), and *fan-in* (maximum number of inputs to a node). *Syntactic degree* of a circuit is defined inductively as follows: Syntactic degree of a leaf is 0 for constants, and 1 for input variables. Syntactic degree of a sum-gate is the maximum of the syntactic degree of its children, moreover, for the product-gate it is the sum of the syntactic degree of its children.

Operation on Complexity Classes. For base classes $\mathcal{C}$ and $\mathcal{D}$ over ring R, a bloated class consists of polynomials from the base classes in any combination of sum, product, and division. For instance, $\mathcal{C}/\mathcal{D} = \{f/g : f \in \mathcal{C}, 0 \neq g \in \mathcal{D}\}$ similarly $\mathcal{C} \cdot \mathcal{D}$ for products, $\mathcal{C} + \mathcal{D}$ for sum, and other possible combinations. The respective computational model for the bloated class is referred to as 'bloated model' in the following text. Also we use $\mathcal{C}_{\mathsf{R}}$ to denote the basic ring R on which $\mathcal{C}$ is being computed over.

Hitting set. A set of points $\mathcal{H} \subseteq \mathbb{F}^n$ is called a *hitting set* for a class $\mathcal{C}$ of n-variate polynomials if for any nonzero polynomial $f \in \mathcal{C}$, there exists a point in $\mathcal{H}$ where f evaluates to a nonzero value. A $T(s)$-time hitting set would mean that the hitting set can be generated in time $\leq T(s)$, for input circuit of size s.

DEFINITION 2.5 (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*

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

where $\mathrm{wt}(\gamma)$ is product of labels over the edges in path γ. The 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 $s \le dw^2\Delta$; in fact, we will take the latter as the ABP-size bound in our calculations.

We remark that ABP is *closed* under both addition and multiplication, which is straightforward from the definition. In fact, we also need to eliminate division in ABPs. Here is an important lemma stated below from [115].

LEMMA 2.6 (Strassen's division elimination). *Let $g(\boldsymbol{x}, y)$ and $h(\boldsymbol{x}, y)$ be computed by ABPs of size s and degree $< d$. Further, assume $h(\boldsymbol{x}, 0) \neq 0$. Then, $g/h \mod y^d$ can be written as $\sum_{i=0}^{d-1} C_i \cdot y^i$, where each C_i is of the form ABP/ABP of size $O(sd^2)$.*

Moreover, in case g/h is a polynomial, then it has an ABP of size $O(sd^2)$.

Proof. ABPs are closed under multiplication, which makes interpolation, with respect to y, possible. Interpolating the coefficient C_i, of y^i, gives a sum of d ABP/ABP's; which can be rewritten as a single ABP/ABP of size $O(sd^2)$.

Next, assume that g/h is a polynomial. For a random $(\boldsymbol{a}, a_0) \in \mathbb{F}^{n+1}$, write $h(\boldsymbol{x} + \boldsymbol{a}, y + a_0) =: h(\boldsymbol{a}, a_0) - \tilde{h}(\boldsymbol{x}, y)$ and define $g' := g(\boldsymbol{x} + \boldsymbol{a}, y + a_0)$. Since $h(\boldsymbol{x}, y)$ is a non-zero polynomial, a random evaluation point such as $(\boldsymbol{a}, a_0)$, guarantees that field element $h(\boldsymbol{a}, a_0) \neq 0$, and $\tilde{h} \in \langle \boldsymbol{x}, y \rangle$. Of course, $\tilde{h}$ has a small ABP. Using the inverse identity in $\mathbb{F}[[\boldsymbol{x}, y]]$, we have $g(\boldsymbol{x} + \boldsymbol{a}, y + a_0)/h(\boldsymbol{x} + \boldsymbol{a}, y + a_0) =$

$$(g'/h(\boldsymbol{a}, a_0))/(1 - \tilde{h}/h(\boldsymbol{a}, a_0)) \equiv (g'/h(\boldsymbol{a}, a_0)) \cdot \left(\sum_{0 \le i < d} (\tilde{h}/h(\boldsymbol{a}, a_0))^i \right) \mod \langle \boldsymbol{x}, y \rangle^d .$$

Note that, the degree blowsup in the above summands to $O(d^2)$ and the ABP-size is $O(sd)$. ABPs are closed under addition/ multiplication; thus, we get an ABP of size $O(sd^2)$ for the polynomial $g(\boldsymbol{x}+\boldsymbol{a}, y+a_0)/h(\boldsymbol{x}+\boldsymbol{a}, y+a_0)$. This implies the ABP-size for g/h as well. ∎

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

DEFINITION 2.7 (Read-once Oblivious Algebraic Branching Program (ROABP)). *An ABP is defined as Read-Once Oblivious Algebraic Branching Program (ROABP) in a variable order $(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for some permutation $\sigma : [n] \to [n]$, if edges of i-th layer of ABP are univariate polynomials in $x_{\sigma(i)}$.*

DEFINITION 2.8 (Any-order ROABP (ARO)). *A polynomial $f \in \mathbb{F}[\boldsymbol{x}]$ 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.*

2.2. Properties of any-order ROABP (ARO). We will start with defining the *partial coefficient space* of a polynomial f to 'characterise' the width of ARO. We can work over any field $\mathbb{F}$.

Let $A(\boldsymbol{x})$ be a polynomial over $\mathbb{F}$ in n variables with individual degree d. Denote the set $M := \{0,\ldots,d\}^n$. Note that, one can write $A(\boldsymbol{x})$ as

$$A(\boldsymbol{x}) \;=\; \sum_{\boldsymbol{\alpha}\in M} \mathsf{coef}_A(\boldsymbol{x}^{\boldsymbol{\alpha}}) \cdot \boldsymbol{x}^{\boldsymbol{\alpha}} \;.$$

Consider a partition of the variables $\boldsymbol{x}$ into two parts $\boldsymbol{y}$ and $\boldsymbol{z}$, with $|\boldsymbol{y}| = k$. Then, $A(\boldsymbol{x})$ can be viewed as a polynomial in variables $\boldsymbol{y}$, where the coefficients are polynomials in $\mathbb{F}[\boldsymbol{z}]$. For monomial $\boldsymbol{y}^{\boldsymbol{a}}$, let us denote the coefficient of $\boldsymbol{y}^{\boldsymbol{a}}$ in $A(\boldsymbol{x})$ by $A_{(\boldsymbol{y},\boldsymbol{a})} \in \mathbb{F}[\boldsymbol{z}]$. The coefficient $A_{(\boldsymbol{y},\boldsymbol{a})}$ can also be expressed as a partial derivative $\partial A/\partial \boldsymbol{y}^{\boldsymbol{a}}$, evaluated at $\boldsymbol{y} = \boldsymbol{0}$ (and multiplied by an appropriate constant), see [51, Section 6]. Moreover, we can also write $A(\boldsymbol{x})$ as

$$A(\boldsymbol{x}) \;=\; \sum_{\boldsymbol{a}\in\{0,\ldots,d\}^k} A_{(\boldsymbol{y},\boldsymbol{a})} \cdot \boldsymbol{y}^{\boldsymbol{a}} \;.$$

One can also capture the space by the coefficient matrix (also known as the partial derivative matrix) where the rows are indexed by monomials p_i from $\boldsymbol{y}$, columns are indexed by monomials q_j from $\boldsymbol{z} = \boldsymbol{x}\backslash\boldsymbol{y}$ and (i,j)-th entry of the matrix is $\mathsf{coef}_{p_i \cdot q_j}(f)$.

The following lemma formalises the connection between ARO width and dimension of the coefficient space (or the rank of the coefficient matrix).

LEMMA 2.9 ([95]). *Let $A(\boldsymbol{x})$ be a polynomial of individual degree d, computed by an* ARO *of width w. Let $k \le n$ and $\boldsymbol{y}$ be any prefix of length k of $\boldsymbol{x}$. Then*

$$\dim_{\mathbb{F}}\{A_{(\boldsymbol{y},\boldsymbol{a})} \;|\; \boldsymbol{a} \in \{0,\ldots,d\}^k\} \;\le\; w \;.$$

We remark that the original statement was for a fixed variable order. Since, ARO affords any-order, the above holds for any-order as well. The following lemma is the converse of the above lemma and shows us that the dimension of the coefficient space is rightly captured by the width.

LEMMA 2.10 (Converse lemma [95]). *Let $A(\boldsymbol{x})$ be a polynomial of individual degree d with $\boldsymbol{x} = (x_1,\ldots,x_n)$, such that for some w, for any $1 \le k \le n$, and $\boldsymbol{y}$, any-order-prefix of length k, we have*

$$\dim_{\mathbb{F}}\{A_{(\boldsymbol{y},\boldsymbol{a})} \;|\; \boldsymbol{a} \in \{0,\ldots,d\}^k\} \;\le\; w \;.$$

Then, there exists an ARO *of width w for $A(\boldsymbol{x})$.*

2.3. Properties of depth-3 diagonal circuits. In this section we will 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 $f, g \in \Sigma\wedge\Sigma$. The following lemma gives us a sum of powers representation of monomial. For proofs see [33, Proposition 4.3].

LEMMA 2.11 (Waring identity for a monomial [33]). *Let $M = x_1^{b_1}\cdots x_k^{b_k}$, where $1 \le b_1 \le \cdots \le b_k$, and* roots of unity *$\mathcal{Z}(i) := \{z \in \mathbb{C} : z^{b_i+1} = 1\}$. Then,*

$$M = \sum_{\varepsilon(i)\in\mathcal{Z}(i):i=2,\cdots,k} \gamma_{\varepsilon(2),\ldots,\varepsilon(k)} \cdot \left(x_1 + \varepsilon(2)x_2 + \ldots + \varepsilon(k)x_k\right)^d \;,$$

where $d := \deg(M) = b_1 + \cdots + b_k$, and $\gamma_{\varepsilon(2),\cdots,\varepsilon(k)}$ are $\prod_{i=2}^k (b_i+1)$ many scalars.

Remark. For fields other than $\mathbb{F} = \mathbb{C}$: We can go to a small extension (at most d^k), for a monomial of degree d, to make sure that $\varepsilon(i)$ exists.

Using this, we show that $\Sigma\wedge\Sigma$ is *closed* under *constant*-fold multiplication.

LEMMA 2.12 ($\Sigma\wedge\Sigma$ closed under multiplication). *Let $f_i \in \mathbb{F}[\boldsymbol{x}]$, of syntactic degree $\le d_i$, 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 $O((d_2+1)\cdots(d_k+1)\cdot s_1 \cdots s_k)$.*

Proof. Let $f_i =: \sum_j \ell_{ij}^{e_{ij}}$; by assumption $e_{ij} \le d_i$. Each summand of $\prod_i f_i$ after expanding can be expressed as $\Sigma\wedge\Sigma$ using Lemma 2.11 of size at most $(d_2+1)\cdots(d_k+1)\cdot\left(\sum_{i\in[k]} \mathsf{size}(\ell_{ij_i})\right)$. Summing up, for all $s_1 \cdots s_k$ many products, gives the upper bound. ◻

Remark. The above lemma, and its proof, hold good for the more general $\Sigma\wedge\Sigma\wedge$ circuits.

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

LEMMA 2.13 ($\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ closed under addition). *Let $f_i \in \mathbb{F}[\boldsymbol{x}]$, of syntactic degree 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 $O((\prod_i d_i)\cdot\prod_i s_i)$.*

Proof. Let $f_i =: u_{i1}/u_{i2}$, where $u_{ij} \in \Sigma\wedge\Sigma$ of size at most s_i. Then

$$f \;=\; \sum_{i\in[k]} f_i \;=\; \left(\sum_{i\in[k]} u_{i1}\prod_{j\ne i} u_{j2}\right) / \left(\prod_{i\in[k]} u_{i2}\right).$$

Use Lemma 2.12 on each product-term in the numerator to obtain $\Sigma\wedge\Sigma$ of size $O((\prod_i d_i)\cdot\prod_i s_i)$. Trivially, $\Sigma\wedge\Sigma$ is closed under addition; so the size of the numerator is $O((\prod_i d_i)\cdot\prod_i s_i)$. Similar argument can be given for the denominator. ◻

Remark. The above holds for $\Sigma\wedge\Sigma\wedge/\Sigma\wedge\Sigma\wedge$ circuits as well.

Using a simple interpolation, the coefficient of y^e can be extracted from $f(\boldsymbol{x}, y) \in \Sigma\wedge\Sigma$ again as a small $\Sigma\wedge\Sigma$ representation.

LEMMA 2.14 ($\Sigma\wedge\Sigma$ coefficient extraction). *Let $f(\boldsymbol{x}, y) \in \mathbb{F}[\boldsymbol{x}][y]$ be computed by a $\Sigma\wedge\Sigma$ circuit of size s and degree d. Then, $\mathsf{coef}_{y^e}(f) \in \mathbb{F}[\boldsymbol{x}]$ is a $\Sigma\wedge\Sigma$ circuit of size $O(sd)$, over $\mathbb{F}[\boldsymbol{x}]$.*

Proof sketch. Let $f =: \sum_i \alpha_i \cdot \ell_i^{e_i}$, with $e_i \le s$ and $\deg_y(f) \le d$. Thus, write $f =: \sum_{i=0}^{d} f_i \cdot y^i$, where $f_i \in \mathbb{F}[\boldsymbol{x}]$. Interpolate using $(d+1)$-many distinct points $y \mapsto \alpha \in \mathbb{F}$, and conclude that f_i has a $\Sigma\wedge\Sigma$ circuit of size $O(sd)$. ◻

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

LEMMA 2.15 ($\Sigma\wedge\Sigma$ differentiation). *Let $f(\boldsymbol{x}, y) \in \mathbb{F}[\boldsymbol{x}][y]$ be computed by a $\Sigma\wedge\Sigma$ circuit of size s and degree d. Then, $\partial_y(f)$ is a $\Sigma\wedge\Sigma$ circuit of size $O(sd^2)$, over $\mathbb{F}[\boldsymbol{x}][y]$.*

Proof sketch. Lemma 2.14 shows that each f_e has $O(sd)$ size circuit where $f =: \sum_e f_e\, y^e$. Doing this for each $e \in [0, d]$ gives a blowup of $O(sd^2)$ and the representation: $\partial_y(f) \;=\; \sum_e f_e \cdot e \cdot y^{e-1}$. ◻

Remark. Same property holds for $\Sigma\wedge\Sigma\wedge$ circuits.

Lastly, we show that $\Sigma\wedge\Sigma$ circuit can be converted into ARO. In fact, we give the proof for a more general model $\Sigma\wedge\Sigma\wedge$. The key ingredient for the lemma is the *duality trick*.

LEMMA 2.16 (Duality trick [106]). *The polynomial $f = (x_1 + \ldots + x_n)^d$ can be* *written as*

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

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

We remark that the above proof works for fields of characteristic $= 0$, or $> d$.

Now, the basic idea is to convert $\wedge\Sigma\wedge$ into $\Sigma\Pi\Sigma^{\{1\}}\wedge$ (i.e. sum-of-product-of-univariates) which is subsumed by ARO [65, Section 2.5.2].

LEMMA 2.17 ($\Sigma\wedge\Sigma\wedge$ as ARO). *Let $f \in \mathbb{F}[\boldsymbol{x}]$ be an n-variate polynomial computable by $\Sigma\wedge\Sigma\wedge$ circuit of size s and syntactic degree D. Then f is computable by an* ARO *of size $O(sn^2D^2)$.*

Proof sketch. Let $g^e = (g_1(x_1) + \cdots + g_n(x_n))^e$, where $\deg(g_i)\cdot e \le D$. Using Lemma 2.16 we get $g^e = \sum_{i=1}^{O(ne)} h_{i1}(x_1)\cdots h_{in}(x_n)$, where each h_{ij} is of degree at most D.

We do this for each power (i.e. each summand of f) individually, to get the final sum-of-product-of-univariates; of top fan-in $O(sne)$ and individual degree at most D. This is an ARO of size $O(sne)\cdot n\cdot D \le O(sn^2D^2)$. ∎

2.4. Basic mathematical tools. For the time-complexity bound, we need to optimize the following function:

LEMMA 2.18. *Let $k \in \mathbb{N}_{\ge 4}$, and $h(x) := x(k-x)7^x$. Then, $\max_{i\in[k-1]} h(i) = h(k-1)$.*

Proof sketch. Differentiate to get $h'(x) = (k-x)7^x - x7^x + x(k-x)(\log 7)7^x = 7^x\cdot[x^2(-\log 7)+x(k\log 7-2)+k]$. It vanishes at $x = \left(\frac{k}{2}-\frac{1}{\log 7}\right)+\sqrt{\left(\frac{k}{2}-\frac{1}{\log 7}\right)^2-\frac{k}{\log 7}}$. Thus, h is maximized at the integer $x = k-1$. ∎

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 2.19 (Valuation). *Let $f \in \mathbb{F}(\boldsymbol{x}, y)$ such that $\mathsf{val}_y(f) \ge 0$. Then, $f \in \mathbb{F}(\boldsymbol{x})[[y]]$*

Proof sketch. Let $f = g/h$ such that $g, h \in \mathbb{F}[\boldsymbol{x}, y]$. Now, $\mathsf{val}_y(f) \ge 0$, implies $\mathsf{val}_y(g) \ge \mathsf{val}_y(h)$. Let $\mathsf{val}_y(g) = d_1$ and $\mathsf{val}_y(h) = d_2$, where $d_1 \ge d_2 \ge 0$. Further, write $g = y^{d_1}\cdot\tilde{g}$ and $h = y^{d_2}\cdot\tilde{h}$. Write, $\tilde{h} = h_0 + h_1\,y + h_2\,y^2 + \cdots + h_d\,y^d$, for some d; with $h_i \in \mathbb{F}[\boldsymbol{x}]$. Note that $h_0 \ne 0$. Thus

$$\begin{aligned} f &= y^{d_1-d_2}\cdot\tilde{g}\cdot\frac{1}{h_0+h_1y+\cdots+h_dy^d} \\ &= \frac{y^{d_1-d_2}\cdot\tilde{g}}{h_0}\cdot\frac{1}{1+(h_1/h_0)y+\cdots+(h_d/h_0)y^d} \;\in\; \mathbb{F}(\boldsymbol{x})[[y]] \end{aligned}$$ ∎

2.5. De-bordering simple models. In this section we will discuss known de-bordering results of restricted models like product of sum of univariates and ARO.

Polynomials approximated by $\Pi\Sigma$ can be easily de-bordered [24, Prop.A.12]. In fact, it is the only constructive de-bordering result known so far. We extend it to show that same holds for polynomials approximated by $\Pi\Sigma\wedge$ circuits. In fact, we start it by showing a much more general theorem.

Let $\mathcal{C}$ and $\mathcal{D}$ be two classes over $\mathbb{F}[\boldsymbol{x}]$. Consider the bloated-class $(\mathcal{C}/\mathcal{C})\cdot(\mathcal{D}/\mathcal{D})$, which has elements of the form $(g_1/g_2)\cdot(h_1/h_2)$, where $g_i \in \mathcal{C}$ and $h_i \in \mathcal{D}$ ($g_2h_2 \ne 0$).

One can also similarly define its border (which will be an element in $\mathbb{F}(\boldsymbol{x})$). Here is an important observation.

Lemma 2.20. $\overline{(\mathcal{C}/\mathcal{C})\cdot(\mathcal{D}/\mathcal{D})} \subseteq (\overline{\mathcal{C}}/\overline{\mathcal{C}})\cdot(\overline{\mathcal{D}}/\overline{\mathcal{D}})$.

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

$$\varepsilon^{a_1-a_2+b_1-b_2}\cdot\left(\frac{\tilde{g}_1}{\tilde{g}_2}\right)\cdot\left(\frac{\tilde{h}_1}{\tilde{h}_2}\right) = f + \varepsilon\cdot Q.$$

Since $\lim_{\varepsilon\to 0}$ exists, the exponent $a_1+b_1-a_2-b_2 \ge 0$. If it is greater than one, then $f = 0$. Moreover, if $a_1+b_1-a_2-b_2 = 0$, then

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

Now, we show an important de-bordering result on $\Pi\Sigma\wedge$ circuits.

Lemma 2.21 (De-bordering $\Pi\Sigma\wedge$). *Consider a polynomial $f \in \mathbb{F}[\boldsymbol{x}]$ which is approximated by $\Pi\Sigma\wedge$ of size s over $\mathbb{F}(\varepsilon)[\boldsymbol{x}]$. Then there exists a $\Pi\Sigma\wedge$ (hence an* ARO*) of size s which exactly computes $f(\boldsymbol{x})$.*

Proof. We will show that $\overline{\Pi\Sigma\wedge} = \Pi\Sigma\wedge \subsetneq$ ARO. From Lemma 2.20, it follows that $\overline{\Pi\Sigma\wedge} \subseteq \prod(\overline{\Sigma\wedge})$. However, we note that $\overline{\Sigma\wedge} = \Sigma\wedge$ and it does not change the size (as it can not increase the sparsity) (refer [24, Prop.A.12]). Therefore, the size does not increase and further it is an ARO. Thus, the conclusion follows. ∎

Next we show that polynomials approximated by ARO can be easily de-bordered. To the best of our knowledge the following lemma was sketched in [46]; also implicitly in [66].

Lemma 2.22 (De-bordering ARO). *Consider a polynomial $f \in \mathbb{F}[\boldsymbol{x}]$ which is approximated by* ARO *of size s over $\mathbb{F}(\varepsilon)[\boldsymbol{x}]$. Then, there exists an* ARO *of size s which exactly computes $f(\boldsymbol{x})$.*

Proof. By definition, there exists a polynomial $g = f + \varepsilon Q$ computable by width w ARO over $\mathbb{F}(\varepsilon)[\boldsymbol{x}]$. Note that $w \le s$. In this proof, we will use the partial derivative matrix. With respect to any-order-prefix $\boldsymbol{y} \subset \boldsymbol{x}$, consider the partial derivative matrix $N(g)$. Using Lemma 2.9 and 2.10, we know $\mathsf{rk}_{\mathbb{F}(\varepsilon)}(N(g)) \le w$. This means determinant of *any* $(w+1)\times(w+1)$ minor of $N(g)$ is identically zero. One can see that the entries of the minor are coefficients of monomials of g which are in $\mathbb{F}[\varepsilon][\boldsymbol{x}\setminus\boldsymbol{y}]$. Thus, determinant polynomial will remain zero even under the limit of $\varepsilon = 0$. Since, $\lim_{\varepsilon\to 0} g = f$, each minor (under limit) captures partial derivative matrix of f of corresponding rows and columns. Thus, we get $\mathsf{rk}_{\mathbb{F}}(N(f)) \le w$. Lemma 2.10 shows that there exists an ARO, of width w over $\mathbb{F}$, which *exactly* computes f. ∎

An obvious consequence of Lemma 2.17 and Lemma 2.22 is the following de-bordering result.

Lemma 2.23 (De-bordering $\Sigma\wedge\Sigma\wedge$). *Consider a polynomial $f \in \mathbb{F}[\boldsymbol{x}]$ which is approximated by $\Sigma\wedge\Sigma\wedge$ of size s over $\mathbb{F}(\varepsilon)[\boldsymbol{x}]$ and syntactic degree D. Then there exists an* ARO *of size $O(sn^2D^2)$ which exactly computes $f(\boldsymbol{x})$.*

2.6. Basic PIT tools. We dedicate this section to discuss some basic PIT tools that we will require in the main section. We will start with the simplest one obtained using PIT lemma of [111, 121, 38, 99].

LEMMA 2.24 (Trivial hitting set). *For a class of n-variate, individual degree $< d$ polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ there exists an explicit hitting set $\mathcal{H} \subseteq \mathbb{F}^n$ of size $d^n + 1$. In other words, there exists a point $\overline{\alpha} \in \mathcal{H}$ such that $f(\overline{\alpha}) \neq 0$ (if $f \neq 0$).*

The above result becomes interesting when $n = O(1)$ as it yields a polynomial-time explicit hitting set. For general n, we have better results for restricted circuits, for example sparse circuits $\Sigma\Pi$, [2, 75] gave a map which reduces multivariate sparse polynomial into univariate polynomial of small degree, while preserving the non-identity. Since testing (low-degree) univariate polynomial is trivial, we get a simple PIT algorithm for sparse polynomials.

Indeed if identity of sparse polynomial can be tested efficiently, product of sparse polynomials $\Pi\Sigma\Pi$ can be tested efficiently. We formalise this in the following lemma.

LEMMA 2.25 ([104, Lemma 2.3]). *For the class of n-variate, degree d polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ computable by $\Pi\Sigma\Pi$ of size s, there exist an explicit hitting set of size $\mathsf{poly}(s, d)$.*

Finally, we state the best known PIT result for ARO, see [66, 60] for more details.

THEOREM 2.26 (ARO hitting set). *For the class of d-degree n-variate polynomials $f \in \mathbb{F}[\boldsymbol{x}]$ computable by size s ARO, there exists an explicit hitting set of size $s^{O(\log \log s)}$.*

The following lemma is useful to construct hitting set for product of two circuit classes when the hitting set of individual circuit is known.

LEMMA 2.27. *Let $\mathcal{H}_1, \mathcal{H}_2 \subseteq \mathbb{F}^n$ of size s_1 and s_2 respectively be the hitting set of the class of n-variate degree d polynomials computable by $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. Then, for the class of polynomials computable by $\mathcal{C}_1 \cdot \mathcal{C}_2$ there is an explicit hitting set $\mathcal{H}$ of size $s_1 \cdot s_2 \cdot O(d)$.*

Proof. Let $f = f_1 \cdot f_2 \in \mathcal{C}_1 \cdot \mathcal{C}_2$ such that $f_1 \in \mathcal{C}_1$ and $f_2 \in \mathcal{C}_2$. For each $\boldsymbol{a}_i \in \mathcal{H}_1$, $\boldsymbol{b}_j \in \mathcal{H}_2$ define a 'formal-sum' evaluation point (over $\mathbb{F}[t]$) $\boldsymbol{c} := (c_\ell)_{1 \leq \ell \leq n}$ such that $c_\ell := a_{i\ell} + t \cdot b_{j\ell}$; where t is a formal variable. Collect these points, going over i, j, in a set H. It can be seen, by shifting and scaling, that non-zeroness is preserved: there exists $\boldsymbol{c} \in H$ such that $0 \neq f(\boldsymbol{c}) \in \mathbb{F}[t]$ and $\deg f(\boldsymbol{c}) = O(d)$. Using trivial hitting set from Lemma 2.24 we obtain the final hitting set $\mathcal{H}$ of size $O(s_1 \cdot s_2 \cdot d)$. ∎

Remark 1. The above argument easily extends to circuit classes $(\mathcal{C}_1/\mathcal{C}_1) \cdot (\mathcal{C}_2/\mathcal{C}_2)$, which compute rationals of the form $(g_1/g_2) \cdot (h_1/h_2)$, where $g_i \in \mathcal{C}_1$ and $h_i \in \mathcal{C}_2$ $(g_2 h_2 \neq 0)$.

Remark 2. The above lemma can be proved alternatively using hitting set generators. These generators are polynomial mapping that certify the non-zeroness of a polynomial by composition. Refer [113, Section 4.1] for detailed discussion.

3. De-bordering depth-3 circuits. In this section we will discuss the proof of de-bordering result (Theorem 1.1). Before moving on, we discuss the bloated model on which we will induct.

DEFINITION 3.1 (Bloated model). *A circuit $\mathcal{C}$ is defined to be in bloated class $\mathsf{Gen}(k, s)$ over the ring of rational functions $\mathsf{R}(\boldsymbol{x})$, with parameter k and size s, if it computes $f \in \mathsf{R}(\boldsymbol{x})$ where $f = \sum_{i \in [k]} T_i$, such that $T_i = (U_i/V_i) \cdot P_i/Q_i$, with $U_i, V_i, P_i, Q_i \in \mathsf{R}[\boldsymbol{x}]$ such that $U_i, V_i \in \Pi\Sigma$ and $P_i, Q_i \in \Sigma\wedge\Sigma$.*

Further, $\mathsf{size}(\mathcal{C}) = \sum_{i\in[k]} \mathsf{size}(T_i)$, *and* $\mathsf{size}(T_i) = \mathsf{size}(U_i) + \mathsf{size}(V_i) + \mathsf{size}(P_i) + \mathsf{size}(Q_i)$.

It is easy to see that size-s $\Sigma^{[k]}\Pi\Sigma$ lies in $\mathsf{Gen}(k,s)$, which will be our general model of induction. Here is the main de-bordering theorem for depth-3 circuits.

THEOREM 3.2 (De-bordering $\overline{\Sigma^{[k]}\Pi\Sigma}$). *Let $f(\boldsymbol{x}) \in \mathbb{F}[x_1,\ldots,x_n]$, such that f can be computed by a $\overline{\Sigma^{[k]}\Pi\Sigma}$-circuit of size s. Then f is also computable by an ABP (over $\mathbb{F}$), of size $s^{O(k\cdot 7^k)}$.*

Proof. We will use DiDIL technique as discussed in subsection 1.4. The $k=1$ case is obvious, as $\overline{\Pi\Sigma} = \Pi\Sigma$ and trivially it has a small ABP. Further, as discussed before, $k=2$ is already non-trivial. Eventually it involves de-bordering $\overline{\mathsf{Gen}(1,s)}$; as DiDIL technique reduces the $k=2$ problem to $\overline{\mathsf{Gen}(1,s)}$ and then we interpolate.

Base step: De-bordering $\overline{\mathsf{Gen}(1,s)}$. Let $g(\boldsymbol{x},\varepsilon) \in R(\boldsymbol{x},\varepsilon)$ be approximating $f \in R(\boldsymbol{x})$; where R is a commutative ring. The specific ring that is needed for the proof to work is defined later in the inductive step. Let d be the maximum of the syntactic degree of the denominator and numerator of the bloated circuit computing g. Here is the de-bordering result.

CLAIM 3.3. $\overline{\mathsf{Gen}(1,s)} \subseteq \mathsf{ABP}/\mathsf{ABP}$, *of size $O(sd^4n)$, while the syntactic degree blows up to $O(nd^2)$.*

Proof. Using Definition 3.1,

$$g(\boldsymbol{x},\varepsilon) =: (U(\boldsymbol{x},\varepsilon)/V(\boldsymbol{x},\varepsilon))\cdot P(\boldsymbol{x},\varepsilon)/Q(\boldsymbol{x},\varepsilon) \;=\; f(\boldsymbol{x}) + \varepsilon\cdot S(\boldsymbol{x},\varepsilon)\;,$$

where $U,V,P,Q \in R(\varepsilon)[\boldsymbol{x}]$ such that $U,V \in \Pi\Sigma$, $P,Q \in \Sigma\wedge\Sigma$. Let $a_1 := \mathsf{val}_\varepsilon(U)$, $a_2 := \mathsf{val}_\varepsilon(V)$, $b_1 := \mathsf{val}_\varepsilon(P)$ and $b_2 := \mathsf{val}_\varepsilon(Q)$. Extracting the maximum ε-power, we get

$$f \,+\, \varepsilon\cdot S \;=\; \varepsilon^{(a_1-a_2)+(b_1-b_2)}\cdot\left(\tilde{U}/\tilde{V}\right)\cdot\left(\tilde{P}/\tilde{Q}\right)\;,$$

where $\widetilde{U},\widetilde{V},\widetilde{P},\widetilde{Q} \in R(\varepsilon)[\boldsymbol{x}]$, and their valuations with respect to. ε are zero i.e. $\lim_{\varepsilon\to 0}\tilde{U}$ exists and is non-zero (similarly for $\tilde{V},\tilde{P},\tilde{Q}$). Since, left side of the equation above is well-defined at $\varepsilon=0$, it must happen that $(a_1-a_2)+(b_1-b_2)\geq 0$. If $(a_1-a_2)+(b_1-b_2)\geq 1$, then $f=0$, and we have trivially de-bordered. Therefore, we can assume $(a_1-a_2)+(b_1-b_2)=0$ which implies that

$$f = (\lim_{\varepsilon\to 0}\tilde{U}/\lim_{\varepsilon\to 0}\tilde{V})\cdot(\lim_{\varepsilon\to 0}\tilde{P}/\lim_{\varepsilon\to 0}\tilde{Q}) \;\in\; (\Pi\Sigma/\Pi\Sigma)\cdot(\text{ARO}/\text{ARO}) \subseteq \mathsf{ABP}/\mathsf{ABP}\;.$$

We have used the fact that $\widetilde{U},\widetilde{V}\in\Pi\Sigma$ and $\widetilde{P},\widetilde{Q}\in\Sigma\wedge\Sigma$ of size at most s, over $R(\varepsilon)[\boldsymbol{x}]$. Further, by Lemma 2.21 and Lemma 2.23, we know that $\overline{\Pi\Sigma}=\Pi\Sigma$ and $\overline{\Sigma\wedge\Sigma}\subseteq \text{ARO}$; therefore f is computable by a ratio of two ABPs of size at most $O(s\cdot d^4 n)$ and the degree gets blown up to atmost $O(nd^2)$. ∎

Bloat out: Reducing $\overline{\Sigma^{[k]}\Pi\Sigma}$ to de-bordering $\overline{\mathsf{Gen}(k-1,\cdot)}$. Let $f_0 := f$ be an arbitrary polynomial in $\overline{\Sigma^{[k]}\Pi\Sigma}$, approximated by $g_0 \in \mathbb{F}(\varepsilon)[\boldsymbol{x}]$, computed by a depth-3 circuit $\overline{C}$ of size s over $\mathbb{F}(\varepsilon)$, i.e. $g_0 := f_0 + \varepsilon\cdot S_0$. Further, assume that $\deg(f_0) < d_0 := d \leq s$; we keep the parameter d separately, to optimize the complexity later. Here, we also stress that one could think of homogeneous circuits and thus the degree can be assumed to be the syntactic degree as well. Then, $g_0 =: \sum_{i\in[k]} T_{i,0}$, such that $T_{i,0}$ is computable by a $\Pi\Sigma$-circuit of size at most s over $\mathbb{F}(\varepsilon)$. Moreover, define $U_{i,0} := T_{i,0}$ and $V_{i,0} := P_{i,0} := Q_{i,0} = 1$ as the base input case (of $\mathsf{Gen}(1,\cdot)$). As explained in the preliminaries, we do a safe division and derivation for reduction.

Φ *homomorphism.* To ensure invertibility and facilitate derivation, we define a homomorphism

$$\Phi : \mathbb{F}(\varepsilon)[\boldsymbol{x}] \to \mathbb{F}(\varepsilon)[\boldsymbol{x}, z]\,, \text{ such that } x_i \mapsto z \cdot x_i + \alpha_i\ ,$$

where α_i are *random* elements in $\mathbb{F}$. Essentially, it suffices to ensure that $\Phi(T_{i,0})|_{\boldsymbol{x}=0} = T_{i,0}(\boldsymbol{\alpha}) \neq 0$ for all $i \in [k]$. We will be working with different ring $\mathcal{R}_i(\boldsymbol{x})$, at i-th step of induction, with $\mathcal{R}_0 := \mathbb{F}[z]/\left\langle z^d \right\rangle$; here think of the z-variable as 'cost-free'. Since Φ is an invertible map, our target is to prove the size upper bound for $\Phi(f_0)$ which is free of mod z^d, and thereby prove upper bound for f_0 by applying the Φ^{-1}.

Divide and derive. Let $v_{i,0} := \mathsf{val}_z(\Phi(T_{i,0}))$. Using the properties of the map we know $v_{i,0} \geq 0$, for each $i \in [k]$. Further, with respect to ε-valuation, assume that $\Phi(T_{i,0}) =: \varepsilon^{a_{i,0}} \cdot \tilde{T}_{i,0}$, where $\tilde{T}_{i,0} =: t_{i,0} + \varepsilon \cdot \tilde{t}_{i,0}(\boldsymbol{x}, z, \varepsilon)$ ($t_{i,0} = \tilde{T}_{i,0}|_{\varepsilon=0}$). Note that, $v_{i,0} = \mathsf{val}_z(\tilde{T}_{i,0})$. With respect to k, we assume $\min_{i\in[k]} \mathsf{val}_z(\tilde{T}_{i,0}) = v_{k,0}$ without loss of generality, else we rearrange the indices to achieve the assumption. Then, we divide $\Phi(g_0)$ by $\tilde{T}_{k,0}$ and derive with respect to z:

$$\Phi(f_0)/\tilde{T}_{k,0} + \varepsilon \cdot \Phi(S_0)/\tilde{T}_{k,0} = \varepsilon^{a_{k,0}} + \sum_{i=1}^{k-1} \Phi(T_{i,0})/\tilde{T}_{k,0} \quad [\textbf{Di}\text{vide}]$$

$$\implies \partial_z \left(\Phi(f_0)/\tilde{T}_{k,0}\right) + \varepsilon \partial_z \left(\Phi(S_0)/\tilde{T}_{k,0}\right) = \sum_{i=1}^{k-1} \partial_z \left(\Phi(T_{i,0})/\tilde{T}_{k,0}\right) \quad [\textbf{D}\text{erive}]$$

$$(3.1) \qquad = \sum_{i=1}^{k-1} \left(\Phi(T_{i,0})/\tilde{T}_{k,0}\right) \cdot \mathsf{dlog}\left(\Phi(T_{i,0})/\tilde{T}_{k,0}\right)$$

$$=: g_1\,.$$

Definability. Let $\mathcal{R}_1 := \mathbb{F}[z]/\langle z^{d_1}\rangle$, and $d_1 := d_0 - v_{k,0} - 1$. For $i \in [k-1]$, define

$$T_{i,1} \,:=\, (\Phi(T_{i,0})/\tilde{T}_{k,0}) \cdot \mathsf{dlog}(\Phi(T_{i,0})/\tilde{T}_{k,0})\,, \text{ and } f_1 \,:=\, \partial_z\left(\Phi(f_0)/t_{k,0}\right)\ .$$

Claim 3.4. *g_1 approximates f_1 correctly, i.e. $\lim_{\varepsilon\to 0} g_1 = f_1$, where g_1 (respectively f_1) are well-defined over $\mathcal{R}_1(\varepsilon, \boldsymbol{x})$ (respectively $\mathcal{R}_1(\boldsymbol{x})$).*

Proof. As we divide by the minimum valuation, by Lemma 2.19 we have

$$\mathsf{val}_z(\Phi(T_{i,0})/\tilde{T}_{k,0}) \geq 0 \implies \Phi(T_{i,0})/\tilde{T}_{k,0} \in \mathbb{F}(\boldsymbol{x}, \varepsilon)[[z]] \implies T_{i,1} \in \mathbb{F}(\boldsymbol{x}, \varepsilon)[[z]]\,.$$

Note that $\mathsf{val}_z(\Phi(f_0) + \varepsilon \cdot \Phi(S_0)) = \mathsf{val}_z(\sum_{i\in[k]} \Phi(T_{i,0})) \geq v_{k,0}$. Setting, $\varepsilon = 0$, implies that $\mathsf{val}_z(\Phi(f_0)) \geq v_{k,0}$ and hence, $\Phi(f_0)/\tilde{T}_{k,0} \in \mathbb{F}(\boldsymbol{x}, \varepsilon)[[z]]$ (by Lemma 2.19). Moreover, $(\Phi(f_0)/\tilde{T}_{k,0})|_{\varepsilon=0} = \Phi(f_0)/t_{k,0} \in \mathbb{F}(\boldsymbol{x})[[z]]$. Combining these it follows that

$$\Phi(f_0)/t_{k,0} \,\in\, \mathbb{F}(\boldsymbol{x})[[z]] \implies f_1 \,\in\, \mathbb{F}(\boldsymbol{x})[[z]]\,.$$

Once we know that each $T_{i,1}$ and f_1 are well-defined power-series, we claim that Eqn. (3.1) holds mod $z^{d_0 - v_{k,0} - 1}$. Note that, $\Phi(f_0) + \varepsilon \cdot \Phi(S_0) = \sum_{i\in[k]} \Phi(T_{i,0})$, holds mod z^d. Thus after dividing by the minimum valuation element (with z-valuation $v_{k,0}$), it holds mod $z^{d_0 - v_{k,0}}$; finally after differentiation it must hold mod $z^{d_0 - v_{k,0} - 1}$.

Further, as $\lim_{\varepsilon\to 0} \tilde{T}_{k,0}$ exists, we must have $\partial_z(\Phi(f_0)/t_{k,0}) = \lim_{\varepsilon\to 0} g_1$; i.e. g_1 approximates f_1 correctly, over $\mathcal{R}_1(\boldsymbol{x})$. ∎

However, we stress that we also think of these as elements over $\mathbb{F}(\boldsymbol{x}, z, \varepsilon)$, with z-degree being 'kept track of' (which could be $> d$). All these different 'lenses' of looking and computing will be important later.

Debordering using reduced fan-in model. To complete the proof we need to show the following – (1) $f_1 \in \overline{\mathsf{Gen}(k-1,\cdot)}$, and (2) assuming we know $\overline{\mathsf{Gen}(k-1,\cdot)}$ has small $\mathsf{ABP}/\mathsf{ABP}$, lift it exactly computes f_0. To prove these claims, we will first show that each $T_{i,1}$ has small $(\Pi\Sigma/\Pi\Sigma)\cdot(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$-circuit over $\mathcal{R}_1(\boldsymbol{x},\varepsilon)$. As for the second part we will interpolate on the bloated model. If the degree of z is carefully controlled, the interpolation would be inexpensive. These two steps are essential in the general reduction as well. Hence we will elaborate on them after showing the fan-in reduction in general.

Inductive step (j-th step): Reducing $\overline{\mathsf{Gen}(k-j,\cdot)}$ to $\overline{\mathsf{Gen}(k-j-1,\cdot)}$. Suppose, we are at the j-th ($j \ge 1$) step. Our induction hypothesis assumes–

1. $\sum_{i\in[k-j]} T_{i,j} =: g_j$, over $\mathcal{R}_j(\boldsymbol{x},\varepsilon)$, such that g_j approximates f_j correctly, where $f_j \in \mathcal{R}_j(\boldsymbol{x})$, where $\mathcal{R}_j := \mathbb{F}[z]/\langle z^{d_j}\rangle$.
2. Here, $T_{i,j} =: (U_{i,j}/V_{i,j})\cdot(P_{i,j}/Q_{i,j})$, where
$$U_{i,j}, V_{i,j} \in \Pi\Sigma \text{ and } P_{i,j}, Q_{i,j} \in \Sigma\wedge\Sigma, \text{ each in } \mathcal{R}_j(\varepsilon)[\boldsymbol{x}]\ .$$
Each can be thought as an element in $\mathbb{F}(\boldsymbol{x},z,\varepsilon)\bigcap\mathbb{F}(\boldsymbol{x},\varepsilon)[[z]]$ as well. Assume that the syntactic degree of each denominator and numerator of $T_{i,j}$ is bounded by D_j.
3. $v_{i,j} := \mathsf{val}_z(T_{i,j}) \ge 0$, for $i \in [k-j]$. Wlog, assume that $\min_i v_{i,j} = v_{k-j,j}$. Moreover, $U_{i,j}|_{z=0} \in \mathbb{F}(\varepsilon)\backslash\{0\}$ (similarly for $V_{i,j}$).

We do like the $j=0$-th step done above, without applying any new homomorphism. Similar to that reduction, we divide and derive to reduce the fan-in further by 1.

Di*vide and* **D***erive.* Let $T_{k-j,j} =: \varepsilon^{a_{k-j,j}}\cdot\tilde{T}_{k-j,j}$, where $\tilde{T}_{k-j,j} =: (t_{k-j,j} + \varepsilon\cdot\tilde{t}_{k-j,j})$ is not divisible by ε. Divide $g_j =: f_j + \varepsilon\cdot S_j$, by $\tilde{T}_{k-j,j}$, to get:

$$f_j/\tilde{T}_{k-j,j} \;+\; \varepsilon\cdot S_j/\tilde{T}_{k-j,j} \;=\; \varepsilon^{a_{k-j,j}} \;+\; \sum_{i=1}^{k-j-1} T_{i,j}/\tilde{T}_{k-j,j}$$

$$\implies \partial_z\left(f_j/\tilde{T}_{k-j,j}\right) \;+\; \varepsilon\cdot\partial_z\left(S_j/\tilde{T}_{k-j,j}\right) \;=\; \sum_{i=1}^{k-j-1}\partial_z\left(T_{i,j}/\tilde{T}_{k-j,j}\right)$$

$$(3.2)\qquad = \sum_{i=1}^{k-j-1}\left(T_{i,j}/\tilde{T}_{k-j,j}\right)\cdot\mathsf{dlog}\left(T_{i,j}/\tilde{T}_{k-j,j}\right)$$

$$=:\; g_{j+1}\,.$$

Definability. Let $\mathcal{R}_{j+1} := \mathbb{F}[z]/\langle z^{d_{j+1}}\rangle$, where $d_{j+1} := d_j - v_{k-j,j} - 1$. For $i \in [k-j-1]$, define

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

Claim 3.5 (**I**nduction hypotheses). *(i) g_{j+1} (respectively f_{j+1}) are well-defined over $\mathcal{R}_{j+1}(\boldsymbol{x},\varepsilon)$ (respectively ,$\mathcal{R}_{j+1}(\boldsymbol{x})$).*

(ii) g_{j+1} approximates f_{j+1} correctly, i.e., $\lim_{\varepsilon\to 0} g_{j+1} = f_{j+1}$.

Proof. Remember, f_j and $T_{i,j}$'s are elements in $\mathbb{F}(\boldsymbol{x}, z, \varepsilon)$ which also belong to $\mathbb{F}(\boldsymbol{x}, \varepsilon)[[z]]$. After dividing by the minimum valuation, by similar argument as in Claim 3.4, it follows that $T_{i,j+1}$ and f_{j+1} are elements in $\mathbb{F}(\boldsymbol{x}, z, \varepsilon) \bigcap \mathbb{F}(\boldsymbol{x}, \varepsilon)[[z]]$, proving the second part of induction-hypothesis-(2). In fact, trivially $v_{i,j+1} \geq 0$, for $i \in [k-j-1]$ proving induction-hypothesis-(3).

Similarly, Eqn. (3.2) holds over $\mathcal{R}_{j+1}(\varepsilon, \boldsymbol{x})$, or equivalently mod $z^{d_{j+1}}$; this is because of the division by z-valuation of $v_{k-j,j}$ and then differentiation, showing induction-hypothesis-(1). So, Eqn. (3.2) being computed mod $z^{d_{j+1}}$ is indeed valid. We also mention that using similar argument as in Claim 3.4, $f_{j+1} \in \mathbb{F}(\boldsymbol{x})[[z]]$.

Finally, as f_{j+1} exists, it is obvious to see that $\lim_{\varepsilon \to 0} g_{j+1} = f_{j+1}$. ∎

Invertibility of $\Pi\Sigma-$*circuits.* In order to prove the second part of induction hypothesis (3) we emphasize the role of dlog and its properties that make the arguments to go through. The action of dlog on $\Sigma\wedge\Sigma$ results in polynomial blow-up in size (Lemma 2.15).

What is the action on $\Pi\Sigma$? As dlog distributes the product *additively*, it suffices to analyse $\mathsf{dlog}(\Sigma)$, and show that $\mathsf{dlog}(\Sigma)$ is in $\Sigma\wedge\Sigma$ with polynomial blow-up in size. Simplifying $T_{i,j+1}$ gives:

$$\begin{aligned} \frac{T_{i,j}}{\tilde{T}_{k-j,j}} &= \varepsilon^{-a_{k-j,j}} \cdot \frac{U_{i,j} \cdot V_{k-j,j}}{V_{i,j} \cdot U_{k-j,j}} \cdot \frac{P_{i,j} \cdot Q_{k-j,j}}{Q_{i,j} \cdot P_{k-j,j}}\,, \\ &= \frac{U_{i,j+1}}{V_{i,j+1}} \cdot \frac{P_{i,j} \cdot Q_{k-j,j}}{Q_{i,j} \cdot P_{k-j,j}} \end{aligned}$$

Where we define $U_{i,j+1} := \varepsilon^{-a_{k-j,j}} \cdot U_{i,j} \cdot V_{k-j,j}$, and $V_{i,j+1} := V_{i,j} \cdot U_{k-j,j}$. Using inductive hypothesis, this directly means:

$$U_{i,j+1}|_{z=0}, V_{i,j+1}|_{z=0} \in \mathbb{F}(\varepsilon) \setminus \{0\}.$$

This proves the second part of induction-hypothesis-(3). The P's and Q's in the equation above will be analysed along with the dlog action on $T_{i,j+1}$ in the upcoming claim.

The overall size blowup. Finally, we show the main step: how to use dlog which is the crux of our reduction. We assume that at the j-th step, $\mathsf{size}(T_{i,j}) \leq s_j$ and by assumption $s_0 \leq s$.

CLAIM 3.6 (Size blowup from DiDIL). *$T_{1,k-1} \in (\Pi\Sigma/\Pi\Sigma)\,(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$ over $\mathcal{R}_{k-1}(\boldsymbol{x}, \varepsilon)$ of size $s^{O(k7^k)}$. It is computed as an element in $\mathbb{F}(\varepsilon, \boldsymbol{x}, z)$, with syntactic degree (in $\boldsymbol{x}, z$) $d^{O(k)}$.*

Proof. Steps $j = 0$ vs $j > 0$ are slightly different because of the homomorphism Φ. However the main idea of using dlog and expand it as a power-series is the same, which eventually shows that $\mathsf{dlog}(\Pi\Sigma)$ is in $\Sigma\wedge\Sigma$ with a controlled blowup.

For $j = 0$, we want to study dlog's effect on $\Phi(T_{i,0})/\tilde{T}_{k,0}$. As dlog distributes over product and thus it suffices to study $\mathsf{dlog}(\ell)$, where $\ell \in \mathcal{R}(\varepsilon)[\boldsymbol{x}]$. However, by the property of Φ, each ℓ must be of the form $\ell = A - zB$, where $A \in \mathbb{F}(\varepsilon)\setminus\{0\}$ and $B \in \mathbb{F}(\varepsilon)[\boldsymbol{x}]$. Using the power series expansion, we have the following, over $\mathcal{R}_1(\boldsymbol{x}, \varepsilon)$:

$$\mathsf{dlog}(\ell) = -\frac{\partial_z\,(A - z \cdot B)}{A\,(1 - z \cdot B/A)} = -\frac{B}{A} \cdot \sum_{j=0}^{d_1-1} \left(\frac{z \cdot B}{A}\right)^j. \tag{3.3}$$

Note,(B/A) and $(-z\cdot B/A)^j$ have a trivial $\wedge\Sigma$ circuits, each of size $O(s)$. For all j use Lemma 2.12 on $(B/A)\cdot(-z\cdot B/A)^j$ to obtain an equivalent $\Sigma\wedge\Sigma$ of size $O(j\cdot d\cdot s)$. Re-indexing gives us the final $\Sigma\wedge\Sigma$ circuit for $\mathsf{dlog}(\ell)$ of size $O(d^3\cdot s)$. We use the fact that $d_1\le d_0=d$. Here the syntactic degree blowsup to $O(d^2)$.

For $j>0$, the above equation holds over $\mathcal{R}_j(\boldsymbol{x})$. However, as mentioned before, the degree could be D_j (possibly $>d_j$) of the corresponding A and B. Thus, the overall size after the power-series expansion would be $O(D_j^2 d\mathsf{size}(\ell))$ [here again we use that $d_j\le d$].

Effect of dlog on $\Sigma\wedge\Sigma$ is, naturally, more straightforward because it is closed under differentiation, as shown in Lemma 2.15. Using Lemma 2.15, we obtain $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ circuit for $\mathsf{dlog}(P_{i,j})$ of size $O\left(D_j^2\cdot s_j\right)$. Similar claim can be made for $\mathsf{dlog}(Q_{i,j})$. Also, $\mathsf{dlog}(U_{i,j}\cdot V_{k-j,j})\in\sum\mathsf{dlog}(\Sigma)$, which could be computed using the above Equation. Thus,

$$\begin{aligned}\mathsf{dlog}(T_{i,j}/\tilde{T}_{k-j,j}) \;&\in\; \mathsf{dlog}(\Pi\Sigma/\Pi\Sigma)\pm\Sigma^{[4]}\mathsf{dlog}(\Sigma\wedge\Sigma)\\ &\subseteq\; \Sigma\wedge\Sigma+\Sigma^{[4]}\Sigma\wedge\Sigma/\Sigma\wedge\Sigma \;=\; \Sigma\wedge\Sigma/\Sigma\wedge\Sigma\,.\end{aligned}$$

Here, $\Sigma^{[4]}$ means sum of 4-many expressions. The first containment is by linearization. Express $\mathsf{dlog}(\Pi\Sigma/\Pi\Sigma)$ as a single $\Sigma\wedge\Sigma$-expression of size $O(D_j^2 d_j s_j)$, by summing up the $\Sigma\wedge\Sigma$-expressions obtained from $\mathsf{dlog}(\Sigma)$. Next, there are 4-many $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ 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 denominator 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 express it as a single $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ expression using Lemma 2.15, with size blowup of $O(D_j^{12}s_j^4)$. Finally, add the single $\Sigma\wedge\Sigma$ expression of size $O(D_j^3 s_j)$, and degree $O(dD_j)$, to get $O(s_j^5 D_j^{16} d)$ size representation.

Also, we need to multiply with $T_{i,j}/\tilde{T}_{k-j,j}$ which is of the form $(\Pi\Sigma/\Pi\Sigma)\cdot(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$, where each $\Sigma\wedge\Sigma$ is basically product of two $\Sigma\wedge\Sigma$ expressions of size s_j and syntantic degree D_j and clubbed together, owing a blowup of $O(D_j s_j^2)$. Hence, multiplying this $(\Pi\Sigma/\Pi\Sigma)\cdot(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$-expression with the $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ expression obtained from dlog-compuation, gives a size blowup of $s_{j+1}:=s_j^7 D_j^{O(1)}d$.

As mentioned before, the main blowup of syntactic degree in the dlog computation could be $O(dD_j)$ and clearing expressions and multiplying the without-dlog expression increases the syntactic degree only by a constant multiple. Therefore, $D_{j+1}:=O(dD_j)\implies D_j=d^{O(j)}$. Hence, $s_{j+1}=s_j^7\cdot d^{O(j)}\implies s_j\le(sd)^{O(j\cdot 7^j)}$. In particular, $s_{k-1}\le s^{O(k\cdot 7^k)}$; here we used that $d\le s$. This calculation quantitatively establishes induction-hypothesis-(2). ∎

Roadmap to trace back f_0. The above claim established that $g_{k-1}\in\mathsf{Gen}(1,\cdot)$ and approximates f_{k-1} correctly. We also know that $\overline{\mathsf{Gen}(1,\cdot)}\in\mathsf{ABP}/\mathsf{ABP}$, from Claim 3.3. Whence, g_{k-1} having $s^{O(k7^k)}$-size bloated-circuit implies: it can be computed as a ratio of ABPs with size $s^{O(k7^k)}\cdot D_{k-1}^4\cdot n=s^{O(k7^k)}$, and syntactic degree $n\cdot D_{k-1}^2=d^{O(k)}$. Now, we recursively 'lift' this quantity, via interpolation, to recover in order, $f_{k-2},f_{k-3},\ldots,f_0$; which we originally wanted.

Interpolation: To integrate and limit. As mentioned above, we will interpolate recursively. We know $f_{k-1}=\partial_z(f_{k-2}/t_{2,k-2})$ has a $\mathsf{ABP}/\mathsf{ABP}$ circuit over $\mathbb{F}(\boldsymbol{x},z)$, i.e. each denominator and numerator is being computed in $\mathbb{F}[\boldsymbol{x},z]$, and size bounded by $\mathcal{S}_{k-1}:=s^{O(k7^k)}$. Here is an important claim about the size of f_{k-2} (we denote it by $\mathcal{S}_{k-2}$).

Claim 3.7 (Tracing back one step). *f_{k-2} can be expressed as*

$$f_{k-2} \ = \sum_{i=0}^{d_{k-2}-1} (\mathsf{ABP}/\mathsf{ABP})\ z^i\ ,$$

of size $s^{O(k7^k)}$ and syntactic degree $d^{O(k)}$.

Proof. Let the degree of both numerator and denominator of f_{k-1} be bounded by $D'_{k-1} := d^{O(k)}$ then we know that it suffices to truncate the power series at $z^{d_{k-1}}$. Further let $e_1, e_2 \leq D'_{k-1}$ be the valuation of f_{k-1} with respect to z. If f_{k-1} is a power series in z, then $e_1 \geq e_2$. The size of the ABPs does not increase after dividing by powers of z, since z and its powers is considered free (equivalent to computing over $\mathbb{F}(z)[\boldsymbol{x}]$). Therefore, $\mathsf{ABP}/\mathsf{ABP}$ can be expressed as $\sum_{i=0}^{d_{k-1}-1} C_{i,k-1} \cdot z^i$, by using the inverse identity: $1/(1-z) \equiv 1 + \ldots + z^{d_{k-1}-1} \bmod z^{d_{k-1}}$. Here, each $C_{i,k-1}$ has an $\mathsf{ABP}/\mathsf{ABP}$ of size at most $O(\mathcal{S}_{k-1} \cdot {D'_{k-1}}^2)$; for details see Lemma 2.6.

Once we get $f_{k-1} = \sum_{i=0}^{d_{k-1}-1} C_{i,k-1} z^i$, definite-integration implies:

$$\frac{f_{k-2}}{t_{2,k-2}} - \left.\frac{f_{k-2}}{t_{2,k-2}}\right|_{z=0} \equiv \sum_{i=1}^{d_{k-1}} \left(\frac{C_{i,k-1}}{i}\right) \cdot z^i \bmod z^{d_{k-1}+1} .$$

The final trick is to get $f_{k-2}/t_{2,k-2}|_{z=0}$ and 'reach' f_{k-2}. As, $f_{k-2}/t_{2,k-2} \in \mathbb{F}(\boldsymbol{x})[[z]]$, substituting $z = 0$ yields an element in $\mathbb{F}(\boldsymbol{x})$. Recall the identity:

$$\begin{aligned} f_{k-2}/t_{2,k-2}|_{z=0} \ &= \lim_{\varepsilon \to 0} (T_{1,k-2}/\tilde{T}_{2,k-2}|_{z=0} + \varepsilon^{a_{2,k-2}}) \\ &\in \lim_{\varepsilon \to 0} \left(\mathbb{F}(\varepsilon) \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma) + \varepsilon^{a_{2,k-2}}\right) . \end{aligned}$$

Since, $\mathbb{F}(\varepsilon) \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma) + \varepsilon^{a_{2,k-2}} \subseteq \Sigma\wedge\Sigma/\Sigma\wedge\Sigma$, over $\mathbb{F}(\varepsilon)(\boldsymbol{x})$. We know that the limit exists and is ARO/ARO ($\subseteq \mathsf{ABP}/\mathsf{ABP}$) of syntactic degree $d^{O(k)}$ and size $s_{k-1} \cdot d^{O(k)}$. Thus, from the above equation, it follows:

$$f_{k-2}/t_{2,k-2} = f_{k-2}/t_{2,k-2}|_{z=0} + \sum_{i=1}^{d_{k-1}} (C_{i,k-1}/i) \cdot z^i \ \in \ \sum_{i=0}^{d_{k-1}} (\mathsf{ABP}/\mathsf{ABP}) \cdot z^i\ ,$$

of size $d_{k-1} \cdot \mathcal{S}_{k-1} D'^2_{k-1} + s_{k-1} \cdot d^{O(k)}$, and degree $D'_{k-1} + d^{O(k)}$. Lastly,

$$t_{2,k-2} \ \in \ \lim_{\varepsilon \to 0} (\Pi\Sigma/\Pi\Sigma) \cdot (\Sigma\wedge\Sigma/\Sigma\wedge\Sigma) \ \subseteq \ (\Pi\Sigma/\Pi\Sigma) \cdot (\text{ARO}/\text{ARO})\ .$$

Thus, it has size s_{k-2}, by previous Claims and degree bound D_{k-2}. Moreover, we know that $\mathsf{val}_z(t_{2,k-2}) \geq v_{2,k-2} = d_{k-2} - d_{k-1} - 1$. Thus, multiply $t_{2,k-2}$ and truncate it till $d_{k-2} - 1$. This gives us the blowup: size $\mathcal{S}_{k-2} = d_{k-1} \cdot \mathcal{S}_{k-1} D'^2_{k-1} + s_{k-1} \cdot d^{O(k)}$ and degree $D'_{k-2} = D'_{k-1} + d^{O(k)}$.

So, we get: f_{k-2} has $\sum_{i=0}^{d_{k-2}-1} (\mathsf{ABP}/\mathsf{ABP}) z^i$ of size $\mathcal{S}_{k-2} = s^{O(k7^k)}$ and degree $D'_{k-2} = d^{O(k)}$. ∎

The $z = 0$-evaluation. To trace back further, we imitate the step as above; and get f_j one by one. But we first need a claim about the $z = 0$ evaluation of $f_j/t_{k-j,j}$.

Claim 3.8 (For definite integration). *$f_j/t_{k-j,j}|_{z=0} \in$ ARO/ARO $\subseteq \mathsf{ABP}/\mathsf{ABP}$ of size $s^{O(k7^k)}$.*

Proof. Note that, $g_j/\tilde{T}_{k-j,j} = \sum_{i\in[k-j]} T_{i,j}/\tilde{T}_{k-j,j} \in \mathbb{F}(\boldsymbol{x})[[z,\varepsilon]]$, as the valuation with respect to z and ε is non-negative. Therefore,

$$\left(\frac{f_j}{t_{k-j,j}}\right)\Big|_{z=0} = \lim_{\varepsilon\to 0}\sum_{i\in[k-j]}\left(\frac{T_{i,j}}{\tilde{T}_{k-j,j}}\right)\Big|_{z=0}$$

$$= \lim_{\varepsilon\to 0}\sum_{i\in[k-j]}\left(\varepsilon^{-a_{k-j,j}}\cdot\frac{U_{i,j}\cdot V_{k-j,j}}{U_{k-j,j}\cdot V_{i,j}}\cdot\frac{P_{i,j}\cdot Q_{k-j,j}}{P_{k-j,j}\cdot Q_{i,j}}\right)\Big|_{z=0}$$

$$\in \lim_{\varepsilon\to 0}\sum_{i\in[k-j]}\left(\mathbb{F}(\varepsilon)\cdot\frac{\Sigma\wedge\Sigma}{\Sigma\wedge\Sigma}\right) = \lim_{\varepsilon\to 0}\left(\frac{\Sigma\wedge\Sigma}{\Sigma\wedge\Sigma}\right) \subseteq \left(\frac{\text{ARO}}{\text{ARO}}\right).$$

Here we crucially used induction-hypothesis-(3) part: each $U_{i,j}, V_{i,j}$ at $z=0$, is an element in $\mathbb{F}(\varepsilon)$. Also, we used that $\Sigma\wedge\Sigma$ is *closed* under constant-fold multiplication (Lemma 2.12). Finally, we take the limit to conclude that $\overline{\Sigma\wedge\Sigma/\Sigma\wedge\Sigma} \subseteq \text{ARO}/\text{ARO}$.

To show the ABP-size upper bound, let us denote the size$(f_j/t_{k-j,j}|_{z=0}) =: S'_j$, and the syntactic degree D'_j. We claim that $S'_j = O(s_j^{O(k-j)}\cdot {D'_j}^4 n)$. Because, we have a sum of $k-j$ many $\Sigma\wedge\Sigma/\Sigma\wedge\Sigma$ expressions each of size s_j; $\Sigma\wedge\Sigma$ is closed under multiplication (Lemma 2.12) and $\Sigma\wedge\Sigma$ to ARO conversion introduces exponent 4 in the degree (Lemma 2.17). Each time the syntactic degree blowup is only a constant multiple, thus $D'_j := d^{O(k)}$ (which is $\le s^{O(k)}$). Therefore, $S'_j = s^{O(k-j)\cdot j7^j} = s^{O(j(k-j)7^j)} = s^{O(k7^k)}$. Here, we use the fact that $\max_{j\in[k-1]} j(k-j)7^j = (k-1)7^{k-1}$ (see Lemma 2.18). This finishes the proof. ∎

Size blowup. Suppose the ABP-size of f_j is $\mathcal{S}_j$; thus we need to estimate $\mathcal{S}_0$. We do not need to eliminate division at each tracing-back-step (which we did to obtain f_{k-2}). Since once we have $\sum_{i=0}^{d_j-1}(\mathsf{ABP}/\mathsf{ABP})\cdot z^i$, it is easy to integrate (with respect to z) without any blowup as we already have all the $\mathsf{ABP}/\mathsf{ABP}$'s in hand (they are z-free). The main size blowup ($= S'_j$) happens due to $z=0$ computation which we calculated above (Claim 3.8). Thus, the final recurrence is $\mathcal{S}_j = \mathcal{S}_{j+1} + S'_j$. This gives $\mathcal{S}_0 = s^{O(k7^k)}$, which is the size of $\Phi(f)$, over $\mathbb{F}(z,\boldsymbol{x})$, being computed as an $\mathsf{ABP}/\mathsf{ABP}$.

Using the degree bound on z, eliminate the division as in the proof of Claim 3.7 to obtain an ε–free ABP over $\mathbb{F}[x,z]$ computing $\Phi(f)$. Apply the map Φ^{-1} to obtain the final ABP of size $s^{O(k7^k)}$ computing the polynomial f. ∎

Remark. In general, we proved that if $f \in \overline{\mathsf{Gen}(k,s)}$, then it can be computed by an ABP of size $s^{O(k7^k)}$.

4. Black-box PIT for border depth-3 circuits. We divide the section into two parts. First subsection deals with proving Theorem 1.2, while the second subsection deals with a better hitting sets in the log-variate regime.

4.1. Quasi-derandomizing $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuits. Integration step of DiDIL is important to give any meaningful upper bound of circuit complexity. However, a hitting set construction demands less—each inductive step of fan-in reduction only needs to preserve non-zeroness. Eventually, we exploit this to give an efficient hitting set construction for $\overline{\Sigma^{[k]}\Pi\Sigma}$, and in the process of reducing the top fan-in analyse the bloated model $\mathsf{Gen}(k,\cdot)$.

THEOREM 4.1 (Hitting set for $\overline{\Sigma^{[k]}\Pi\Sigma}$). *There exists an explicit $s^{O(k\cdot 7^k\cdot\log\log s)}$ time hitting set for $\overline{\Sigma^{[k]}\Pi\Sigma}$-circuits of size s. For constant k, the algorithm runs in quasi-polynomial time.*

Proof. The basic fan-in reduction strategy is same as in section 3. Let $f_0 := f$ be an arbitrary polynomial in $\overline{\Sigma^{[k]}\Pi\Sigma}$, approximated by $g_0 \in \mathbb{F}(\varepsilon)[\boldsymbol{x}]$, computed by a depth-3 circuit $\overline{C}$ of size s over $\mathbb{F}(\varepsilon)$, i.e. $g_0 := f_0 + \varepsilon \cdot S_0$. Further, assume that $\deg(f_0) < d_0 := d \le s$. Let $g_0 =: \sum_{i\in[k]} T_{i,0}$, such that $T_{i,0}$ is computable by a $\Pi\Sigma$-circuit of size at most s over $\mathbb{F}(\varepsilon)$. As before, define $\mathcal{R}_0 := \mathbb{F}[z]/\langle z^d\rangle$. Thus, $f_0 + \varepsilon \cdot S_0 = \sum_{i\in[k]} T_{i,0}$, holds over $\mathcal{R}_0(\boldsymbol{x}, \varepsilon)$.

Define $U_{i,0} := T_{i,0}$ and $V_{i,0} := P_{i,0} := Q_{i,0} = 1$ to set the input instance of $\mathsf{Gen}(k, s)$. Of course, we assume that each $T_{i,0} \neq 0$ (otherwise it is a smaller fan-in than k).

The homomorphism Φ*.* To ensure invertiblity and facilitate derivation, we define the same Φ as in section 3, i.e. $\Phi : \mathbb{F}(\varepsilon)[\boldsymbol{x}] \to \mathbb{F}(\varepsilon)[\boldsymbol{x}, z]$ such that $x_i \mapsto z \cdot x_i + \alpha_i$. For the upper bound proof, we took $\alpha_i \in \mathbb{F}$ to be random; but for the PIT purpose, we cannot work with a random shift. The purpose of shifting was to ensure the invertibility, i.e., $\mathbb{F}(\varepsilon) \ni T_{i,0}(\boldsymbol{\alpha}) \neq 0$; that is easy to ensure since $\ell(y, y^2, \ldots, y^n) \neq 0$, for any linear polynomial ℓ, over any field. Since, $\deg(\prod_i T_{i,0}) \le s$, there exists an $i \in [s]$ such that $\boldsymbol{\alpha} = (i, i^2, \ldots, i^n)$ hits $T_{i,0}$! In the proof, we will work with every such $\boldsymbol{\alpha}$ (s-many), and for the right value, non-zeroness will be preserved, which suffices.

0-th step: Reduction from k to $k-1$. We will use the same notation as in section 3. We know that g_1 approximates f_1 correctly over $\mathcal{R}_1(\boldsymbol{x}, \varepsilon)$. Rewriting the same, we have

(4.1)
$$f_0 + \varepsilon \cdot S_0 \;=\; \sum_{i\in[k]} T_{i,0}\,, \ \text{ over } \mathcal{R}_0(\boldsymbol{x}, \varepsilon) \implies f_1 + \varepsilon \cdot S_1 \;=\; \sum_{i\in[k-1]} T_{i,1}\,, \ \text{ over } \mathcal{R}_1(\boldsymbol{x}, \varepsilon)\,.$$

Here, define $T_{i,1} := (\Phi(T_{i,0})/\tilde{T}_{k,0}) \cdot \mathsf{dlog}(\Phi(T_{i,0})/\tilde{T}_{k,0})$, for $i \in [k-1]$ and $f_1 := \partial_z\,(\Phi(f_0)/t_{k,0})$, same as before. Also, we will consider $T_{i,1}$ as an element of $\mathbb{F}(\boldsymbol{x}, z, \varepsilon)$ and keep track of $\deg(z)$.

The "iff" condition. Note that the equality in (4.1) over $\mathcal{R}_1(\varepsilon, \boldsymbol{x})$ is only "one-sided". Whereas, to reduce the problem of identity testing to smaller fan-in case, we need a necessary and sufficient condition: If $f_0 \neq 0$, we *would like* to claim that $f_1 \neq 0$ (over $\mathcal{R}_1(\boldsymbol{x})$). However, it may not be directly true because of the loss of z-free terms of f_0, due to differentiation. Note that $f_1 \neq 0$ implies $\mathsf{val}_z(f_1) < d =: d_1$. Further, $f_1 = 0$, over $R_1(\boldsymbol{x})$, implies–

1. Either $\Phi(f_0)/t_{k,0}$ is z-free. This implies $\Phi(f_0)/t_{k,0} \in \mathbb{F}(\boldsymbol{x})$, which further implies it is in $\mathbb{F}$, because z-free implies $\boldsymbol{x}$-free, by substituting $z = 0$, by the definition of Φ. Also, note that $f_0, t_{k,0} \neq 0$ implies $\Phi(f_0)/t_{k,0}$ is a *nonzero* element in $\mathbb{F}$. Thus, it suffices to check whether $\Phi(f_0)|_{z=0} = f_0(\boldsymbol{\alpha})$ is non-zero or not.
2. Or $\partial_z(\Phi(f_0)/t_{k,0}) = z^{d_1} \cdot p$ where $p \in \mathbb{F}(z, \boldsymbol{x})$ s.t. $\mathsf{val}_z(p) \ge 0$. By simple power series expansion, one can conclude that $p \in \mathbb{F}(\boldsymbol{x})[[z]]$ (Lemma 2.19). Hence,
$$\Phi(f_0)/t_{k,0} = z^{d_1+1} \cdot \tilde{p}, \text{where}\, \tilde{p} \in \mathbb{F}(\boldsymbol{x})[[z]] \implies \mathsf{val}_z(\Phi(f_0)) \ge d\,,$$
a contradiction. Here we used the simple fact that differentiation decreases the valuation by 1.

Conversely, it is obvious that $f_0 = 0$ implies $f_1 = 0$. Thus, we have proved the following:

$$f_0 \neq 0 \text{ over } \mathbb{F}[\boldsymbol{x}] \iff f_1 \neq 0 \text{ over } \mathcal{R}_1(\boldsymbol{x}), \quad \text{or} \quad 0 \neq \Phi(f_0)|_{z=0} \in \mathbb{F}.$$

Recall, Claim 3.6 shows that $T_{i,1} \in (\Pi\Sigma/\Pi\Sigma)\,(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$ with a polynomial blowup. Therefore, subject to $z = 0$ test, we have reduced the identity testing problem to $k-1$. We will recurse over this until we reach $k = 1$.

Induction step. Assume that we are at the end of j-th step ($j \geq 1$). Our inductive hypothesis assumes the following invariants:

1. $\sum_{i\in[k-j]} T_{i,j} = f_j + \varepsilon \cdot S_j$ over $\mathcal{R}_j(\varepsilon, \boldsymbol{x})$, where $T_{i,j} \neq 0$ and $\mathcal{R}_j := \mathbb{F}[z]/\langle z^{d_j}\rangle$.
2. Each $T_{i,j} = (U_{i,j}/V_{i,j})\cdot(P_{i,j}/Q_{i,j})$ where $U_{i,j}, V_{i,j} \in \Pi\Sigma$ and $P_{i,j}, Q_{i,j} \in \Sigma\wedge\Sigma$.
3. $\mathsf{val}_z(T_{i,j}) \geq 0$, for all $i \in [k-j]$. Moreover, $U_{i,j}|_{z=0} \in \mathbb{F}(\varepsilon)\backslash\{0\}$ (similarly $V_{i,j}$).
4. $f_0 \neq 0$ iff: $f_j \neq 0$ over $\mathcal{R}_j(\boldsymbol{x})$, or there exists $1 \leq i \leq j-1$ such that $f_i/t_{k-i,i}|_{z=0} \neq 0$, over $\mathbb{F}(\boldsymbol{x})$

Reducing the problem to $k-j-1$. We will follow the $j = 0$ case, *without* applying any homomorphism. Again, this reduction step is exactly the same as before, which yields: $f_j + \varepsilon \cdot S_j = \sum_{i\in[k-j]} T_{i,j}$, over $\mathcal{R}_j(\boldsymbol{x}, \varepsilon) \implies$

$$f_{j+1} + \varepsilon \cdot S_{j+1} = \sum_{i\in[k-j-1]} T_{i,j+1}, \text{ over } \mathcal{R}_{j+1}(\boldsymbol{x}, \varepsilon). \tag{4.2}$$

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

It remains to show that, all the invariants assumed are still satisfied for $j+1$. The first 3 invariants are already shown in section 3. The 4-th invariant is the iff condition to be shown below.

The "iff" condition in the induction. The above (4.2) reduces $k-j$-summands to $k-j-1$. But we want an 'iff' condition to efficiently reduce the identity testing. If $f_{j+1} \neq 0$, then $\mathsf{val}_z(f_{j+1}) < d_{j+1}$. Further, $f_{j+1} = 0$, over $R_{j+1}(\boldsymbol{x})$ implies–

1. Either $f_j/t_{k-j,j}$ is z-free, i.e. $f_j/t_{k-j,j} \in \mathbb{F}(\boldsymbol{x})$. Now, if indeed $f_0 \neq 0$, then $t_{k-j,j}$ as well as f_j *must be non-zero* over $\mathbb{F}(z, \boldsymbol{x})$, by induction hypothesis (assuming they are non-zero over $\mathcal{R}_j(\boldsymbol{x})$). We will eventually show that $f_j/t_{k-j,j}|_{z=0}$ has a small ARO/ARO circuit; which helps us to construct a quasi-polynomial size hitting set using Theorem 2.26.
2. Or $\partial_z(f_j/t_{k-j,j}) = z^{d_{j+1}} \cdot p$, where $p \in \mathbb{F}(z, \boldsymbol{x})$ s.t. $\mathsf{val}_z(p) \geq 0$. By simple power series expansion, one concludes that $p \in \mathbb{F}(\boldsymbol{x})[[z]]$ (Lemma 2.19). Hence,

$$\frac{f_j}{t_{k-j,j}} \in z^{d_{j+1}+1} \cdot \tilde{p}, \text{ where } \tilde{p} \in \mathbb{F}(\boldsymbol{x})[[z]] \implies \mathsf{val}_z(f_j) \geq d_j$$
$$\implies f_j = 0\,, \text{ over } \mathcal{R}_j(\boldsymbol{x}).$$

Conversely, $f_j = 0$, over $\mathcal{R}_j(\boldsymbol{x})$, implies $\mathsf{val}_z(f_j/\tilde{T}_{k-j,j}) \geq d_j - v_{k-j,j} \implies \mathsf{val}_z(\partial_z(f_j/\tilde{T}_{k-j,j})) \geq d_j - v_{k-j,j} - 1 = d_{j+1} \implies \partial_z(f_j/\tilde{T}_{k-j,j}) = 0$, over $\mathcal{R}_{j+1}(\varepsilon, \boldsymbol{x})$. Fixing $\varepsilon = 0$ we deduce $f_{j+1} = \partial_z(f_j/t_{k-j,j}) = 0$.

Thus, we have proved that $f_j \neq 0$ over $\mathcal{R}_j(\boldsymbol{x})$ iff

$$f_{j+1} \neq 0 \text{ over } R_{j+1}(\boldsymbol{x})\,, \text{ or}\,,\ 0 \neq (f_j/t_{k-j,j})|_{z=0} \in \mathbb{F}(\boldsymbol{x})\,.$$

This concludes the proof of the 4-th invariant.

Note: In the expression above $f_j/t_{k-j,j}$ may be undefined at $z = 0$. However, we keep track of z–degree to show that it is bounded in both numerator and denominator, as in Claim 3.6 . Later when we show that $(f_j/t_{k-j,j})|_{z=0} \in \mathsf{ABP}/\mathsf{ABP}$, we use the degree bound to interpolate and cancel out z–power to get a ratio which is well-defined at $z = 0$.

Constructing the hitting set. The above discussion has reduced the problem of testing $\Phi(f)$ to testing f_{k-1} or $f_j/t_{k-j,j}|_{z=0}$, for $j \in [k-2]$. We know that $f_{k-1} \in (\Pi\Sigma/\Pi\Sigma) \cdot (\mathrm{ARO}/\mathrm{ARO})$, of size $s^{O(k7^k)}$, from Claim 3.6. We obtain the hitting set of $\Pi\Sigma$ from Lemma 2.25, and for $\Sigma\wedge\Sigma$ we obtain the hitting set from Theorem 2.26 (due to Lemma 2.17). Finally we combine the two hitting sets using Lemma 2.27 and use the fact that the syntactic degree is bounded by $s^{O(k)}$ to obtain a hitting set $\mathcal{H}_{k-1}$ of size $s^{O(k7^k \log\log s)}$.

However, it remains to show– (1) efficient hitting set for $f_j/t_{k-j,j}|_{z=0}$, for $j \in [k-2]$, and most importantly (2) how to translate these hitting sets to that of $\Phi(f)$.

Recall: Claim 3.8 shows that $f_k/t_{k-j,j}|_{z=0} \in \mathrm{ARO}/\mathrm{ARO}$, of size $s^{O(k7^k)}$ (over $\mathbb{F}(\boldsymbol{x})$). Thus, it has a hitting set $\mathcal{H}_j$ of size $s^{O(k7^k \log\log s)}$, for all $j \in [k-2]$ (Theorem 2.26).

To translate the hitting set, we need a small property which will bridge the gap of lifting the hitting set to f_0.

CLAIM 4.2 (Fix $\boldsymbol{x}$). *For $\boldsymbol{b} \in \mathbb{F}^n$, if the following two things hold: (i) $f_{j+1}|_{\boldsymbol{x}=\boldsymbol{b}} \neq 0$, over $\mathcal{R}_{j+1}$, and (ii) $\mathsf{val}_z(\tilde{T}_{k-j,j}|_{\boldsymbol{x}=\boldsymbol{b}}) = v_{k-j,j}$, then $f_j|_{\boldsymbol{x}=\boldsymbol{b}} \neq 0$, over $\mathcal{R}_j$.*

Proof. Suppose the hypothesis holds, and $f_j|_{\boldsymbol{x}=\boldsymbol{b}} = 0$, over $\mathcal{R}_j$. Then,

$$\mathsf{val}_z\left(\left(\frac{f_j}{\tilde{T}_{k-j,j}}\right)\Bigg|_{\boldsymbol{x}=\boldsymbol{b}}\right) \geq d_j - v_{k-j,j} \implies \mathsf{val}_z(\partial_z\left(\left(\frac{f_j}{\tilde{T}_{k-j,j}}\right)\Bigg|_{\boldsymbol{x}=\boldsymbol{b}}\right) \geq d_{j+1}.$$

The last condition implies that $\partial_z(f_j/\tilde{T}_{k-j,j})|_{\boldsymbol{x}=\boldsymbol{b}} = 0$, over $\mathcal{R}_{j+1}(\boldsymbol{x})$. Fixing $\varepsilon = 0$ we deduce $f_{j+1}|_{\boldsymbol{x}=\boldsymbol{b}} = 0$. This is a contradiction! □

Finally, we have already shown in section 3 that $\tilde{T}_{k-j,j} \in (\Pi\Sigma/\Pi\Sigma)\cdot(\Sigma\wedge\Sigma/\Sigma\wedge\Sigma)$, and $t_{k-j,j} \in (\Pi\Sigma/\Pi\Sigma) \cdot (\mathrm{ARO}/\mathrm{ARO})$, of size $s^{O(k7^k)}$, which is similar to f_{k-1}.

Joining the dots: The final hitting set. We now have all the ingredients to construct the hitting set for $\Phi(f_0)$. We know $\mathcal{H}_{k-1}$ works for f_{k-1} (as well as $t_{2,k-2}$, because they both are of the same size and belong to $(\Pi\Sigma/\Pi\Sigma) \cdot (\mathrm{ARO}/\mathrm{ARO})$). This lifts to f_{k-2}. But from the 4-th invariant, we know that $\mathcal{H}_{k-2}$ works for the $z = 0$ part. Eventually, lifting this using Claim 4.2, the final hitting set (in $\boldsymbol{x}$) will be $\mathcal{H} := \bigcup_{j\in[k-1]} \mathcal{H}_j$. We remark that we do not need extra hitting set for each $t_{k-j,j}$, because it is already covered by $\mathcal{H}_{k-1}$. We have also kept track of $\deg(z)$ which is bounded by $s^{O(k)}$. We use a trivial hitting set for z which does not change the size. Thus, we have successfully constructed a $s^{O(k7^k \log\log s)}$-time hitting set for $\overline{\Sigma^{[k]}\Pi\Sigma}$. □

Remark. The set $\mathcal{H}$ constructed is a $s^{O(k7^k \log\log s)}$-time hitting set for $\overline{\mathsf{Gen}(k,s)}$, over fields of large characteristic.

4.2. Border PIT for log-variate depth-3 circuits.

In this section, we prove Theorem 1.3. This proof is dependent on adapting and extending proof of Forbes, Ghosh, and Saxena [49], by showing that there is a $\mathsf{poly}(s)$-time hitting set for log-variate $\overline{\Sigma\wedge\Sigma}$-circuits.

THEOREM 4.3 (Derandomizing log-variate $\overline{\Sigma\wedge\Sigma}$). *There is a $\mathsf{poly}(s)$-time hitting set for $n = O(\log s)$ variate $\overline{\Sigma\wedge\Sigma}$-circuits of size s.*

Proof sketch. Let $g = f + \varepsilon \cdot Q$, such that $g \in \Sigma\wedge\Sigma$, over $\mathbb{F}(\varepsilon)$, approximates $f \in \overline{\Sigma\wedge\Sigma}$. The idea is the same as [49]— (1) dimension of the space generated by all partial derivatives of f is $\mathsf{poly}(s,d)$, (2) low partial derivative space implies low cone-size monomials, (3) we can extract low cone-size monomials efficiently, (4) number of low cone-size monomials is at most $\mathsf{poly}(sd)$-many.

We remark that (2) is direct from [47, Corollary 4.14] (with origins in [50]); see Theorem 2.2. (4) is also directly taken from [49, Lemma 5] once we assume (1); for the full statement we refer to Lemma 2.3.

To show (1), we know that g has $\mathsf{poly}(s,d)$–dimensional partial-derivative space over $\mathbb{F}(\varepsilon)$. Denote

$$V_\varepsilon := \left\langle \frac{\partial g}{\partial \boldsymbol{x}^{\boldsymbol{a}}} \mid \boldsymbol{a} < \infty \right\rangle_{\mathbb{F}(\varepsilon)}, \text{ and } V := \left\langle \frac{\partial f}{\partial \boldsymbol{x}^{\boldsymbol{a}}} \mid \boldsymbol{a} < \infty \right\rangle_{\mathbb{F}}.$$

Consider the matrix M_ε, where we index the rows by $\partial_{\boldsymbol{x}^{\boldsymbol{a}}}$, while columns are indexed by monomials in the support of g, and the entries are the value of partial derivative operator. Suppose, $\dim(V_\varepsilon) =: r \le \mathsf{poly}(s,d)$ (because g has a size–s $\Sigma\wedge\Sigma$ circuit). That means, all $(r+1)$ polynomials $\frac{\partial g}{\partial \boldsymbol{x}^{\boldsymbol{a}}}$ are linearly dependent. In other words, the determinant of any $(r+1) \times (r+1)$ minor of M_ε is 0. Note that $\lim_{\varepsilon \to 0} M_\varepsilon = M$, the corresponding partial-derivative matrix for f. Crucially, the zeroness of the determinant of any $(r+1) \times (r+1)$ minor of M_ε translates to the corresponding $(r+1) \times (r+1)$ submatrix of M as well (one can also think of det as a "continuous" function, yielding this property). In particular, $\dim(V) \le r \le \mathsf{poly}(s,d)$.

Finally, to show (3), we note that the coefficient extraction lemma [49, Lemma 4] also holds over $\mathbb{F}(\varepsilon)$. Thus, given the circuit of g, we can decide whether the coefficient of $m =: \boldsymbol{x}^{\boldsymbol{a}}$ is zero or not, in $\mathsf{poly}(\mathrm{cs}(m), s, d)$-time; see Lemma 2.4. Note: the coefficient is an arbitrary element in $\mathbb{F}(\varepsilon)$; however we are only interested in its non-zeroness, which is merely 'unit-cost' for us.

We only extract monomials with cone-size $\mathsf{poly}(s,d)$ (property (2)) and there are only $\mathsf{poly}(s,d)$ many such monomials. Therefore, we have a $\mathsf{poly}(s)$-time hitting set for $\overline{\Sigma\wedge\Sigma}$. ∎

Once we have Theorem 4.3, we argue that this polynomial-time hitting set can be used to give a poly-time hitting set for $\overline{\Sigma^{[k]}\Pi\Sigma}$. We restate Theorem 1.3 with proper complexity below.

THEOREM 4.4 (Efficient hitting set for log-variate $\overline{\Sigma^{[k]}\Pi\Sigma}$). *There exists an explicit $s^{O(k7^k)}$-time hitting set for $n = O(\log s)$ variate, size-s, $\overline{\Sigma^{[k]}\Pi\Sigma}$ circuits.*

Proof sketch. We proceed similarly as in subsection 4.1, with same notations. The reduction and branching out (or conditions) remains exactly the same; in the end, we get that $f_{k-1} \in (\Pi\Sigma/\Pi\Sigma) \cdot (\mathrm{ARO}/\mathrm{ARO})$. Crucially, observe that this ARO is not a generic poly-sized ARO; these AROs are de-bordered log-variate $\overline{\Sigma\wedge\Sigma}$ circuits. From Theorem 4.3, we know that there is a $s^{O(k7^k)}$-time hitting set (because of the size blowup, as seen in section 3). Combining this hitting set with $\Pi\Sigma$-hitting set is easy, by Lemma 2.27.

Moreover, $t_{k-j,j}$ are also of the form $(\Pi\Sigma/\Pi\Sigma) \cdot (\mathrm{ARO}/\mathrm{ARO})$, where again these AROs are de-bordered log-variate $\overline{\Sigma\wedge\Sigma}$ circuits and $s^{O(k7^k)}$-time hitting set exists. Therefore, take the union of the hitting sets (as before), each of size $s^{O(k7^k)}$. This gives the final hitting set which is again $s^{O(k7^k)}$-time constructible. ∎

5. Gentle leap into depth-4: De-bordering $\overline{\Sigma^{[k]}\Pi\Sigma\wedge}$ circuits. The main content of this section is to sketch the de-bordering theorem for $\overline{\Sigma^{[k]}\Pi\Sigma\wedge}$. We intend to extend DiDIL and induct on a slightly more general bloated model, as sketched in subsection 1.4.

THEOREM 5.1 ($\overline{\Sigma^{[k]}\Pi\Sigma\wedge}$ upper bound). *Let $f(\boldsymbol{x}) \in \mathbb{F}[x_1, \ldots, x_n]$, such that f can be computed by a $\overline{\Sigma^{[k]}\Pi\Sigma\wedge}$-circuit of size s. Then f is also computable by an ABP (over $\mathbb{F}$), of size $s^{O(k \cdot 7^k)}$.*

Proof sketch. We will go through the proof of Theorem 3.2 (see section 3), while reusing the notations, and point out the important changes for the DiDIL technique to work on this more general bloated-model $(\Pi\Sigma\wedge/\Pi\Sigma\wedge) \cdot (\Sigma\wedge\Sigma\wedge/\Sigma\wedge\Sigma\wedge)$. As earlier, we induct on the top fan-in parameter k.

Base case. The analysis remains unchanged. We merely have to de-border $\Pi\Sigma\wedge$ and $\Sigma\wedge\Sigma\wedge$ for the numerator and the denominator separately using Lemma 2.21 and Lemma 2.23. Then use the product lemma (Lemma 2.20) to conclude:

$$\overline{(\Pi\Sigma\wedge/\Pi\Sigma\wedge) \cdot (\Sigma\wedge\Sigma\wedge/\Sigma\wedge\Sigma\wedge)} \;\subseteq\; (\Pi\Sigma\wedge/\Pi\Sigma\wedge) \cdot (\mathrm{ARO}/\mathrm{ARO}) \;\subseteq\; \mathsf{ABP}/\mathsf{ABP}\,.$$

Reducing the problem to $k-1$. To facilitate DiDIL, we use the same $\Phi : \mathbb{F}(\varepsilon)[\boldsymbol{x}] \longrightarrow \mathbb{F}(\varepsilon)[\boldsymbol{x}, z]$; since α_i are *random*, the bottom $\Sigma\wedge$ circuits are 'invertible' $(\bmod\ z^d)$. For the same reasons as Theorem 3.2, it suffices to upper bound the size of $\Phi(f)$.

We will apply again divide and derive to reduce the fan-in step by step. We just need to understand $T_{i,j}$. Similar to Claim 3.6, we claim the following.

CLAIM 5.2. *$T_{1,k-1} \in \frac{\Pi\Sigma\wedge}{\Pi\Sigma\wedge} \cdot \frac{\Sigma\wedge\Sigma\wedge}{\Sigma\wedge\Sigma\wedge}$, an element in the ring $\mathcal{R}_{k-1}(\boldsymbol{x}, \varepsilon)$, of size at most $s^{O(k7^k)}$.*

Proof. The main part is to show that dlog acts on $\Pi\Sigma\wedge$ circuits "well". To elaborate, we note that (3.3) can be written for $\Sigma\wedge$ circuits, giving a $\Sigma\wedge\Sigma\wedge$ circuit. To elaborate, let $A - z \cdot B =: h \in \Sigma\wedge$, such that $0 \neq A \in \mathbb{F}(\varepsilon)$. Therefore, over $\mathcal{R}_1(\boldsymbol{x})$, we have

$$\mathsf{dlog}(h) = -\frac{\partial_z\,(z \cdot B)}{A\,(1 - z \cdot B/A)} = -\frac{\partial_z\,(z \cdot B)}{A} \cdot \sum_{j=0}^{d_1-1} \left(\frac{z \cdot B}{A}\right)^j\,.$$

Once we use the fact that $\Sigma\wedge\Sigma\wedge$ is closed under multiplication (Lemma 2.12), it readily follows that $\mathsf{dlog}(\Pi\Sigma\wedge) \in \Sigma\wedge\Sigma\wedge$. Moreover, the derivative of $\Sigma\wedge\Sigma\wedge$ is again a $\Sigma\wedge\Sigma\wedge$ circuit, due to easy interpolation (Lemma 2.15). Following the same proof arguments (as for Theorem 3.2), we can establish the above claim.

It was already remarked that properties shown in subsection 2.3 hold for $\Sigma\wedge\Sigma\wedge$ circuits as well. Therefore, the rest of the calculations remain unchanged, and the size claim holds. ∎

Interpolation & Definite integration. It is again not hard to see that

$$f_j/t_{k-j,j}|_{z=0} \in \lim_{\varepsilon \to 0} \sum_{i \in [k-j]} \mathbb{F}(\varepsilon) \cdot (\Sigma\wedge\Sigma\wedge/\Sigma\wedge\Sigma\wedge) \;\subseteq\; \mathrm{ARO}/\mathrm{ARO} \;\subseteq\; \mathsf{ABP}/\mathsf{ABP}\,.$$

Here, we have used the facts that $\Sigma\wedge\Sigma\wedge$ is closed under multiplication (Lemma 2.12) and $\overline{\Sigma\wedge\Sigma\wedge} \subseteq \mathrm{ARO}$ (Lemma 2.23). The remaining steps also follow similarly once we have the $\mathsf{ABP}/\mathsf{ABP}$ form of de-bordered expressions.

We remark that in all the steps the size and degree claims remain the same and hence the final size of the circuit for $\Phi(f)$ immediately follows. ∎

6. Black-box PIT for border depth-4 circuits. The DiDIL-paradigm that works for depth-3 circuits can be used to give hitting set for border depth-4 $\overline{\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}}$ and $\overline{\Sigma^{[k]}\Pi\Sigma\wedge}$ circuits. But before that, we have to argue that we have efficient hitting set for the wedge model $\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$, which we discuss in the next subsection. Later, we will sketch the proof of the hitting set for border of bounded depth-4 circuits.

6.1. Efficient hitting set for $\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$. Forbes [48] gave quasipolynomial-time black-box PIT for $\Sigma\wedge\Sigma\Pi^{[\delta]}$; using a *rank*-based method. We will make some small observations to extend the same for $\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$ as well. We encourage interested readers to refer to [48] for details. First, we need some definitions and properties.

Shifted Partial Derivative measure $\boldsymbol{x}^{\leq \ell}\boldsymbol{\partial}_{\leq m}$ is a linear operator first introduced in [72, 63] as:

$$\boldsymbol{x}^{\leq \ell}\boldsymbol{\partial}_{\leq m}(g) \;:=\; \{\boldsymbol{x}^{\boldsymbol{c}}\partial_{\boldsymbol{x}^{\boldsymbol{b}}}\,(g)\}_{\deg \boldsymbol{x}^{\boldsymbol{c}}\leq \ell, \deg \boldsymbol{x}^{\boldsymbol{b}}\leq m}\;.$$

It was shown in [48] that the rank of shifted partial derivatives of a polynomial computed by $\Sigma\wedge\Sigma\Pi^{[\delta]}$ is small. We state the result formally in the next lemma. Consider the fractional field $\mathcal{R} := \mathbb{F}(\varepsilon)$.

Lemma 6.1 (Measure upper bound). *Let $g(\varepsilon,\boldsymbol{x}) \in \mathcal{R}[x_1,\ldots,x_n]$ be computable by $\Sigma\wedge\Sigma\Pi^{[\delta]}$ circuit of size s. Then*

$$\mathsf{rkspan}\boldsymbol{x}^{\leq \ell}\boldsymbol{\partial}_{\leq m}(g) \leq s\cdot m\cdot \binom{n+(\delta-1)m+\ell}{(\delta-1)m+\ell}.$$

Further it was observed in [48] that, the rank can be lower bounded using the *Trailing Monomial* (ref [37, Section 2]). Under any *monomial ordering*, the trailing monomial of g denoted by $\mathsf{TM}(g)$ is the smallest monomial in the set $\mathsf{support}(g) := \{\boldsymbol{x}^{\boldsymbol{a}} : \mathsf{coef}_{\boldsymbol{x}^{\boldsymbol{a}}}(g) \neq 0\}$.

Proposition 6.2 (Measure the trailing monomial). *Consider $g \in \mathcal{R}[\boldsymbol{x}]$. For any $\ell, m \geq 0$,*

$$\mathsf{rkspan}\boldsymbol{x}^{\leq \ell}\boldsymbol{\partial}_{\leq m}(g) \geq \mathsf{rkspan}\boldsymbol{x}^{\leq \ell}\boldsymbol{\partial}_{\leq m}\left(\mathsf{TM}(g)\right).$$

For fields of characteristic zero, a lower bound on a monomial was obtained.

Lemma 6.3 (Monomial lower bound). *Consider a monomial $\boldsymbol{x}^{\boldsymbol{a}} \in \mathcal{R}[x_1,\ldots,x_n]$. Then,*

$$\mathsf{rkspan}\left(\boldsymbol{x}^{\leq \ell}\boldsymbol{\partial}_{\leq m}\left(\boldsymbol{x}^{\boldsymbol{a}}\right)\right) \geq \binom{\eta}{m}\binom{\eta-m+\ell}{\ell}$$

where $\eta := |\mathsf{support}\left(\boldsymbol{x}^{\boldsymbol{a}}\right)|$.

In [48] the above results were combined to show that the trailing monomial of polynomials computed by $\Sigma\wedge\Sigma\Pi^{[\delta]}$ circuits have logarithmically small support size. Using the same idea we show that if such a polynomial approximates f, then the support of $\mathsf{TM}(f)$ is also small. We formalize this in the next lemma.

Lemma 6.4 (Trailing monomial support). *Let $g(\varepsilon,\boldsymbol{x}) \in \mathcal{R}[x_1,\ldots,x_n]$ be computable by a $\Sigma\wedge\Sigma\Pi^{[\delta]}$ circuit of size s such that $g = f + \varepsilon\cdot Q$ where $f \in \mathbb{F}[\boldsymbol{x}]$ and $Q \in \mathbb{F}[\varepsilon,\boldsymbol{x}]$. Let $\eta := |\mathsf{support}(\mathsf{TM}(f))|$. Then $\eta = O(\delta\log s)$.*

Proof. Let $\boldsymbol{x}^{\boldsymbol{a}} := \mathsf{TM}(f)$ and $S := \{i \,|\, a_i \neq 0\}$. Define a substitution map ρ such that $x_i \to y_i$ for $i \in S$ and $x_i \to 0$ for $i \notin S$. It is easy to observe that $\mathsf{TM}(\rho(f)) = \rho(\mathsf{TM}(f)) = \boldsymbol{y}^{\boldsymbol{a}}$. Using Lemma 6.1 we know:

$$\mathsf{rk}_{\mathcal{R}}\boldsymbol{y}^{\leq \ell}\boldsymbol{\partial}_{\leq m}(\rho(g)) \;\leq\; s\cdot m\cdot \binom{\eta+(\delta-1)m+\ell}{(\delta-1)m+\ell} =: R\,.$$

To obtain the upper bound for $\rho(f)$ we use the following claim.

CLAIM 6.5. $\mathsf{rk}_{\mathbb{F}} \boldsymbol{y}^{\leq \ell} \boldsymbol{\partial}_{\leq m}(\rho(f)) \leq R$.

Proof. Define the *coefficient matrix* $N(\rho(g))$ with respect to $\boldsymbol{y}^{\leq \ell} \boldsymbol{\partial}_{\leq m}(\rho(g))$ as follows: the rows are indexed by the operators $\boldsymbol{y}^{=\ell_i}\, \partial_{\boldsymbol{y}^{=m_i}}$, while the columns are indexed by the terms present in $\rho(g)$; and the entries are the respective operator-action on the respective term in $\rho(g)$. Note that $\mathsf{rk}_{\mathbb{F}(\varepsilon)} N(\rho(g)) \leq R$. Similarly define $N(\rho(f))$ with respect to $\boldsymbol{y}^{\leq \ell} \boldsymbol{\partial}_{\leq m}(\rho(f))$, then it suffices to show that $\mathsf{rk}_{\mathbb{F}} N(\rho(f)) \leq R$.

For any $r > R$, let $\mathcal{N}(\rho(g))$ be a $r \times r$ sub-matrix of $N(\rho(g))$. The rank bound ensures: $\det \mathcal{N}(\rho(g)) = 0$. This will remain true under the limit $\varepsilon = 0$; thus, $\det(\mathcal{N}(\rho(f))) = 0$.
Since $r > R$ was arbitrary and linear dependence is preserved, we deduce:

$$\mathsf{rk}_{\mathbb{F}} N(\rho(f)) \leq R\,.$$

For lower bound, recall $\boldsymbol{y}^{\boldsymbol{a}} = \mathsf{TM}(\rho(f))$. Then, by Proposition 6.2 and Lemma 6.3, we get:

$$\mathsf{rk}_{\mathbb{F}} \boldsymbol{y}^{\leq \ell} \boldsymbol{\partial}_{\leq m}(\rho(f)) \geq \binom{\eta}{m}\binom{\eta - m + \ell}{\ell}. \tag{6.1}$$

Comparing Claim 6.5 and (6.1) we get:

$$s \geq \frac{1}{m} \cdot \binom{\eta}{m} \cdot \binom{\eta - m + \ell}{\ell} \Big/ \binom{\eta + (\delta-1)m + \ell}{(\delta-1)m + \ell}.$$

For $\ell := (\delta-1)(\eta+(\delta-1)m)$ and $m := \lfloor n/e^3\delta \rfloor$, [48, Lem.A.6] showed $\eta \leq O(\delta \log s)$. ∎

The existence of a small support monomial in a polynomial which is being approximated, is a structural result which will help in constructing a hitting set for this larger class. The idea is to use a map that reduces the number of variables to the size of the support of the trailing monomial, and then invoke Lemma 2.24.

THEOREM 6.6 (Hitting set for $\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$). *For the class of n-variate, degree d polynomials* approximated *by $\Sigma\wedge\Sigma\Pi^{[\delta]}$ circuits of size s, there is an explicit hitting set $\mathcal{H} \subseteq \mathbb{F}^n$ of size $s^{O(\delta \log s)}$ i.e., for every such nonzero polynomial f there exists an $\boldsymbol{\alpha} \in \mathcal{H}$ for which $f(\boldsymbol{\alpha}) \neq 0$.*

Proof. Let $g(\varepsilon, \boldsymbol{x}) \in \mathcal{R}[x_1, \ldots, x_n]$ be computable by a $\Sigma\wedge\Sigma\Pi^{[\delta]}$ circuit of size s such that $g =: f + \varepsilon \cdot Q$, where $f \in \mathbb{F}[\boldsymbol{x}]$ and $Q \in \mathbb{F}[\varepsilon, \boldsymbol{x}]$. Then Lemma 6.4 shows that there exists a monomial $\boldsymbol{x}^{\boldsymbol{a}}$ of f such that $\eta := |\mathsf{support}(\boldsymbol{x}^{\boldsymbol{a}})| = O(\delta \log s)$.

Let $S \in \binom{[n]}{\eta}$. Define a substitution map ρ_S such that $x_i \to y_i$ for $i \in S$ and $x_i \to 0$ for $i \notin S$. Note that, under this substitution non-zeroness of f is preserved for some S; because monomials of support $S \supseteq \mathsf{support}(\boldsymbol{x}^{\boldsymbol{a}})$ will survive for instance. Essentially $\rho_S(f)$ is an η-variate degree-d polynomial, for which Lemma 2.24 gives a trivial hitting set of size $O(d^\eta)$. Therefore, with respect to S we get a hitting set $\mathcal{H}_S$ of size $O(d^\eta)$. To finish, we do this for all such S, to obtain the final hitting set $\mathcal{H}$ of size:

$$\binom{n}{\eta} \cdot O\left(d^\eta\right) \;\leq\; O((nd)^\eta)\,.$$ ∎

Remark 6.7. Unlike the PIT result for the border of depth 3 circuits, we obtained this result without de-bordering the circuit at all.

6.2. DiDIL on depth-4 models. The DiDIL-paradigm along with the branching idea, in subsection 4.1, can be used to give hitting set for border depth-4 $\overline{\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}}$ and $\overline{\Sigma^{[k]}\Pi\Sigma\wedge}$ circuits. For brevity, we denote these two types of (non-border) depth-4 circuits by $\Sigma^{[k]}\Pi\Sigma\Upsilon$ circuits where $\Upsilon \in \{\wedge, \Pi^{[\delta]}\}$. We will give a separate hitting set for the border of each class, while analysing them together.

THEOREM 6.8 (Hitting set for bounded border depth-4). *There exists an explicit $s^{O(k\cdot 7^k \cdot \log\log s)}$ (respectively $s^{O(\delta^2 k 7^k \log s)}$)-time hitting set for $\overline{\Sigma^{[k]}\Pi\Sigma\wedge}$ (respectively $\overline{\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}}$)-circuits of size s.*

Proof sketch. We will again follow the same notation as subsection 4.1. Let $g_0 := \sum_{i\in[k]} T_{i,0} = f_0 + \varepsilon S_0$ such that g_0 is computable by $\Sigma^{[k]}\Pi\Sigma\Upsilon$ over $\mathbb{F}(\varepsilon)$. As earlier, we will instead work with a bloated model that preserves the structure when applying the DiDIL technique. The bloated model we consider is

$$\Sigma^{[k]}\left(\Pi\Sigma\Upsilon/\Pi\Sigma\Upsilon\right)\left(\Sigma\wedge\Sigma\Upsilon/\Sigma\wedge\Sigma\Upsilon\right) .$$

Using the hitting set of product of sparse polynomials (refer [75]), we can obtain a point $\boldsymbol{\alpha} = (a_1, \ldots, a_n) \in \mathbb{F}(\varepsilon)^n$ such that $\Pi\Sigma\Upsilon$ evaluated at $\boldsymbol{\alpha}$ is non-zero. This evaluation point helps in maintaining its invertibility. We capture the non-zeroness in a 1-1 invertible homomorphism $\Phi : \mathbb{F}(\varepsilon)[\boldsymbol{x}] \to \mathbb{F}(\varepsilon)[\boldsymbol{x}, z]$ such that $x_i \to z \cdot x_i + \alpha_i$. The invertibility of the map implies: $f_0 \neq 0$ if and only if $\Phi(f_0) \neq 0$.

The next steps are essentially the same: reduce k to the bloated $k-1$, and inductively to the bloated $k = 1$ case. There will be 'branches' and for each branch we will give efficient hitting sets; taking their union will give the final hitting set.

By **Di**vide and **D**erive, we will eventually show that: $f_0 \neq 0 \iff f_{k-1} \neq 0$ over $\mathcal{R}_j(\boldsymbol{x})$, or there exists $1 \leq i \leq k-2$ such that $(f_i/t_{k-i,i}|_{z=0} \neq 0$, over $\mathbb{F}(\boldsymbol{x}))$. Similar to Claim 5.2 we can show that

$$T_{1,k-1} \in \left(\Pi\Sigma\Upsilon/\Pi\Sigma\Upsilon\right)\left(\Sigma\wedge\Sigma\Upsilon/\Sigma\wedge\Sigma\Upsilon\right),$$

over $\mathcal{R}_{k-1}(\boldsymbol{x}, \varepsilon)$. The trick is again to use dlog and show that $\mathsf{dlog}(\Pi\Sigma\Upsilon) \in \Sigma\wedge\Sigma\Upsilon$. However the size blowup behaves slightly differently. To prove it formally, we need the following claim that upper bounds the blow-up from applying the map ψ on $\Sigma\Pi^{[\delta]}$.

CLAIM 6.9. *Let $g \in \Sigma\Pi^{[\delta]}$, then $\Psi(g) \in \Sigma\Pi^{[\delta]}$ of size at most $3^\delta \cdot \mathsf{size}(g)$, when number of variables $n \gg \delta$.*

Proof sketch. Let $\boldsymbol{x}^{\boldsymbol{a}}$ be a monomial of degree δ, such that $\sum_i a_i \leq \delta$. Then the number of monomials produced by Ψ can be upper bounded by the AM-GM inequality:

$$\prod_i (a_i + 1) \leq \left(\frac{\sum_i a_i + n}{n}\right)^n \leq (1 + \delta/n)^n$$

As $\delta/n \to 0$, we have $(1 + \delta/n)^n \to e^\delta$. As $e < 3$, the upper bound follows. ☐

We claim that $T_{1,k-1}$ is in the bloated model with reasonable blowup in size.

CLAIM 6.10. *For $\Sigma^{[k]}\Pi\Sigma\wedge$, respectively $\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}$, we have*

$$T_{1,k-1} \in \left(\frac{\Pi\Sigma\wedge}{\Pi\Sigma\wedge}\right) \cdot \left(\frac{\Sigma\wedge\Sigma\wedge}{\Sigma\wedge\Sigma\wedge}\right) \textit{ respectively } \left(\frac{\Pi\Sigma\Pi^{[\delta]}}{\Pi\Sigma\Pi^{[\delta]}}\right) \cdot \left(\frac{\Sigma\wedge\Sigma\Pi^{[\delta]}}{\Sigma\wedge\Sigma\Pi^{[\delta]}}\right),$$

over $\mathcal{R}_{k-1}(\boldsymbol{x}, \varepsilon)$ of size $s^{O(k7^k)}$ respectively $(s3^\delta)^{O(k7^k)}$.

Proof sketch. We will follow the line of arguments from the proof of Claim 5.2 and explain it for one step i.e. over $\mathcal{R}_1(\boldsymbol{x},\varepsilon)$. After applying the map, let $A - z\cdot B = h \in \Sigma\Upsilon$, such that $A \in \mathbb{F}(\varepsilon)$. Therefore, over $\mathcal{R}_1(\boldsymbol{x})$, we have

$$\mathsf{dlog}(h) = -\frac{\partial_z\,(z\cdot B)}{A\,(1 - z\cdot B/A)} = -\frac{B}{A}\cdot\sum_{j=0}^{d_1-1}\left(\frac{z\cdot B}{A}\right)^j \,.$$

Here, use the fact that $\Sigma\wedge\Sigma\Upsilon$ is closed under multiplication. For $\Sigma\wedge\Sigma\wedge$ circuits, the calculations remains the same as in section 5. However, for $\Sigma\wedge\Sigma\Pi^{[\delta]}$ circuits, note that as h is shifted, $\mathrm{size}(B)$ is no longer $\mathsf{poly}(s)$; but it is at most $3^\delta\cdot s$, see Claim 6.9. Therefore, the claim follows. ◻

Eventually, one can show (using Lemma 2.20 to distribute):

$$f_{k-1} \in \overline{(\Pi\Sigma\Upsilon/\Pi\Sigma\Upsilon)\cdot(\Sigma\wedge\Sigma\Upsilon/\Sigma\wedge\Sigma\Upsilon)} \subseteq (\Pi\Sigma\Upsilon/\Pi\Sigma\Upsilon)\cdot\overline{(\Sigma\wedge\Sigma\Upsilon/\Sigma\wedge\Sigma\Upsilon)}\,.$$

When $\Upsilon = \wedge$, we know $\overline{\Sigma\wedge\Sigma\wedge} \subseteq$ ARO and thus this has a hitting set of size $s^{O(k7^k\log\log s)}$ (Theorem 2.26). We also know hitting set for $\Pi\Sigma\wedge$ (Lemma 2.25). Combining them using Lemma 2.27, we have a quasipolynomial-time hitting set of size $s^{O(k7^k\log\log s)}$.

As seen before, we also need to understand the evaluation at $z = 0$. By a similar argument, it will follow that

$$f_j/t_{k-j,j}|_{z=0} \in \lim_{\varepsilon\to 0}\sum_{i\in[k-j]}\mathbb{F}(\varepsilon)\cdot(\Sigma\wedge\Sigma\Upsilon/\Sigma\wedge\Sigma\Upsilon) \subseteq \overline{\Sigma\wedge\Sigma\Upsilon}\,.$$

When $\Upsilon = \wedge$, we can de-border and this can be shown to be an ARO. Thus, in that case $f_j/t_{k-j,j}|_{z=0} \in \mathrm{ARO}/\mathrm{ARO}$, where hitting set is known (similarly as before) giving hitting set for each additional check in each step. Once we have hitting set for each step, we can take a union (similar to Claim 4.2) to finally give the desired hitting set.

Unfortunately, we do not know the size complexity upper bound of $\overline{\Sigma\wedge\Sigma\Upsilon}$, when $\Upsilon = \Pi^{[\delta]}$, as the duality trick cannot be directly applied. However, as we know a hitting set for $\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$, from Theorem 6.6; we will use it to get the final hitting set. To see why this works, note that we need to *hit* $f_{k-1} \in \left(\Pi\Sigma\Pi^{[\delta]}/\Pi\Sigma\Pi^{[\delta]}\right)\cdot \overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}/\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$. We know hitting sets for both $\Pi\Sigma\Pi^{[\delta]}$ (Lemma 2.25) and $\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$ (Theorem 6.6), thus combining them is easy using Lemma 2.27.

To get the final estimate, define $s' := s^{O(\delta k 7^k)}$; which signifies the size blow-up due to DiDIL. Next, the hitting set $\mathcal{H}_{k-1}$ for f_{k-1} has size $(nd)^{O(\delta\log s')} \leq s^{O(\delta^2 k 7^k\log s)}$. We know that a similar bound also holds for each branch. Taking their union gives the final hitting set of the size as claimed. ◻

7. Conclusion & future direction. This work introduces the DiDIL-technique and successfully de-borders as well as derandomizes PIT for $\overline{\Sigma^{[k]}\Pi\Sigma}$. Further we extend this to subclasses of depth-4 as well. This opens a variety of questions which would enrich border-complexity theory.

1. Does $\overline{\Sigma^{[k]}\Pi\Sigma} \subseteq \Sigma\Pi\Sigma$, or $\overline{\Sigma^{[k]}\Pi\Sigma} \subseteq \mathsf{VF}$, i.e. does it have small formulas?
2. Can we show that $\mathsf{VBP} \neq \overline{\Sigma^{[k]}\Pi\Sigma}$? [1]

[1] Very recently, Dutta and Saxena [40] showed an exponential gap between the two classes.

3. Can we improve the current hitting set of $s^{\exp(k)\cdot\log\log s}$ to $s^{O(\mathsf{poly}(k)\cdot\log\log s)}$, or even a $\mathsf{poly}(s)$-time hitting set? The current technique seems to blow-up the exponent.
4. Can we de-border $\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$, or $\overline{\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}}$, for constant k and δ? Note that we already have partially derandomized PIT for the class (Theorem 6.8).
5. Can we show that $\overline{\Sigma^{[k]}\wedge\Sigma}\subseteq\Sigma\wedge\Sigma$ for constant k? To show that polynomials of constant border-Waring rank have waring rank which is polynomially bounded by the degree and the number of variables.
6. Can we de-border $\overline{\Sigma^{[2]}\Pi\Sigma\wedge^{[2]}}$? i.e. the bottom layer has bi-variate polynomials.

De-bordering vs. Derandomization. In this work, we have successfully de-bordered and (quasi)-derandomized $\overline{\Sigma^{[k]}\Pi\Sigma}$. Here, we remark that de-bordering did not directly give us a hitting set, since the de-bordering result was more general than the models for which explicit hitting sets are known. However, we were still able to do it because of the DiDIL-technique. Moreover, while extending this to depth-4, we could quasi-derandomize $\overline{\Sigma^{[k]}\Pi\Sigma\Pi^{[\delta]}}$, because eventually hitting set for $\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$ is known. However we could not de-border $\overline{\Sigma\wedge\Sigma\Pi^{[\delta]}}$, because the duality-trick *fails* to give an ARO. This whole paradigm suggests that de-bordering *may be* harder than derandomization.

Acknowledgments. This work was mostly carried when the first author was a research scholar at CMI, and a visiting scholar at CSE, IIT Kanpur funded by Google PhD Fellowship (2018-2022). The authors would also like to thank the anonymous reviewers for their useful comments and suggestions that improved the presentation of the paper.

REFERENCES

[1] M. Agrawal, *Proving Lower Bounds Via Pseudo-random Generators*, in FSTTCS 2005, 2005, pp. 92–105, https://doi.org/10.1007/11590156_6.
[2] M. Agrawal and S. Biswas, *Primality and identity testing via chinese remaindering*, J. ACM, 50 (2003), pp. 429–443, https://doi.org/10.1145/792538.792540.
[3] M. Agrawal, S. Ghosh, and N. Saxena, *Bootstrapping variables in algebraic circuits*, Proceedings of the National Academy of Sciences, 116 (2019), pp. 8107–8118, https://doi.org/10.1073/pnas.1901272116.
[4] M. Agrawal, R. Gurjar, A. Korwar, and N. Saxena, *Hitting-sets for ROABP and sum of set-multilinear circuits*, SIAM J. Comput., 44 (2015), pp. 669–697, https://doi.org/10.1137/140975103.
[5] M. Agrawal, N. Kayal, and N. Saxena, *Primes is in p*, Annals of mathematics, (2004), pp. 781–793, https://doi.org/10.4007/annals.2004.160.781.
[6] M. Agrawal, C. Saha, R. Saptharishi, and N. Saxena, *Jacobian hits circuits: Hitting sets, lower bounds for depth-d occur-k formulas and depth-3 transcendence degree-k circuits*, SIAM J. Comput., 45 (2016), pp. 1533–1562, https://doi.org/10.1137/130910725.
[7] M. Agrawal and V. Vinay, *Arithmetic circuits: A chasm at depth four*, in FOCS 2008, 2008, pp. 67–75, https://doi.org/10.1109/FOCS.2008.32.
[8] Z. Allen-Zhu, A. Garg, Y. Li, R. M. de Oliveira, and A. Wigderson, *Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing*, in Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC, 2018, pp. 172–181, https://doi.org/10.1145/3188745.3188942.
[9] E. Allender and F. Wang, *On the power of algebraic branching programs of width two*, Computational Complexity, 25 (2016), pp. 217–253, https://doi.org/10.1007/s00037-015-0114-7.
[10] R. Andrews, *Algebraic hardness versus randomness in low characteristic*, in 35th Computational Complexity Conference, CCC, 2020, pp. 37:1–37:32, https://doi.org/10.4230/LIPICS.CCC.2020.37.

[11] R. Andrews and M. A. Forbes, *Ideals, determinants, and straightening: proving and using lower bounds for polynomial ideals*, in 54th Annual ACM SIGACT Symposium on Theory of Computing, STOC, 2022, pp. 389–402, https://doi.org/10.1145/3519935.3520025.

[12] M. Beecken, J. Mittmann, and N. Saxena, *Algebraic independence and blackbox identity testing*, Inf. Comput., 222 (2013), pp. 2–19, https://doi.org/10.1016/j.ic.2012.10.004.

[13] M. Ben-Or and R. Cleve, *Computing algebraic formulas using a constant number of registers*, SIAM J. Comput., 21 (1992), pp. 54–58, https://doi.org/10.1137/0221006.

[14] M. Ben-Or and P. Tiwari, *A deterministic algorithm for sparse multivariate polynominal interpolation (extended abstract)*, in Proceedings of the 20th Annual ACM Symposium on Theory of Computing, STOC, ACM, 1988, pp. 301–309, https://doi.org/10.1145/62212.62241.

[15] A. Bernardi, E. Carlini, M. V. Catalisano, A. Gimigliano, and A. Oneto, *The hitchhiker guide to: Secant varieties and tensor decomposition*, Mathematics, 6 (2018), p. 314, https://doi.org/10.3390/math6120314.

[16] V. Bhargava and S. Ghosh, *Improved hitting set for orbit of roabps*, Comput. Complex., 31 (2022), p. 15, https://doi.org/10.1007/S00037-022-00230-9.

[17] D. Bini, *Relations between exact and approximate bilinear algorithms. applications*, Calcolo, 17 (1980), pp. 87–97, https://doi.org/10.1007/BF02575865.

[18] D. Bini, M. Capovani, F. Romani, and G. Lotti, *$O(n^{2.7799})$ complexity for $n*n$ approximate matrix multiplication*, Inf. Process. Lett., 8 (1979), pp. 234–235, https://doi.org/10.1016/0020-0190(79)90113-3.

[19] P. Bisht and N. Saxena, *Blackbox identity testing for sum of special roabps and its border class*, Comput. Complex., 30 (2021), p. 8, https://doi.org/10.1007/S00037-021-00209-Y.

[20] M. Bläser, J. Dörfler, and C. Ikenmeyer, *On the complexity of evaluating highest weight vectors*, in 36th Computational Complexity Conference, CCC, 2021, pp. 29:1–29:36, https://doi.org/10.4230/LIPICS.CCC.2021.29.

[21] M. Bläser, C. Ikenmeyer, V. Lysikov, A. Pandey, and F. Schreyer, *On the orbit closure containment problem and slice rank of tensors*, in Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA, 2021, pp. 2565–2584, https://doi.org/10.1137/1.9781611976465.152.

[22] M. Bläser, C. Ikenmeyer, M. Mahajan, A. Pandey, and N. Saurabh, *Algebraic Branching Programs, Border Complexity, and Tangent Spaces*, in 35th Computational Complexity Conference CCC, 2020, pp. 21:1–21:24, https://doi.org/10.4230/LIPIcs.CCC.2020.21.

[23] M. Boij, E. Carlini, and A. Geramita, *Monomials as sums of powers: the real binary case*, Proceedings of the American Mathematical Society, 139 (2011), pp. 3039–3043, https://doi.org/10.1090/S0002-9939-2011-11018-9.

[24] K. Bringmann, C. Ikenmeyer, and J. Zuiddam, *On algebraic branching programs of small width*, J. ACM, 65 (2018), pp. 32:1–32:29, https://doi.org/10.1145/3209663.

[25] P. Bürgisser, *The complexity of factors of multivariate polynomials*, Found. Comput. Math., 4 (2004), pp. 369–396, https://doi.org/10.1007/s10208-002-0059-5.

[26] P. Bürgisser, *Correction to: The complexity of factors of multivariate polynomials*, Found. Comput. Math., 20 (2020), pp. 1667–1668, https://doi.org/10.1007/s10208-020-09477-6.

[27] P. Bürgisser, M. Clausen, and M. A. Shokrollahi, *Algebraic complexity theory*, vol. 315, Springer Science & Business Media, 2013, https://doi.org/10.1007/978-3-662-03338-8.

[28] P. Bürgisser, C. Franks, A. Garg, R. M. de Oliveira, M. Walter, and A. Wigderson, *Towards a theory of non-commutative optimization: Geodesic 1st and 2nd order methods for moment maps and polytopes*, in 60th IEEE Annual Symposium on Foundations of Computer Science, FOCS, 2019, pp. 845–861, https://doi.org/10.1109/FOCS.2019.00055.

[29] P. Bürgisser, A. Garg, R. M. de Oliveira, M. Walter, and A. Wigderson, *Alternating minimization, scaling algorithms, and the null-cone problem from invariant theory*, in 9th Innovations in Theoretical Computer Science Conference, ITCS, 2018, pp. 24:1–24:20, https://doi.org/10.4230/LIPICS.ITCS.2018.24.

[30] P. Bürgisser and C. Ikenmeyer, *Explicit lower bounds via geometric complexity theory*, in Symposium on Theory of Computing Conference, STOC, 2013, pp. 141–150, https://doi.org/10.1145/2488608.2488627.

[31] P. Bürgisser, J. M. Landsberg, L. Manivel, and J. Weyman, *An overview of mathematical issues arising in the geometric complexity theory approach to $VP \neq VNP$*, SIAM J. Comput., 40 (2011), pp. 1179–1209, https://doi.org/10.1137/090765328.

[32] A. Carbery and J. Wright, *Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$*, Mathematical research letters, 8 (2001), pp. 233–248.

[33] E. Carlini, M. V. Catalisano, and A. V. Geramita, *The solution to the waring problem for monomials and the sum of coprime monomials*, Journal of Algebra, 370 (2012), pp. 5–14,

https://doi.org/10.1016/j.jalgebra.2012.07.028.

[34] P. Chatterjee, M. Kumar, C. Ramya, R. Saptharishi, and A. Tengse, *On the existence of algebraically natural proofs*, in 61st IEEE Annual Symposium on Foundations of Computer Science, FOCS, 2020, pp. 870–880, https://doi.org/10.1109/FOCS46700.2020.00085.

[35] C. Chou, M. Kumar, and N. Solomon, *Hardness vs randomness for bounded depth arithmetic circuits*, in 33rd Computational Complexity Conference, CCC, 2018, pp. 13:1–13:17, https://doi.org/10.4230/LIPICS.CCC.2018.13.

[36] D. Coppersmith and S. Winograd, *Matrix multiplication via arithmetic progressions*, J. Symb. Comput., 9 (1990), pp. 251–280, https://doi.org/10.1016/S0747-7171(08)80013-2.

[37] D. A. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms - an introduction to computational algebraic geometry and commutative algebra*, Undergraduate texts in mathematics, Springer, 2015, https://doi.org/10.1007/978-3-319-16721-3.

[38] R. A. DeMillo and R. J. Lipton, *A probabilistic remark on algebraic program testing*, Information Processing Letters, 7 (1978), pp. 193–195, https://doi.org/10.1016/0020-0190(78)90067-4.

[39] P. Dutta, P. Dwivedi, and N. Saxena, *Deterministic identity testing paradigms for bounded top-fanin depth-4 circuits*, in 36th Computational Complexity Conference, CCC, 2021, pp. 11:1–11:27, https://doi.org/10.4230/LIPICS.CCC.2021.11.

[40] P. Dutta and N. Saxena, *Separated borders: Exponential-gap fanin-hierarchy theorem for approximative depth-3 circuits*, in 63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS, 2022, pp. 200–211, https://doi.org/10.1109/FOCS54457.2022.00026.

[41] P. Dutta, N. Saxena, and A. Sinhababu, *Discovering the roots: Uniform closure results for algebraic classes under factoring*, J. ACM, 69 (2022), pp. 18:1–18:39, https://doi.org/10.1145/3510359.

[42] P. Dutta, N. Saxena, and T. Thierauf, *A largish sum-of-squares implies circuit hardness and derandomization*, in 12th Innovations in Theoretical Computer Science Conference, ITCS, 2021, pp. 23:1–23:21, https://doi.org/10.4230/LIPICS.ITCS.2021.23.

[43] Z. Dvir and A. Shpilka, *Locally decodable codes with two queries and polynomial identity testing for depth 3 circuits*, SIAM J. Comput., 36 (2007), pp. 1404–1434, https://doi.org/10.1137/05063605X.

[44] Z. Dvir, A. Shpilka, and A. Yehudayoff, *Hardness-randomness tradeoffs for bounded depth arithmetic circuits*, SIAM J. Comput., 39 (2009), pp. 1279–1293, https://doi.org/10.1137/080735850.

[45] S. A. Fenner, R. Gurjar, and T. Thierauf, *A deterministic parallel algorithm for bipartite perfect matching*, Commun. ACM, 62 (2019), pp. 109–115, https://doi.org/10.1145/3306208.

[46] M. Forbes, *Some concrete questions on the border complexity of polynomials. presentation given at the workshop on algebraic complexity theory WACT 2016 in Tel Aviv*, 2016, https://www.youtube.com/watch?v=1HMogQIHT6Q.

[47] M. A. Forbes, *Polynomial Identity Testing of Read-Once Oblivious Algebraic Branch-ing Programs*, PhD thesis, Massachusetts Institute of Technology, (2014), https://dspace.mit.edu/handle/1721.1/89843.

[48] M. A. Forbes, *Deterministic divisibility testing via shifted partial derivatives*, in IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS, 2015, pp. 451–465, https://doi.org/10.1109/FOCS.2015.35.

[49] M. A. Forbes, S. Ghosh, and N. Saxena, *Towards blackbox identity testing of log-variate circuits*, in 45th International Colloquium on Automata, Languages, and Programming, ICALP, 2018, pp. 54:1–54:16, https://doi.org/10.4230/LIPICS.ICALP.2018.54.

[50] M. A. Forbes and A. Shpilka, *Explicit Noether normalization for simultaneous conjugation via polynomial identity testing*, in Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 16th International Workshop, APPROX, and 17th International Workshop, RANDOM, 2013, pp. 527–542, https://doi.org/10.1007/978-3-642-40328-6_37.

[51] M. A. Forbes and A. Shpilka, *Quasipolynomial-time identity testing of non-commutative and read-once oblivious algebraic branching programs*, in 54th Annual IEEE Symposium on Foundations of Computer Science, FOCS, 2013, pp. 243–252, https://doi.org/10.1109/FOCS.2013.34.

[52] M. A. Forbes and A. Shpilka, *A PSPACE construction of a hitting set for the closure of small algebraic circuits*, in Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC, 2018, pp. 1180–1192, https://doi.org/10.1145/3188745.

3188792.
[53] A. Garg, L. Gurvits, R. M. de Oliveira, and A. Wigderson, *A deterministic polynomial time algorithm for non-commutative rational identity testing*, in IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS, 2016, pp. 109–117, https://doi.org/10.1109/FOCS.2016.95.
[54] F. Gesmundo and J. M. Landsberg, *Explicit polynomial sequences with maximal spaces of partial derivatives and a question of k. mulmuley*, Theory Comput., 15 (2019), pp. 1–24, https://doi.org/10.4086/TOC.2019.V015A003.
[55] S. Ghosh, *Low Variate Polynomials: Hitting Set and Bootstrapping*, PhD thesis, PhD thesis. Indian Institute of Technology Kanpur, 2019.
[56] J. A. Grochow, *Unifying known lower bounds via geometric complexity theory*, Computational Complexity, 24 (2015), pp. 393–475, https://doi.org/10.1007/s00037-015-0103-x.
[57] J. A. Grochow, M. Kumar, M. Saks, and S. Saraf, *Towards an algebraic natural proofs barrier via polynomial identity testing*, arXiv preprint arXiv:1701.01717, (2017).
[58] J. A. Grochow, K. D. Mulmuley, and Y. Qiao, *Boundaries of VP and VNP*, in 43rd International Colloquium on Automata, Languages, and Programming, ICALP, 2016, pp. 34:1–34:14, https://doi.org/10.4230/LIPICS.ICALP.2016.34.
[59] Z. Guo, *Variety evasive subspace families*, in 36th Computational Complexity Conference, CCC, 2021, pp. 20:1–20:33, https://doi.org/10.4230/LIPICS.CCC.2021.20.
[60] Z. Guo and R. Gurjar, *Improved explicit hitting-sets for roabps*, in Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM, 2020, pp. 4:1–4:16, https://doi.org/10.4230/LIPICS.APPROX/RANDOM.2020.4.
[61] Z. Guo, M. Kumar, R. Saptharishi, and N. Solomon, *Derandomization from algebraic hardness*, SIAM J. Comput., 51 (2022), pp. 315–335, https://doi.org/10.1137/20M1347395, https://doi.org/10.1137/20m1347395.
[62] Z. Guo, N. Saxena, and A. Sinhababu, *Algebraic dependencies and PSPACE algorithms in approximative complexity over any field*, Theory Comput., 15 (2019), pp. 1–30, https://doi.org/10.4086/toc.2019.v015a016.
[63] A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi, *Approaching the chasm at depth four*, Journal of the ACM, 61 (2014), pp. 33:1–33:16, http://doi.acm.org/10.1145/2629541.
[64] A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi, *Arithmetic circuits: A chasm at depth 3*, SIAM J. Comput., 45 (2016), pp. 1064–1079, https://doi.org/doi/10.1137/140957123.
[65] R. Gurjar, *Derandomizing PIT for ROABP and Isolation Lemma for Special Graphs*, PhD thesis, Indian Institute of Technology Kanpur, 2015.
[66] R. Gurjar, A. Korwar, and N. Saxena, *Identity testing for constant-width, and any-order, read-once oblivious arithmetic branching programs*, Theory Comput., 13 (2017), pp. 1–21, https://doi.org/10.4086/toc.2017.v013a002.
[67] J. Heintz and C. Schnorr, *Testing polynomials which are easy to compute (extended abstract)*, in Proceedings of the 12th Annual ACM Symposium on Theory of Computing, STOC, 1980, pp. 262–272, https://doi.org/10.1145/800141.804674.
[68] J. Hüttenhain and P. Lairez, *The boundary of the orbit of the 3-by-3 determinant polynomial*, Comptes Rendus Mathematique, 354 (2016), pp. 931–935, https://doi.org/10.1016/j.crma.2016.07.002.
[69] G. Ivanyos, Y. Qiao, and K. V. Subrahmanyam, *Non-commutative edmonds' problem and matrix semi-invariants*, Comput. Complex., 26 (2018), pp. 717–763, https://doi.org/10.1007/s00037-016-0143-x.
[70] V. Kabanets and R. Impagliazzo, *Derandomizing polynomial identity tests means proving circuit lower bounds*, Comput. Complex., 13 (2004), pp. 1–46, https://doi.org/10.1007/S00037-004-0182-6.
[71] Z. S. Karnin and A. Shpilka, *Black box polynomial identity testing of generalized depth-3 arithmetic circuits with bounded top fan-in*, Comb., 31 (2011), pp. 333–364, https://doi.org/10.1007/S00493-011-2537-3.
[72] N. Kayal, *An exponential lower bound for the sum of powers of bounded degree polynomials*, Electronic Colloquium on Computational Complexity (ECCC), 19 (2012), p. 81, http://eccc.hpi-web.de/report/2012/081.
[73] N. Kayal and S. Saraf, *Blackbox polynomial identity testing for depth 3 circuits*, in 50th Annual IEEE Symposium on Foundations of Computer Science, FOCS, 2009, pp. 198–207, https://doi.org/10.1109/FOCS.2009.67.
[74] N. Kayal and N. Saxena, *Polynomial identity testing for depth 3 circuits*, computational complexity, 16 (2007), pp. 115–138, https://doi.org/10.1007/S00037-007-0226-9.

[75] A. R. Klivans and D. A. Spielman, *Randomness efficient identity testing of multivariate polynomials*, in Proceedings on 33rd Annual ACM Symposium on Theory of Computing, STOC, 2001, pp. 216–223, https://doi.org/10.1145/380752.380801.

[76] P. Koiran, *Arithmetic circuits: The chasm at depth four gets wider*, Theor. Comput. Sci., 448 (2012), pp. 56–65, https://doi.org/10.1016/j.tcs.2012.03.041.

[77] S. Kopparty, S. Saraf, and A. Shpilka, *Equivalence of polynomial identity testing and polynomial factorization*, Comput. Complex., 24 (2015), pp. 295–331, https://doi.org/10.1007/S00037-015-0102-Y.

[78] M. Kumar, *On the power of border of depth-3 arithmetic circuits*, ACM Trans. Comput. Theory, 12 (2020), pp. 5:1–5:8, https://doi.org/10.1145/3371506.

[79] M. Kumar, C. Ramya, R. Saptharishi, and A. Tengse, *If VNP is hard, then so are equations for it*, in 39th International Symposium on Theoretical Aspects of Computer Science, STACS, 2022, pp. 44:1–44:13, https://doi.org/10.4230/LIPICS.STACS.2022.44.

[80] M. Kumar and R. Saptharishi, *Hardness-randomness tradeoffs for algebraic computation*, Bulletin of EATCS, 3 (2019).

[81] M. Kumar, R. Saptharishi, and A. Tengse, *Near-optimal bootstrapping of hitting sets for algebraic circuits*, in Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA, 2019, https://doi.org/10.1137/1.9781611975482.40.

[82] J. M. Landsberg and G. Ottaviani, *New lower bounds for the border rank of matrix multiplication*, Theory Comput., 11 (2015), pp. 285–298, https://doi.org/10.4086/toc.2015.v011a011.

[83] T. Lehmkuhl and T. Lickteig, *On the order of approximation in approximative triadic decompositions of tensors*, Theor. Comput. Sci., 66 (1989), pp. 1–14, https://doi.org/10.1016/0304-3975(89)90141-2.

[84] N. Limaye, S. Srinivasan, and S. Tavenas, *Superpolynomial lower bounds against low-depth algebraic circuits*, Commun. ACM, 67 (2024), pp. 101–108, https://doi.org/10.1145/3611094.

[85] L. Lovász, *On determinants, matchings, and random algorithms*, in Fundamentals of Computation Theory, FCT, Proceedings of the Conference on Algebraic, Arthmetic, and Categorial Methods in Computation Theory, L. Budach, ed., 1979, pp. 565–574.

[86] M. Mahajan, *Algebraic complexity classes*, CoRR, abs/1307.3863 (2013), https://arxiv.org/abs/1307.3863.

[87] D. Medini and A. Shpilka, *Hitting sets and reconstruction for dense orbits in vp_{e} and $\Sigma\Pi\Sigma$ circuits*, in 36th Computational Complexity Conference, CCC, 2021, pp. 19:1–19:27, https://doi.org/10.4230/LIPICS.CCC.2021.19.

[88] P. Mukhopadhyay, *Depth-4 identity testing and Noether's normalization lemma*, in Computer Science - Theory and Applications - 11th International Computer Science Symposium in Russia, CSR, 2016, https://doi.org/10.1007/978-3-319-34171-2_22.

[89] K. Mulmuley, *Geometric complexity theory VI: The flip via positivity*, arXiv preprint arXiv:0704.0229, (2010).

[90] K. Mulmuley, *The GCT program toward the* P *vs.* NP *problem*, Commun. ACM, 55 (2012), pp. 98–107, https://doi.org/10.1145/2184319.2184341.

[91] K. Mulmuley, *Geometric complexity theory V: equivalence between blackbox derandomization of polynomial identity testing and derandomization of noether's normalization lemma*, in FOCS 2012, 2012, pp. 629–638, https://doi.org/10.1109/FOCS.2012.15.

[92] K. Mulmuley and M. A. Sohoni, *Geometric complexity theory I: an approach to the P vs. NP and related problems*, SIAM J. Comput., 31 (2001), pp. 496–526, https://doi.org/10.1137/S009753970038715X.

[93] K. Mulmuley, U. V. Vazirani, and V. V. Vazirani, *Matching is as easy as matrix inversion*, Comb., 7 (1987), pp. 105–113, https://doi.org/10.1007/BF02579206.

[94] D. Mumford, *Algebraic geometry I: complex projective varieties*, Springer Science & Business Media, 1995.

[95] N. Nisan, *Lower bounds for non-commutative computation (extended abstract)*, in Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, STOC, 1991, pp. 410–418, https://doi.org/10.1145/103418.103462.

[96] N. Nisan and A. Wigderson, *Hardness vs Randomness*, Journal of Computer and System Sciences, 49 (1994), pp. 149–167, https://doi.org/10.1016/S0022-0000(05)80043-1.

[97] I. Niven, *Formal power series*, The American Mathematical Monthly, 76 (1969), pp. 871–889, http://www.jstor.org/stable/2317940.

[98] I. C. Oliveira, *Open Problems*, Algebraic Methods, Simons Institute for the Theory of Computing, (2018). Emailed by author.

[99] Ø. Ore, *Über höhere kongruenzen*, Norsk Mat. Forenings Skrifter, 1 (1922), p. 15.

[100] S. Peleg and A. Shpilka, *A generalized sylvester-gallai type theorem for quadratic polynomials*, in 35th Computational Complexity Conference, CCC, 2020, pp. 8:1–8:33, https://doi.org/10.4230/LIPICS.CCC.2020.8.

[101] S. Peleg and A. Shpilka, *Polynomial time deterministic identity testing algorithm for $\Sigma^{[3]}\Pi\Sigma\Pi^{[2]}$ circuits via edelstein-kelly type theorem for quadratic polynomials*, in 53rd Annual ACM SIGACT Symposium on Theory of Computing, STOC, 2021, pp. 259–271, https://doi.org/10.1145/3406325.3451013.

[102] C. Saha, R. Saptharishi, and N. Saxena, *A case of depth-3 identity testing, sparse factorization and duality*, Comput. Complex., 22 (2013), pp. 39–69, https://doi.org/10.1007/S00037-012-0054-4.

[103] C. Saha and B. Thankey, *Hitting sets for orbits of circuit classes and polynomial families*, in Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM, 2021, pp. 50:1–50:26, https://doi.org/10.4230/LIPICS.APPROX/RANDOM.2021.50.

[104] R. Saptharishi, *Unified Approaches to Polynomial Identity Testing and Lower Bounds*, PhD thesis, Chennai Mathematical Institute, 2013, https://www.tifr.res.in/~ramprasad.saptharishi/assets/pubs/phd_thesis.pdf.

[105] R. Saptharishi, *A survey of lower bounds in arithmetic circuit complexity*, Github Survey, (2019), https://github.com/dasarpmar/lowerbounds-survey/releases/tag/v8.0.7.

[106] N. Saxena, *Diagonal Circuit Identity Testing and Lower Bounds*, in ICALP 2008, 2008, pp. 60–71, https://doi.org/10.1007/978-3-540-70575-8_6.

[107] N. Saxena, *Progress on polynomial identity testing-II*, in Perspectives in Computational Complexity, Springer, 2014, pp. 131–146, https://doi.org/10.1007/978-3-319-05446-9_7.

[108] N. Saxena and C. Seshadhri, *An almost optimal rank bound for depth-3 identities*, SIAM J. Comput., 40 (2011), pp. 200–224, https://doi.org/10.1137/090770679.

[109] N. Saxena and C. Seshadhri, *Blackbox identity testing for bounded top-fanin depth-3 circuits: The field doesn't matter*, SIAM J. Comput., 41 (2012), pp. 1285–1298, https://doi.org/10.1137/10848232.

[110] N. Saxena and C. Seshadhri, *From sylvester-gallai configurations to rank bounds: Improved blackbox identity test for depth-3 circuits*, J. ACM, 60 (2013), pp. 33:1–33:33, https://doi.org/10.1145/2528403.

[111] J. T. Schwartz, *Fast probabilistic algorithms for verification of polynomial identities*, J. ACM, 27 (1980), pp. 701–717, https://doi.org/10.1145/322217.322225.

[112] A. Shpilka, *Sylvester-gallai type theorems for quadratic polynomials*, in Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC, 2019, pp. 1203–1214, https://doi.org/10.1145/3313276.3316341.

[113] A. Shpilka and A. Yehudayoff, *Arithmetic circuits: A survey of recent results and open questions*, Foundations and Trends in Theoretical Computer Science, 5 (2010), pp. 207–388, http://dx.doi.org/10.1561/0400000039.

[114] A. K. Sinhababu, *Power series in complexity: Algebraic Dependence, Factor Conjecture and Hitting Set for Closure of VP*, PhD thesis, PhD thesis, Indian Institute of Technology Kanpur, 2019.

[115] V. Strassen, *Vermeidung von divisionen.*, Journal für die reine und angewandte Mathematik, 264 (1973), pp. 184–202.

[116] V. Strassen, *Polynomials with rational coefficients which are hard to compute*, SIAM J. Comput., 3 (1974), pp. 128–149, https://doi.org/10.1137/0203010, https://doi.org/10.1137/0203010.

[117] J. J. Sylvester, *On the principles of the calculus of forms*, éditeur inconnu, 1852.

[118] S. Tavenas, *Improved bounds for reduction to depth 4 and depth 3*, Inf. Comput., 240 (2015), pp. 2–11, https://doi.org/10.1016/J.IC.2014.09.004.

[119] L. G. Valiant, *Completeness classes in algebra*, in Proceedings of the 11h Annual ACM Symposium on Theory of Computing, 1979, pp. 249–261, https://doi.org/10.1145/800135.804419.

[120] L. G. Valiant, S. Skyum, S. Berkowitz, and C. Rackoff, *Fast Parallel Computation of Polynomials Using Few Processors*, SIAM Journal of Computing, 12 (1983), pp. 641–644, https://doi.org/10.1137/0212043. MFCS 1981.

[121] R. Zippel, *Probabilistic algorithms for sparse polynomials*, in Symbolic and Algebraic Computation, EUROSAM '79, An International Symposiumon Symbolic and Algebraic Computation, 1979, pp. 216–226, https://doi.org/10.1007/3-540-09519-5_73.