Electronic Colloquium on Computational Complexity, Report No. 46 (2011)

Black-Box Identity Testing of Depth-4 Multilinear Circuits

Shubhangi Saraf [*] Ilya Volkovich [†]

Abstract

We study the problem of identity testing for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits, i.e. multilinear depth-4 circuits with fan-in k at the top + gate. We give the first polynomial-time deterministic identity testing algorithm for such circuits. Our results also hold in the black-box setting.

The running time of our algorithm is $(ns)^{\mathcal{O}(k^3)}$, where n is the number of variables, s is the size of the circuit and k is the fan-in of the top gate. The importance of this model arises from [AV08], where it was shown that derandomizing black-box polynomial identity testing for general depth-4 circuits implies a derandomization of polynomial identity testing (PIT) for general arithmetic circuits. Prior to our work, the best PIT algorithm for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits [KMSV10] ran in quasi-polynomial-time, with the running time being $n^{\mathcal{O}(k^6 \log(k) \log^2 s)}$.

We obtain our results by showing a strong *structural result* for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits that compute the zero polynomial. We show that under some mild technical conditions, any gate of such a circuit must compute a *sparse* polynomial. We then show how to combine the structure theorem with a result by Klivans and Spielman [KS01], on the identity testing for sparse polynomials, to yield the full result.

[*] CSAIL, MIT. Email: `shibs@mit.edu`.

[†] Faculty of Computer Science, Technion, Haifa 32000, Israel. Email: `ilyav@cs.technion.ac.il`. Research received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 257575.

ISSN 1433-8092

1 Introduction

A central problem in algebraic complexity theory and algorithms design is the problem of Polynomial Identity Testing (PIT): given an arithmetic circuit C over a field $\mathbb{F}$, with input variables $x_1, x_2, \ldots, x_n$, determine whether C computes the identically zero polynomial. Numerous applications and connections to other algorithmic and number theoretic problems further emphasize the significance of PIT. Among the examples are algorithms for finding perfect matchings in graphs [Lov79, MVV87], primality testing [AKS04], and many more. In addition, PIT also shows up in many fundamental results in complexity theory such as $\mathsf{IP} = \mathsf{PSPACE}$ [LFKN92, Sha90] and the PCP theorem [AS98, ALM$^+$98].

PIT is one of the most basic and natural questions for which a very simple randomized solution is known: Schwartz and Zippel [Sch80, Zip79] independently showed that if one evaluates the circuit at a randomly chosen point from a sufficiently large domain, then with high probability any non-zero circuit will evaluate to a non-zero value. It has been a long standing open question to derandomize the algorithm.

The main open question is to come up with an efficient (i.e. polynomial-time or at least subexponential-time) *deterministic* algorithm for the problem. Indeed, Kabanets and Impagliazzo [KI04] showed that any deterministic algorithm for identity testing implies *super polynomial* circuit lower bounds: either $\mathsf{NEXP} \not\subseteq \mathsf{P/poly}$ or the Permanent has no polynomial size arithmetic circuits. Other connections between deterministic PIT algorithms and circuits lower bounds were given in [HS80, DSY09].

A very natural and often desirable setting to consider the PIT question is in the *black-box* model. The connection to lower bounds is even more natural and strong in this case. In the black-box setting, one is not given the full description of the circuit C but only allowed black-box (oracle) access to C. The problem of derandomizing identity testing in this setting reduces to that of finding for every s an explicit set of points $\mathcal{H} \subseteq \mathbb{F}^n$ of size $\mathrm{poly}(s)$ such that any non-zero circuit of size s does not vanish on $\mathcal{H}$. We refer to such sets as *hitting sets*. The Schwartz-Zippel test, in fact, provides an exponential-size hitting set. Furthermore, applying standard probabilistic arguments one can show existence of “small” hitting sets. Interestingly, any explicit construction of a hitting set (for any class of circuits) immediately gives, via interpolation, an explicit polynomial that cannot be computed by that class of circuits [Agr05].

In a recent surprising result by Agrawal and Vinay [AV08], it was shown that a complete derandomization of black-box identity testing for just depth-4 ($\Sigma\Pi\Sigma\Pi$) arithmetic circuits already implies a near complete derandomization for the general PIT problem. More precisely, they showed that black-box identity testing for depth-4 ($\Sigma\Pi\Sigma\Pi$) arithmetic circuits implies *exponential* lower bounds for general arithmetic circuits, which in turn implies a quasi-polynomial-time algorithm for the general PIT problem. This makes black-box identity testing for even very low depth circuits a very rewarding pursuit!

For a long time, black-box identity tests were only known for depth-2 circuits (equivalently circuits computing sparse polynomials) [BOT88, KS01, LV03] (and references within). In light of the Agrawal-Vinay result, studying black-box identity testing for depth-3 and depth-4 circuits seems to be a very promising direction and line of attack for the general PIT problem. In recent times there has been a surge of results on black-box (and non-black-box) identity testing for some classes of depth-3 circuits such as depth-3 circuits with bounded top fan-in (also known as $\Sigma\Pi\Sigma(k)$ circuits) [DS06, KS07, KS08, KS09, SS09, AM10, SS10, SS11], and even some restricted classes of depth-4 circuits [Sax08, SV09, KMSV10, AM10]. For more information on PIT we refer the reader to the survey [SY10].

In the current paper, we study multilinear depth-4 $\Sigma\Pi\Sigma\Pi(k)$ circuits. Very recently Karnin et al. [KMSV10] gave the first deterministic subexponential-time (in fact, quasi-polynomial-time) PIT algorithm for this model. Their result was in the black-box setting. We further investigate this class of circuits and give the first deterministic *polynomial-time* black-box PIT algorithm for it. Our approach is quite different from that taken in [KMSV10], and we believe that the techniques might be useful to understand other, more general classes of circuits as well.

Following the same approach as in [AV08], it can be shown that derandomizing black-box identity testing for *multilinear* depth-4 ($\Sigma\Pi\Sigma\Pi$) circuits implies an *exponential* lower bound for general *multilinear* arithmetic circuits. Getting explicit lower bounds is one of the biggest challenges of complexity theory and has been the focus of much research. So far, the best known lower bounds are: $\Omega(n^{4/3}/\log^2 n)$ for multilinear circuits due to Raz et al. [RSY08], and $n^{\Omega(\log n)}$ for multilinear formulas due to Raz [Raz09]. It is an interesting open question to improve any of those bounds. All the above makes the study of PIT for depth-4 circuits, even in the multilinear case, a really interesting and challenging open question.

We now define the model of multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits formally. Similar definitions were given in [KMSV10], however we repeat them for the sake of completeness. A depth-4 circuit has 4 layers of alternating $(+, \times)$ gates and it computes a polynomial of the form

$$C(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{k} F_i = \sum_{i=1}^{k} \prod_{j=1}^{d_i} P_{ij}$$

where k is the fan-in of the top Σ gate and d_i are the fan-ins of the Π gates at the second level. We refer to F_i-s as the multiplication gates and P_{ij}-s are the polynomials computed at the third level of the circuit (which is a $\Sigma\Pi$ component). This implies that if the size of C is s then, clearly, each P_{ij} in C can be computed by a depth-2 ($\Sigma\Pi$) circuit of size at most s. Such polynomials are known as *s-sparse* polynomials as they contain at most s non-zero monomials (see e.g [BOT88, KS01, LV03]). In other words, each P_{ij} is s-sparse. We define $\gcd(C) \triangleq \gcd(F_1, \ldots, F_k)$, that is, the gcd of the set of polynomials computed by the multiplication gates. We say that C is *simple* if $\gcd(C) = 1$. A $\Sigma\Pi\Sigma\Pi(k)$ circuit is called *minimal* if for every proper subset $\emptyset \subsetneq A \subsetneq [k]$, the corresponding subcircuit $C_A \triangleq \sum_{i \in A} F_i$ of C is non-zero. Multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits are circuits in which the fan-in of the top Σ gate is a constant k and each multiplication gate F_i computes a multilinear polynomial. We say that a polynomial is *s-dense* if it contains more than s monomials.

The main technical contribution in the proof of both black-box and non black-box identity testing algorithms for $\Sigma\Pi\Sigma\Pi(k)$ circuits is a new structural theorem for identically zero multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits. We refer to it as the *Sparsity Bound*. This result can be viewed as a natural (though unsuspected) generalization of the previously shown structural theorems for depth-3 $\Sigma\Pi\Sigma(k)$ circuits known as the *Rank Bound* (see Section 1.3). Our result lends optimism to the hope that similar structural results should also hold for general $\Sigma\Pi\Sigma\Pi(k)$ circuits (without the restriction of multilinearity). At a very high level, we show that the only way a multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit can completely cancel itself out and compute the zero polynomial is that the circuit must not be computing very "complex" polynomials. In particular, we show that in any simple and minimal multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit that computes the identically zero polynomial, the polynomials computed at the multiplication gates must be *sparse*.

Theorem 1 (Sparsity Bound for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits)**.** *Let $k \geq 2$ and let $C(\bar{x}) = \sum_{i=1}^{k} F_i(\bar{x})$ be a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit of size s computing the zero polynomial. Then each F_i is $s^{\mathcal{O}(k^2)}$-sparse.*

One way to interpret the theorem is as follows: for a fixed k the sparsity of each multiplication gate in a simple and minimal, identically zero multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit is at most polynomially large in the size of the circuit. Note, that for general circuits this sparsity can be exponentially large. Later on (Section 4.1) we show a lower bound on the multiplication gate's sparsity indicating that our result is nearly optimal. Once we have the structure theorem we exploit it to design PIT algorithms in both black-box and non black-box settings, thus proving the following theorems.

Theorem 2 (Black-Box PIT for $\Sigma\Pi\Sigma\Pi(k)$ circuits)**.** *Let k, n, s be integers. There is an explicit set $\mathcal{H}$ of size $n^{\mathcal{O}(k)} \cdot s^{\mathcal{O}(k^3)}$, that can be constructed in time $n^{\mathcal{O}(k)} \cdot s^{\mathcal{O}(k^3)}$, such that the following holds. Let $P \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be a non-zero polynomial computed by a multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit of size s on n variables. Then $P|_{\mathcal{H}} \not\equiv 0$.*

In our construction we heavily use the black-box PIT algorithm of [KS01] for sparse polynomials as a subroutine [1]. Using their PIT algorithm we show how to make any "non sparse" circuit into a "somewhat sparse" circuit. Our structure theorem then guarantees that within this process, we do not inadvertently end up making a non-zero circuit into a zero circuit. Once we have a "somewhat sparse" non-zero circuit, we use the above PIT algorithm coupled with some techniques from [KMSV10] to find a non-zero evaluation point for it, and hence get a black-box identity tester. In the non black-box setting (e.g. when the circuit is given to us explicitly) we get a slight improvement in the running time.

Theorem 3 (Non Black-Box PIT for $\Sigma\Pi\Sigma\Pi(k)$ circuits)**.** *Let k, n, s be integers. Given a multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit C of size s computing a polynomial over $\mathbb{F}[x_1, x_2, \ldots, x_n]$ there exists an algorithm that runs in time $\mathrm{poly}(n) \cdot s^{\mathcal{O}(k^2)}$ and determines whether $C \equiv 0$.*

1.1 Overview of the Proof of The Structure Theorem

As mentioned earlier, our algorithm is based on a new structure theorem for simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits $C = \sum_{i=1}^{k} F_i = \sum_{i=1}^{k} \prod_{j=1}^{d_i} P_{ij}$. We now give an overview of the proof of the structural theorem. We wish to find an upper bound on the sparsity (i.e. the number of non-zero monomials) of the polynomials computed by the multiplication gates (F_i-s). At a high level, our strategy will be to set a multiplication gate to zero (thus obtaining a circuit with less multiplication gates) and then to use an inductive argument on the resulting circuit. To do so we will find a partial zero assignment $\bar{a}$ to some P_{ij} (i.e. $P_{ij}(\bar{a}) = 0$) and substitute it into C. Let $C' = \sum_{i=1}^{k} F'_i$ be the resulting circuit obtained by the substitution of $\bar{a}$. First of all, note that C' may not satisfy the conditions of the inductive claim. In other words, the substitution may compromise either simplicity or minimality of the circuit (or both of them). Furthermore, note that a partial substitution may decrease the sparsity of the multiplication gates. That is, for some $i \in [k]$ the sparsity of F'_i might be much smaller than the sparsity of F_i. The main issue here is that it is not clear how to bound this difference in sparsity accurately enough. Hence, any upper bound obtained on the sparsity of the gates of C' may not provide us with any useful information on the sparsity of the gates in the original circuit C. This makes the offered strategy problematic.

[1] In fact, our construction works with any black-box PIT algorithm for sparse polynomials.

Nevertheless, we show that we can still work around these problems. Our main idea behind effectively bounding the sparsity is that, instead of identifying a single P_{ij} and setting it to zero, we will go over all P_{ij}-s and set them to zero one at a time. Once we select a P_{ij} to set to zero, we will look for a zero assignment of P_{ij} that preserves certain special properties of C. More specifically, we will find a zero assignment of P_{ij} such that after the substitution, the resulting circuit C' is simple and minimal, and in addition, the aforementioned decrease in sparsity is brought to a minimum. In order to find such an assignment we construct a polynomial Φ such that the above conditions (e.g. "simplicity", "minimality" and "sparsity difference minimization") can be formulated in terms of being a non-zero assignment of Φ. Consequently, the desired assignment $\bar{a}$ will be a *zero* assignment of P_{ij} which is simultaneously also a *non-zero* assignment of Φ. However, such an assignment may not even exist (for example if P_{ij} is a factor of Φ). To handle this problem we introduce a new technique. In fact, we settle on finding a zero assignment $\bar{a}$ of P_{ij} which is "almost good", paying some "small price" for it. The main issue turns out to be the estimation of the sparsity difference that results from the partial substitution (i.e. Sparsity(F_i) - Sparsity(F_i')). For this purpose we use Shearer's Lemma (see Lemma 2.5). To apply the lemma, we map each multilinear polynomial P to a family of sets corresponding to P's non-zero monomials (see Definition 2.1). Note that the lemma suggests that the more distinct partial substitutions we have, the tighter is the bound. This is the reason why we combine the information received from all different partial substitutions to C (i.e. by trying out all the P_{ij}).

1.2 Overview of the Black-Box Algorithm

Our black-box algorithm is based on the Sparsity Bound. For the moment, let us ignore the issues of simplicity and minimality. These do create issues which will have to have to be addressed, but just to understand the motivation for the algorithm suppose that we were a given a simple and minimal $\Sigma\Pi\Sigma\Pi(k)$ circuit C. There can be two cases: either (i) each multiplication gate of C is sparse (e.g. the circuit is "sparse") or (ii) C has a dense multiplication gate (e.g. the circuit is "dense"). In case (i) the polynomial computed by C is sparse (as a sum of a small number of sparse polynomials) therefore one can invoke a PIT algorithm for sparse polynomials to check if $C \equiv 0$. In case (ii) the Sparsity Bound implies that $C \not\equiv 0$. This observation gives rise to a non black-box PIT algorithm: Compute the sparsity of each gate separately. If there is a dense gate, then conclude that $C \not\equiv 0$. Otherwise, check whether the monomials cancel each other out. Again, the conditions of simplicity and minimality create some issues, but they can be dealt with (for more details see Section 6).

The black-box setting is trickier, since we do not get to "see" the circuit, and hence cannot carry out the procedure described above. Indeed, the main problem is to decide in which of the cases we are. Our strategy will be to walk on the edge between the two cases. Given a non-zero circuit C we are going to gradually reduce the sparsities of the gates, step-by-step, until we reach case (i), while preserving the properties of C (simplicity and minimality). In each step the sparsities of gates will reduce by a "small" factor. Through the entire process the Sparsity Bound will guarantee that C remains non-zero as long as we are in case (ii). Now, let us consider the "edge step" that is, the last reduction step before we reach case (i). We claim that in this step the circuit is dense, but not "too dense". On one hand, we are still in case (ii) where the Sparsity Bound guarantees that $C \not\equiv 0$. But on the other hand, a reduction by a "small" factor makes the circuit sparse. Hence, we can conclude that the circuit is non-zero and "somewhat sparse". To make the argument a bit more formal, consider a non-zero $\Sigma\Pi\Sigma\Pi(k)$ circuit C of size s. If C

is $s^{\mathcal{O}(k^2)}$-sparse, then we are done. Otherwise, let $\bar{a} = (a_1, \ldots, a_n) \in \mathbb{F}^n$ be an assignment such that for every $1 \leq t \leq n$ the circuit $C(a_1, \ldots, a_t, x_{t+1}, \ldots, x_n)$ - resulting from setting $x_j = a_j$ for $1 \leq j \leq t$ - is simple and minimal. In particular, $F_i(\bar{a}) \neq 0$ for each $1 \leq i \leq k$, and hence each $F_i(\bar{a})$ is 1-sparse. Let $0 \leq t \leq n$ be the largest index such that the circuit $C(a_1, \ldots, a_t, x_{t+1}, \ldots, x_n)$ has an $s^{\mathcal{O}(k^2)}$-dense multiplication gate. Such an index exists by a hybrid argument. From the Sparsity Bound, we get that $C(a_1, \ldots, a_t, x_{t+1}, \ldots, x_n) \not\equiv 0$. One the other hand, from the choice of t we get that each multiplication gate in $C(a_1, \ldots, a_t, a_{t+1}, x_{t+2} \ldots, x_n)$ is $s^{\mathcal{O}(k^2)}$-sparse. Observe that since each F_i is multilinear all the P_{ij}-s in it must be variable disjoint. Thus, setting $x_{t+1} = a_{t+1}$ can affect at most one P_{ij} in each F_i, and hence reduce the sparsity of each F_i by a factor of at most s (recall that each P_{ij} is s-sparse and that $P_{ij}(\bar{a}) \neq 0$). Consequently, each multiplication gate in $C(a_1, \ldots, a_t, x_{t+1}, \ldots, x_n)$ is $s \cdot s^{\mathcal{O}(k^2)}$-sparse. A priori we do not know what this index t is. However, it turns out not to matter, since we can just run over all possible indices.

In all this discussion, we hid many technical issues under the rug such as the simplicity and minimality of the original circuits and how to find such an assignment $\bar{a}$. We note that in the black-box setting minimality comes for free since we can assume w.l.o.g that the black-box contains a minimal circuit. To handle simplicity and find an assignment $\bar{a}$ as above we use the method of *generators* (see Section 2.3). This method was previously used in [SV09] and [KMSV10]. In [KMSV10] this method has been already used to work around the aforementioned problems and to ensure that all the steps of the proof go through smoothly.

1.3 Related Previous Results

Since the vast majority of the work on low depth circuits focused on circuits of depths 2 and 3, and since our results can be viewed as an extension of this body of work, we formally define depth-3 circuits and some related and relevant notions. A depth-3 $\Sigma\Pi\Sigma(k)$ circuit C of degree d computes a polynomial of the form

$$C(\bar{x}) = \sum_{i=1}^{k} F_i(\bar{x}) = \sum_{i=1}^{k} \prod_{j=1}^{d_i} L_{ij}(\bar{x})$$

where the $L_{ij}(\bar{x})$-s are linear functions: $L_{ij}(\bar{x}) = \sum\limits_{t=1}^{n} a_{ij}^t x_t + a_{ij}^0$ with $a_{ij}^t \in \mathbb{F}$, and $d_i \leq d$. Note that L_{ij}-s are irreducible polynomials. A multilinear $\Sigma\Pi\Sigma(k)$ circuit has the additional requirement that each F_i is a multilinear polynomial. We refer to the F_i-s as the multiplication gates of the circuit. A *subcircuit* of C is defined as a sum of a subset of the multiplication gates in C. Let $\gcd(C) \triangleq \gcd(F_1, F_2, \ldots, F_k)$. We say that a circuit is *simple* if $\gcd(C) = 1$. We say that a circuit is *minimal* if no proper subcircuit of C computes the zero polynomial. Define the *rank* of C, denoted by $\mathrm{rank}(C)$, as the rank of its linear functions, viewed as $(n+1)$-dimensional vectors over $\mathbb{F}^{n+1}$. Formally: $\mathrm{rank}(C) \triangleq \dim\left(\mathrm{span}\{L_{ij}\}_{i \in [k], j \in [d_i]}\right)$. Clearly, $\Sigma\Pi\Sigma(k)$ circuits are a restricted case of $\Sigma\Pi\Sigma\Pi(k)$ circuits.

The first non black-box [DS06] and almost all the black-box PIT algorithms for $\Sigma\Pi\Sigma(k)$ circuits [KS08, KS09, SS09, SS10] were designed based on the following structural property of $\Sigma\Pi\Sigma(k)$ known as the *Rank Bound*.

Lemma 1.1. *There exists an increasing function $R(k, d)$ such that if C is a simple and minimal $\Sigma\Pi\Sigma(k)$ circuit of degree d computes the zero polynomial then* $\mathrm{rank}(C) < R(k, d)$.

Up until recently, all the resulting black-box PIT algorithms for $\Sigma\Pi\Sigma(k)$ circuits were based on the black-box PIT algorithm of [KS08] that admits a running time of $\mathrm{poly}(n) \cdot d^{\mathcal{O}(R(k,d))}$. The recent result of [SS11], giving a $\mathrm{poly}(n) \cdot d^{\mathcal{O}(k)}$ time black-box PIT algorithm, was obtained by a different approach. In the non black-box setting Kayal & Saxena [KS07] presented a $\mathrm{poly}(n) \cdot d^{\mathcal{O}(k)}$ time PIT algorithm by another methodology.

The first result of Dvir & Shpilka [DS06] gave an upper bound of $R(k,d) = 2^{\mathcal{O}(k^2)} \log^{k-2} d$ over general fields. It was later improved by Saxena & Seshadhri [SS09, SS10] to $R(k,d) = \mathcal{O}(k^2 \log d)$. Based on an example of [KS07] they, however, also illustrated a limitation of this approach by exhibiting a lower bound of $\mathrm{rank}(C) = \Omega(k \log d)$ over finite fields, implying that the best black-box PIT algorithm for finite fields achieved via this approach will be quasi-polynomial-time for constant values of k. However, over the field of reals $\mathbb{R}$ the bound was significantly improved by Kayal & Saraf [KS09] to $k^{\mathcal{O}(k)}$ and later on by Saxena & Seshadhri [SS10] to $R(k) = \mathcal{O}(k^2)$. Note that there is no dependence on d, thus implying a polynomial-time algorithm.

For multilinear $\Sigma\Pi\Sigma(k)$ circuits the best upper bound of $R_{ML}(k) = \mathcal{O}(k^2 \log k)$ for general fields was shown in [SS10]. Yet, the PIT algorithm with the best running time of $n^{\mathcal{O}(k)}$ obtained in [SV09] does not rely on the Rank Bound.

Much less is understood about depth-4 circuits. The existing deterministic PIT algorithms had covered only very restricted classes of those circuits [Sax08, SV09, AM10]. The first PIT algorithm for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits was given by Karnin et al. in [KMSV10]. The algorithm is in the black-box setting and has a running time of $n^{\mathcal{O}(k^3 \cdot R_{ML}(k) \cdot \log^2 s)}$, when s is the size of the circuit and $R_{ML}(k)$ is the multilinear Rank Bound (see above). This implies a quasi-polynomial-time algorithm for constant values of k. The main ingredient of the algorithm is a new structural theorem for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits suggesting that there is a non-zero multilinear $\Sigma\Pi\Sigma(k)$ circuit "hiding" in every non-zero multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit. The idea is to "search" for the hidden $\Sigma\Pi\Sigma(k)$ circuit using the multilinear Rank Bound. This search is what makes the algorithm quasi-polynomial.

The Rank Bound relies strongly on the properties of linear functions. That is, one can define a linear space spanned by the circuit components and benefit from its structure. This notion is absent when moving to depth-4 circuits. In the current paper we suggest a natural generalization of this notion - the *Sparsity Bound*. The idea is to bound the sparsities of the polynomials computed by each multiplication gate in terms of the sparsities of the circuit components (P_{ij}-s). This approach can be seen as an extension of the Rank Bound approach. The following lemma demonstrates this point and, in fact, can be seen as a "sanity check".

Lemma 1.2. *Let $C(\bar{x}) = \sum_{i=1}^{k} F_i(\bar{x}) = \sum_{i=1}^{k} \prod_{j=1}^{d_i} L_{ij}(\bar{x})$ be a simple and minimal, multilinear $\Sigma\Pi\Sigma(k)$ circuit computing the zero polynomial. Let s denote the maximal sparsity of a L_{ij} appearing in C. Then each F_i is $s^{R_{ML}(k)}$-sparse.*

Proof. Fix $i \in [k]$ and consider $F_i = \prod_{j=1}^{d_i} L_{ij}(\bar{x})$. By definition, each F_i is s^{d_i}-sparse. As F_i is a multilinear polynomial, it must be the case that all the L_{ij}-s of it are variable-disjoint and hence linearly independent. Therefore, by the Rank Bound: $d_i = \dim\left(\mathrm{span}\{L_{ij}\}_{j \in [d_i]}\right) \leq \mathrm{rank}(C) < R_{ML}(k)$ and whence F_i is $s^{R_{ML}(k)}$-sparse. □

Finally, we note that in the light of the connections between deterministic PIT and circuit lower bounds, the recent results of [Raz06, Raz09, RSY08, RY09], showing lower bounds for multilinear circuits and formulas, suggest that efficient identity testers for multilinear formulas may be at reach. Indeed, a major progress in this question has been made in the recent of work of Anderson et al. [AvMV11]. In this work, multilinear read-k formulas [2] were studied, resulting in polynomial and quasi-polynomial time PIT algorithms in the black-box and the non black-box settings, respectively, for constant values of k. In fact, they have studied a broader model - multilinear read-k formulas in which each leaf (variable) can be replaced by a sparse polynomial. The model is referred to as "sparse-substituted" and it extends the model considered in this paper. However, the resulting black-box PIT algorithm for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits admits a quasi-polynomial running time.

1.4 Organization

We start by some basic definitions and notation in Section 2. In Section 3 we formally introduce our model and give some related definition. In Section 4, we prove our structural theorem (Theorem 1) and exhibit a lower bound. We give our main result - a black-box PIT algorithm for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits, in Section 5, thus proving Theorem 2. Finally, in Section 6 we consider a non black-box PIT algorithm for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits, proving Theorem 3.

2 Preliminaries

For a positive integer n, let $[n]$ denote the set $\{1, \ldots, n\}$. Let $\mathbb{F}$ be the underlying field, and $\overline{\mathbb{F}}$ be its algebraic closure. For a polynomial $P(x_1, \ldots, x_n)$, a variable x_i, and $\alpha \in \mathbb{F}$, let $P|_{x_i=\alpha}$ denote the polynomial that results upon setting $x_i = \alpha$. We now give some definitions that apply to polynomials $P, Q \in \mathbb{F}[x_1, x_2, \ldots, x_n]$. We say that P *depends* on x_i if there exist $\bar{a} \in \overline{\mathbb{F}}^n$ and $b \in \overline{\mathbb{F}}$ such that: $P(a_1, a_2, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n) \neq P(a_1, a_2, \ldots, a_{i-1}, b, a_{i+1}, \ldots, a_n)$. We denote $\mathrm{var}(P) \triangleq \{i \ : \ P \text{ depends on } x_i\}$. Intuitively, P depends on x_i if x_i appears when P is written as a sum of monomials. Given a subset $I \subseteq [n]$ and an assignment $\bar{a} \in \mathbb{F}^n$ we define $P|_{\bar{x}_I = \bar{a}_I}$ to be the polynomial resulting from setting $x_i = a_i$ for every $i \in I$. We say that P *divides* Q, or equivalently Q is *divisible by* P, and denote it by $P \mid Q$ if there exists a polynomial $h \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ such that $Q = P \cdot h$. Otherwise, we say that P does not divide Q (or Q is not divisible by P) and denote it by $P \not| \, Q$. Given the notion of divisibility we define the gcd of a set of polynomials in the natural way.

2.1 Sparsity of a Polynomial

In this section we formally define the notion "sparsity of a polynomial". We also show how to upper bound the sparsity of a given polynomial via the sparsities of its different partial substitutions. For simplicity we concentrate on multilinear polynomials, however the definitions can be generalized to all polynomials. A multilinear polynomial $P \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ can be (uniquely) written as

$$P = \sum_{A \subseteq [n]} \alpha_A \cdot X^A \tag{1}$$

where $\alpha_A \in \mathbb{F}$ (the coefficients) and X^A denotes $\prod\limits_{i \in A} x_i$.

[2] read-k formulas are arithmetic formulas in which each variable can appear at most k times.

Definition 2.1. *We define the* characteristic set *of a multilinear polynomial P as* $\chi_P \triangleq \{A \;:\; A \subseteq [n],\ \alpha_A \neq 0\}$. *The* sparsity *of P is defined as the number of (the non-zero) monomials of P and denoted by $\|P\|$. Clearly, $\|P\| = |\chi_P|$. For the purposes of connecting the sparsity of a given polynomial with the sparsities of its partial substitutions we extend these notions. For every $I \subseteq [n]$ we define $\chi_{P|_I} \triangleq \{A \setminus I \;:\; A \subseteq [n],\ \alpha_A \neq 0\}$ and $\|P\|_I \triangleq \left|\chi_{P|_I}\right|$.*

The following is immediate from the definitions and will be used implicitly.

Corollary 2.2. *Let $P, Q \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be variable-disjoint polynomials and let $I \subseteq [n]$. Then $\|P\|_I \leq \|P\|$ and $\|P \cdot Q\|_I = \|P\|_I \cdot \|Q\|_I$.*

Intuitively, $\chi_{P|_I}$ captures the distinct monomials left after we "erase" the variables of $\bar{x}_I$. In fact, the same effect is achieved by setting these variables to field elements. However, different substitutions may lead to different results, as monomials may cancel out. For example: $\|(x_1x_3 + x_1x_2x_3)|_{x_3=1}\| = 2 = \|x_1x_3 + x_1x_2x_3\|_{\{3\}}$, however $\|(x_1x_3 + x_1x_2x_3)|_{x_3=0}\| = 0$. It is not hard too see that a random assignment to the variables of $\bar{x}_I$ should not lead to any unwanted cancellations, and thus achieve the same effect in sparsity as "erasing" the variables of I.

We show that $\|P\|_I$ is, in fact, attained by the substitution that maximizes the sparsity of the resulting polynomial. We do so by formulating this condition (of not creating any unwanted cancellations and hence maximizing the sparsity) as that of being a non-zero assignment to a certain polynomial Ψ_P. Observe that a random assignment would indeed satisfy this condition. Before giving the proof, we need some more definitions.

Definition 2.3. *Let $P \in \mathbb{F}[x_1, x_2, \ldots, x_n]$. For $B, I \subseteq [n]$ such that $B \cap I = \emptyset$ we define:*

$$P_{B,I} \triangleq \sum_{J \subseteq I} \alpha_{B \cup J} \cdot X^J \quad \textit{and} \quad \Psi_P \triangleq \prod_{P_{B,I} \not\equiv 0} P_{B,I}.$$

Intuitively, $P_{B,I}$ represents the monomials of P that could be affected by setting the variables in $\bar{x}_I$ to field elements. Ψ_P captures all the non-zero $P_{B,I}$-s.

Lemma 2.4 (Max sparsity Condition)**.** *Let $I \subseteq [n]$ and let $\bar{a} \in \overline{\mathbb{F}}^n$ be such that $\Psi_P|_{\bar{x}_I = \bar{a}_I} \not\equiv 0$. Then $\|P|_{\bar{x}_I=\bar{a}_I}\| = \max\limits_{\bar{b} \in \overline{\mathbb{F}}^n} \|P|_{\bar{x}_I=\bar{b}_I}\| = \|P\|_I$.*

Proof. Clearly, $\|P|_{\bar{x}_I=\bar{a}_I}\| \leq \max\limits_{\bar{b} \in \overline{\mathbb{F}}^n} \|P|_{\bar{x}_I=\bar{b}_I}\| \leq \|P\|_I$. We now show that $\|P\|_I \leq \|P|_{\bar{x}_I=\bar{a}_I}\|$. In fact, we show that $\chi_{P|_I} \subseteq \chi_{(P|_{\bar{x}_I=\bar{a}_I})}$. Let $P|_{\bar{x}_I=\bar{a}_I} = \sum\limits_{B \subseteq [n] \setminus I} \beta_B \cdot X^B$ be the representation of $P|_{\bar{x}_I=\bar{a}_I}$ as in From 1. Observe that $\beta_B = P_{B,I}|_{\bar{x}_I=\bar{a}_I}$. Now, let $B \in \chi_{P|_I}$. By the definition of $\chi_{P|_I}$ there exists $A \subseteq [n]$ such that $B = A \setminus I$ and $\alpha_A \neq 0$. Set $J \triangleq A \setminus B$. Note that $J \subseteq I$. This implies that $P_{B,I} \not\equiv 0$ as the coefficient of X^J in $P_{B,I}$ is $\alpha_{B \cup J} = \alpha_A \neq 0$. From the choice of $\bar{a}$, it holds that $\beta_B = P_{B,I}|_{\bar{x}_I=\bar{a}_I} \neq 0$ which implies that $B \in \chi_{(P|_{\bar{x}_I=\bar{a}_I})}$. □

As mentioned earlier, the main technical task will be to bound $\|F\|$ in terms of the different $\|F\|_I$-s. For this purpose we use Shearer's Lemma (see e.g. [CGFS86]).

Lemma 2.5 (Shearer)**.** *Let $n \in N$ and let $I_1, \ldots I_m \subseteq [n]$ be subsets of $[n]$ such that every element of $i \in [n]$ is contained in at least k of $I_1, \ldots I_m$. Let $\mathcal{F}$ be a collection of subsets of $[n]$ and let $\mathcal{F}_j \triangleq \{A \cap I_j \;:\; A \in \mathcal{F}\}$ for $j \in [m]$. Then we have $|\mathcal{F}|^k \leq \prod\limits_{j=1}^{m} |\mathcal{F}_j|$.*

The following corollary of Shearer's Lemma connects the sparsity of a polynomial with the sparsities of its different partial substitutions.

Corollary 2.6. *Let $F \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be a polynomial, $d \geq 2$ and $I_1, I_2, \ldots, I_d \subseteq [n]$ disjoint sets. Then $\|F\|^{d-1} \leq \prod_{j=1}^{d} \|F\|_{I_j}$.*

Proof. Set $\mathcal{F} = \chi_F$ and $\mathcal{F}_j = \left\{ A \cap \bar{I}_j \ : \ A \in \mathcal{F} \right\}$, and apply Shearer's lemma. Note that since the I_j-s are disjoint, every $i \in [n]$ appears in every complement set $\bar{I}_j$, except at most one. □

We will need the following recent result of [SV10] that gives an efficient factorization algorithm for sparse multilinear polynomials. In particular we can see that a factor of a multilinear s-sparse polynomial is also a multilinear s-sparse polynomial.

Lemma 2.7 (Corollary from [SV10])**.** *Given a multilinear polynomial $P \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ there is a* $\mathrm{poly}(n, \|P\|)$ *time deterministic algorithm that outputs the* irreducible *factors, $h_1, \ldots, h_k$ of P. Furthermore, $\|h_1\| \cdot \|h_2\| \cdot \ldots \cdot \|h_k\| = \|P\|$.*

The following is a simple property of the sparsity of the gcd of several polynomials. For a proof see Appendix A.

Observation 2.8. *Let $\{F_i\}, \{G_i\} \subseteq \mathbb{F}[x_1, x_2, \ldots, x_n]$ be such that $F_i, G_i \not\equiv 0$. Then*

$$\|\mathrm{gcd}(F_1 \cdot G_1, F_2 \cdot G_2, \ldots, F_k \cdot G_k)\| \leq \|\mathrm{gcd}(F_1, F_2, \ldots, F_k)\| \cdot \|G_1\| \cdot \|G_2\| \cdot \ldots \cdot \|G_k\|.$$

2.2 The Operator D_ℓ

This operator was defined and used in [KMSV10] for the purpose of finding a non-zero assignment of a polynomial, that preserves certain properties. In this paper we extend the usage of this operator to finding *zero* assignments of polynomials. This task of finding zero assignments turns out to be much trickier. In this section we formally define the operator and list some properties that immediately follow (and will be used later).

Definition 2.9. *For $\ell \in [n]$ let $D_\ell(P, Q)$ be the polynomial defined as follows:*

$$D_\ell(P, Q)(\bar{x}) \triangleq \left| \begin{pmatrix} P & P|_{x_\ell=0} \\ Q & Q|_{x_\ell=0} \end{pmatrix} \right| (\bar{x}) = (P \cdot Q|_{x_\ell=0} - P|_{x_\ell=0} \cdot Q)(\bar{x}).$$

Note that D_ℓ is a bilinear transformation. The following lemma lists several useful properties of D_ℓ that are easy to verify.

Lemma 2.10. *Let $P, Q, R \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be multilinear polynomials and let $\ell \in [n]$. Then the following properties hold:*

1. *$D_\ell(R + Q, P) = D_\ell(R, P) + D_\ell(Q, P)$.*
2. *Let R be such that $\ell \notin$ $\mathrm{var}(R)$ then $D_\ell(R \cdot Q, P) = R \cdot D_\ell(Q, P)$.*
3. *$D_\ell(Q, P) = -D_\ell(P, Q)$.*
4. *Let $i \neq \ell$ then $D_\ell(P|_{x_i=\alpha}, Q|_{x_i=\alpha}) = (D_\ell(P, Q))\,|_{x_i=\alpha}$.*

5. *Let $\alpha, \beta \in \mathbb{F}$ then $\|D_\ell(Q, \alpha \cdot x_\ell + \beta)\| \leq \|Q\|$.*

6. *Let P be irreducible and let $\ell \in \mathrm{var}(P)$ then $D_\ell(Q, P) \equiv 0$ iff $P \mid Q$.*

The last property can be easily generalized to yield a condition for a set of polynomials to have a (non-trivial) gcd. This condition was implicitly used in [KMSV10]. For the sake of completeness we give a proof in Appendix A.

Lemma 2.11. *Let $F_1, F_2, \ldots, F_k \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be multilinear polynomials. Then $\gcd(F_1, F_2, \ldots, F_k) \neq 1$ iff there exists $\ell \in \mathrm{var}(F_1)$ such that $D_\ell(F_i, F_1) \equiv 0$ for every $i \in [k]$.*

2.3 Mappings and Generators for Arithmetic Circuits

In this section, we formally define the notion of generators and hitting sets for polynomials and describe a few useful properties. For a further discussion see [SV09, KMSV10].

A mapping $\mathcal{G} = (\mathcal{G}^1, \ldots, \mathcal{G}^n) : \mathbb{F}^q \to \mathbb{F}^n$, is a *generator* for the circuit class $\mathcal{M}$ if for every non-zero n-variate polynomial $P \in \mathcal{M}$, it holds that $P(\mathcal{G}) \not\equiv 0$. The image of the mapping $\mathcal{G}$ is denoted as $\mathrm{Im}\,(\mathcal{G}) = \mathcal{G}(\bar{\mathbb{F}}^q)$. Ideally, q should be very small compared to n. A set $\mathcal{H} \subseteq \mathbb{F}^n$ is a *hitting set* for a circuit class $\mathcal{M}$, if for every non-zero polynomial $P \in \mathcal{M}$, there exists $\bar{a} \in \mathcal{H}$, such that $P(\bar{a}) \neq 0$. A generator can also be viewed as a mapping containing a hitting set for $\mathcal{M}$ in its image. That is, for every non-zero $P \in \mathcal{M}$ there exists $\bar{a} \in \mathrm{Im}\,(\mathcal{G})$ such that $P(\bar{a}) \neq 0$. In identity testing generators and hitting sets play the same role. Given a generator one can easily construct a hitting set by evaluating the generator on a large enough set of points. Conversely, in [SV09] an efficient method of constructing a generator from a hitting set, was given. The following is an immediate and important property of a generator:

Observation 2.12. *Let $P = P_1 \cdot P_2 \cdot \ldots \cdot P_k$ be a product of non-zero polynomials $P_i \in \mathcal{M}$ and let $\mathcal{G}$ be a generator for $\mathcal{M}$. Then $P(\mathcal{G}) \not\equiv 0$.*

The following Lemma from [KMSV10] establishes a generator for multilinear sparse polynomials.

Lemma 2.13 (Lemma 2.7 in [KMSV10])**.** *For every $m \geq 1$ there exists a generator $\mathcal{S}_m \triangleq \left(\mathcal{S}_m^1, \mathcal{S}_m^2, \ldots, \mathcal{S}_m^n\right) : \mathbb{F}^q \to \mathbb{F}^n$ for m-sparse multilinear polynomials where the individual degree of each $\mathcal{S}_m^i$ is bounded by $n-1$ and $q(n,m) = \mathcal{O}(\log_n m)$.*

We conclude this section with a well-known lemma concerning polynomials, giving a trivial (yet possibly large) hitting set. A proof can be found in [Alo99].

Lemma 2.14. *Let $P \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be a polynomial. Suppose that for every $i \in [n]$ the individual degree of x_i is bounded by d_i, and let $S_i \subseteq \mathbb{F}$ be such that $|S_i| > d_i$. We denote $S = S_1 \times S_2 \times \cdots \times S_n$ then $P \equiv 0$ iff $P|_S \equiv 0$.*

3 Depth-4 Multilinear Circuits

In this section, we formally present the model of depth-4 multilinear circuits and some related definitions. Similar definitions were given in [KMSV10].

Definition 3.1. *A multilinear depth-4 $\Sigma\Pi\Sigma\Pi(k)$ circuit C has four layers of alternating Σ and Π gates (the top Σ gate is at level one) and it computes a polynomial of the form*

$$C(\bar{x}) = \sum_{i=1}^{k} F_i(\bar{x}) = \sum_{i=1}^{k} \prod_{j=1}^{d_i} P_{ij}(\bar{x})$$

where the $P_{ij}(\bar{x})$-s are multilinear polynomials computed by the last two layers of $\Sigma\Pi$ gates of the circuit and are the inputs to the Π gates at the second level. In addition, each multiplication gate F_i computes a multilinear polynomial.

The requirement that the F_i-s compute multilinear polynomials implies that for each $i \in [n]$ the polynomials $\{P_{ij}\}_{j\in[d_i]}$ are variable-disjoint. Note that if the circuit is of size s then each P_{ij} is s-sparse. For every $A \subseteq [k]$ we define the subcircuit C_A of C as $C_A \triangleq \sum_{i\in A} F_i$. We define $\gcd(C) \triangleq \gcd(F_1, \ldots, F_k)$. We say that the circuit C is *simple* if $\gcd(C) = 1$. We define the *simplification* of C to be $\mathrm{sim}(C) \triangleq C/\gcd(C)$. Note that $\mathrm{sim}(C)$ is a simple $\Sigma\Pi\Sigma\Pi(k)$ circuit. We say that the circuit C is *minimal* if no proper subcircuit of C computes the zero polynomial. That is, for every $\emptyset \subsetneq A \subsetneq [k]$ it holds that $C_A \not\equiv 0$. For $P \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ we say that the circuit C is *P-minimal* if no proper subcircuit of C is divisible by P.

Lemma 2.7 implies that all the irreducible factors of the P_{ij}-s are s-sparse. Moreover, if C is given to us explicitly those factors can be computed efficiently. Consequently, we can assume w.l.o.g that all the P_{ij}-s are irreducible.

4 The Sparsity Bound

In this section we prove the main technical result (Theorem 1) an upper bound on the sparsity of the multiplication gates in a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit C computing the zero polynomial. To complete the picture we also give a lower bound on the gate's sparsity.

We first present an outline of the proof. As mentioned earlier, we assume w.l.o.g that all P_{ij}-s are irreducible and use the circuit size s to bound their sparsities (i.e. $\|P_{ij}\| \leq s$). The proof is by induction on k (the fan-in).

Step 1: We show that for every $\emptyset \subsetneq A \subsetneq [k]$, $\gcd(C_A)$ is $s^{5(k-|A|+1)^2}$-sparse. We do so by "embedding" $\gcd(C_A)$ into a circuit with a smaller k and applying the inductive argument. In particular, we conclude that for every $i \neq j$ it holds that $\gcd(F_i, F_j)$ is $s^{5(k-1)^2}$-sparse.

Step 2: Applying the previous argument, we conclude for each $1 \leq i \leq k-1$ there are "many" P_{kj}-s that do not divide F_i. More specifically, if all the gates of C are s^{5k^2}-sparse, then we are done. Otherwise, w.l.o.g F_k is s^{5k^2}-dense. Let $1 \leq i \leq k-1$. Write $F_k = F'_k \cdot \gcd(F_i, F_k)$. By definition, F'_k is the product of all the P_{kj}-s that do not divide F_i. As F_k is s^{5k^2}-dense and $\gcd(F_i, F_k)$ is $s^{5(k-1)^2}$-sparse we get that F'_k is $s^{5k^2-5(k-1)^2} = s^{10k-5}$-dense. Finally, we observe that since each P_{kj} is s-sparse, F'_k must be a product of at least $10k-5$ of them.

Step 3: For every $1 \leq i \leq k-1$ and j such that P_{kj} does not divide F_i we find a zero assignment $\bar{a}$ of P_{kj} that preserves certain properties of C (see discussion in Section 1.1). In particular, $\bar{a}$

maximizes the sparsity of F_i' - the polynomial resulting from a substitution of $\bar{a}$ into F_i. Afterwards, using the inductive argument we obtain a "good" estimation for the sparsity of F_i'. Formally, we show that F_i' is $s^{5(k-1)^2+k+19}$-sparse. This analysis is the heart of the proof of the sparsity bound.

Step 4: Based on the information collected in Step 3 we use Shearer's Lemma to show that for every $1 \leq i \leq k-1$, F_i is s^{5k^2-1}-sparse. As $-F_k = F_1 + F_2 + \ldots + F_{k-1}$ we can upper bound the sparsity of F_k by the sum of the sparsities of F_i-s, for $1 \leq i \leq k-1$. We conclude that F_k is $s^{5k^2-1} \cdot (k-1) < s^{5k^2}$-sparse, thus reaching a contradiction to our assumption. The "large" number of distinct partial substitutions used in Shearer's Lemma makes our bound nearly optimal.

We now give a more formal statement and then prove Theorem 1.

Theorem 4.1 (The Sparsity Bound)**.** *There exists an non-decreasing function $\varphi(k,s)$ such that if $C(\bar{x}) = \sum_{i=1}^{k} F_i(\bar{x})$ is a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit of size s computing the zero polynomial, then for each $i \in [k]$ it holds that $\|F_i\| \leq \varphi(k,s)$ and $\varphi(k,s) \leq s^{5k^2}$.*

Proof. The proof is by induction on k. The base case is $k = 2$. Note that in this case C must be of the form $C = \alpha - \alpha$ for some $\alpha \in \mathbb{F}$. Therefore, $\|F_1\| = \|F_2\| = 1$. Assume that $k \geq 3$.

We first state and prove some lemmas that will be useful for the proof. Note that all lemma are proven as part of the inductive argument we apply in the proof of the theorem. We start by showing an upper bound on the sparsity of the gcd of any subcircuit of a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit computing the zero polynomial. We do so by 'embedding' the gcd as a multiplication gate into a $\Sigma\Pi\Sigma\Pi$ circuit of a smaller fan-in. Informally, we do this be setting all the variables that do not appear in the gcd to field elements in such a way that the simplicity and minimality of the circuit is preserved (and hence induction can be applied). By setting the variables, the subcircuit corresponding to the gcd collapses to a single multiplication gate which is just a scaled version of the gcd term. In fact, setting those variables to *random* field elements would work with high probability if the underlying field is large enough. Below we present a formal argument.

Lemma 4.2. *Let $C(\bar{x}) = \sum_{i=1}^{k} F_i(\bar{x})$ be a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit of size s computing the zero polynomial and let $G \triangleq \gcd(F_1, F_2, \ldots, F_t)$ for some $2 \leq t \leq k-1$. Then $\|G\| \leq \varphi(k-t+1, s)$.*

Proof. Denote $V = [n] \setminus \mathrm{var}(G)$ and $F_i = f_i \cdot G$ for $i \in [t]$. Observe that $\mathrm{var}(f_i) \subseteq V$ for each f_i. We will show now that there exists a partial assignment $\bar{a} \in \overline{\mathbb{F}}^n$ to $\bar{x}_V$ that preserves the properties of minimality and the simplicity of the circuit. For that purpose, recalling Definition 2.9, we define the polynomial:

$$\Phi = \prod_{\emptyset \subsetneq A \subsetneq [k]} C_A \cdot \prod_{\ell, i \,:\, D_\ell(F_i, F_1) \not\equiv 0} D_\ell(F_i, F_1).$$

Let $\bar{a} \in \overline{\mathbb{F}}^n$ such that $\Phi|_{\bar{x}_V = \bar{a}_V} \not\equiv 0$. Let $F_i' \triangleq F_i|_{\bar{x}_V = \bar{a}_V}$ for $i \in [k]$. Consider $C' = C|_{\bar{x}_V = \bar{a}_V} \triangleq \sum_{i=1}^{k} F_i'$. From the definition of $\bar{a}$ it follows that C' is minimal since $C_A' = C_A|_{\bar{x}_V = \bar{a}_V} \not\equiv 0$ for every $\emptyset \subsetneq$

$A \subsetneq [k]$. Next, we argue that C' is simple. As C is simple we have that $\gcd(F_1, F_2, \ldots, F_k) = 1$. Therefore, Lemma 2.11 implies that for every $\ell \in \text{var}(F_1)$ there exists $i \in [k]$ such that $D_\ell(F_i, F_1) \not\equiv 0$. From the choice of $\bar{a}$ we obtain that for every $\ell \in \text{var}(F_1)$ there exists $i \in [k]$ such that $D_\ell(F'_i, F'_1) \not\equiv 0$, and thus, by the second direction of Lemma 2.11 $\gcd(F'_1, F'_2, \ldots, F'_k) = 1$. Now, set $H_1 \triangleq \sum_{i=1}^{t} F'_i$ and $H_i \triangleq F'_{t+i-1}$ for $2 \leq i \leq k-t+1$. Recall that $\text{var}(f_i) \subseteq V$ for each $i \in [t]$, therefore

$$H_1 = \sum_{i=1}^{t} F_i|_{\bar{x}_V=\bar{a}_V} = \sum_{i=1}^{t} (f_i \cdot G)\,|_{\bar{x}_V=\bar{a}_V} = \left(\sum_{i=1}^{t} f_i|_{\bar{x}_V=\bar{a}_V}\right) \cdot G = \alpha \cdot G$$

for some $\alpha \in \mathbb{F}$. We now can define $\hat{C} \triangleq \sum_{i=1}^{k-t+1} H_i = \alpha \cdot G + \sum_{i=t+1}^{k} F'_i$ - the circuit obtained from C' by joining together the first t summands. Indeed, we embedded G into a circuit with a smaller fan-in. We argue $\hat{C}$ satisfies the required properties so we can apply the induction hypothesis. First, note that $\hat{C} \equiv 0$. The minimality of $\hat{C}$ follows from the minimality of C'. In particular, $\alpha \cdot G = H_1 = C'_{[t]} \not\equiv 0$ and thus $\alpha \neq 0$. Finally, observe that

$$\gcd(\hat{C}) = \gcd(\alpha \cdot G, F'_{t+1}, \ldots, F'_k) = \gcd(F'_1, \ldots, F'_t, F'_{t+1}, \ldots, F'_k) = \gcd(C')$$

and hence $\hat{C}$ is simple. We can conclude that $\hat{C}$ is a simple and minimal, $\Sigma\Pi\Sigma\Pi(k-t+1)$ circuit of size s computing the zero polynomial. In addition, note that $2 \leq k-t+1 \leq k-1$. Consequently, we can apply the induction hypothesis on $\hat{C}$. We obtain that $\|G\| = \|H_1\| \leq \varphi(k-t+1, s)$. □

Next is a technical lemma that will allow us to decrease, in some sense, the top fan-in of the circuit. This step is required in order to use the inductive argument. Recall that a P-minimal circuit is one where no proper subcircuit is divisible by P.

Lemma 4.3. *Let $C(\bar{x}) = \sum_{i=1}^{k} F_i(\bar{x})$ be a multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit computing the zero polynomial and let $P \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be a factor of F_k (i.e. $P \mid F_k$) such that $P \nmid F_1$. Then there exists a set $A \subseteq [k]$ of size $2 \leq |A| \leq k-1$ such that the following holds: $1 \in A$, the subcircuit $C_A(\bar{x}) = \sum_{i \in A} F_i(\bar{x})$ is P-minimal and $P \mid C_A$.*

Proof. As C computes the zero polynomial we have that $P \mid C$. Therefore, C can be partitioned into subcircuits that are minimal w.r.t. this property. Formally, there exists a partition $\bigcup_i A_i = [k]$, $A_i \bigcap A_j = \emptyset$ such that for every i the subcircuit C_{A_i} is P-minimal and $P \mid C_{A_i}$. Let w.l.o.g A_1 be such that $1 \in A_1$. It is only left to show that $2 \leq |A_1| \leq k-1$. First of all, note that since $P \mid F_k$ there must be $A_j = \{k\}$ for some $j \neq i$ and hence $|A_1| \leq k-1$. For the second condition, note that since $P \nmid F_1$ it must be the case that $A_1 \neq \{1\}$ and hence $|A_1| \geq 2$. □

Finally, we give the heart of our argument. The lemma shows how to transform a given simple and P-minimal circuit C, such that $P \mid C$, into a simple and minimal circuit of a smaller fan-in, computing the zero polynomial. As before, the transformation is carried out by finding a partial assignment $\bar{a}$ to $\bar{x}_{\text{var}(P)}$ that preserves the minimality and the simplicity of C. However, the most important property of $\bar{a}$ is that it maximizes the sparsity of the circuit, resulting upon the partial substitution into $\bar{x}_{\text{var}(P)}$. This fact allows us to apply Shearer's Lemma. It can be easily seen that $P(\bar{a}) = 0$. To find the required assignment we present a new technique (see Sec 2.2).

Lemma 4.4. *Let $2 \leq t \leq k-1$. Let $P \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be a non-constant, irreducible, multilinear polynomial and let $C(\bar{x}) = \sum\limits_{i=1}^{t} F_i(\bar{x})$ be a simple and P-minimal, multilinear $\Sigma\Pi\Sigma\Pi(t)$ circuit of size s such that $P \mid C$. Then $\|F_1\|_{\mathrm{var}(P)} \leq \varphi(t,s) \cdot s^t$.*

Proof. Our goal is to upperbound $\|F_1\|_{\mathrm{var}(P)}$. As previously, we wish to do so by an appropriate embedding. What we would like to do is to fix the variables in $\mathrm{var}(P)$ in a way that will make P evaluate to zero, and at the same time, will result in a simple and minimal circuit. This way the circuit would have fewer multiplication gates, and we could apply induction. Unfortunately, this scenario might not be possible to implement. We will look to approximate it paying a 'small price'.

Pick $\ell \in \mathrm{var}(P)$. For each $i \in [t]$ we can write $F_i = H_i \cdot Q_i$ such that $\|Q_i\| \leq s$ and $\ell \notin \mathrm{var}(H_i)$: If $\ell \in \mathrm{var}(F_i)$ set Q_i to be the irreducible factor of F_i that depends on ℓ, otherwise set $Q_i = 1$. Now, recalling Definition 2.9, consider $C' = D_\ell(C, P) \overset{\Delta}{=} \sum\limits_{i=1}^{t} D_\ell(F_i, P)$. By Lemma 2.10

$$C' = \sum_{i=1}^{t} H_i \cdot D_\ell(Q_i, P) \equiv 0.$$

In addition, note that C' is minimal since $C'_A = D_\ell(C_A, P) \not\equiv 0$ for every $\emptyset \subsetneq A \subsetneq [t]$ from the P-minimality of C and Lemma 2.10. By definition, the polynomials H_i and Q_i are variable-disjoint. However, this might not be the case for P and H_i. Consequently, C' might be non-multilinear. Furthermore, $\|D_\ell(Q_i, P)\|$ might be large. In order to resolve the aforementioned problems we will use a partial assignment $\bar{a}$ with properties similar to the ones in Lemma 4.2. Let $V = \mathrm{var}(P) \setminus \{\ell\}$. We recall Definition 2.3 and define:

$$\Phi = \prod_{\emptyset \subsetneq A \subsetneq [t]} C'_A \cdot \prod_{j,i \,:\, D_j(F_i, F_1) \not\equiv 0} D_j(F_i, F_1) \cdot \Psi_{H_1}.$$

Let $\bar{a} \in \overline{\mathbb{F}}^n$ such that $\Phi|_{\bar{x}_V = \bar{a}_V} \not\equiv 0$. Set $H'_i \overset{\Delta}{=} H_i|_{\bar{x}_V = \bar{a}_V}$ and $Q'_i \overset{\Delta}{=} D_\ell(Q_i, P)|_{\bar{x}_V = \bar{a}_V}$ for $i \in [t]$. Consider $C'' = C'|_{\bar{x}_V = \bar{a}_V} \overset{\Delta}{=} \sum\limits_{i=1}^{k} H'_i \cdot Q'_i$. By a reasoning similar to Lemma 4.2 we obtain that C'' is minimal and that $\gcd(H'_1, H'_2, \ldots, H'_t) = 1$. By Lemma 2.10 for each $i \in [t]$ we have:

$$Q'_i = D_\ell(Q_i|_{\bar{x}_V = \bar{a}_V}, P|_{\bar{x}_V = \bar{a}_V}) = D_\ell(Q_i|_{\bar{x}_V = \bar{a}_V}, \alpha_P \cdot x_\ell + \beta_P)$$

for some $\alpha_P, \beta_P \in \mathbb{F}$ and hence $\|Q'_i\| \leq \|Q_i\| \leq s$. Consequently, we get that C'' is a minimal, multilinear $\Sigma\Pi\Sigma\Pi(t)$ circuit of size s computing the zero polynomial. In addition, from the choice of $\bar{a}$ and Lemma 2.4 we obtain (recall that $\ell \notin \mathrm{var}(H_1)$):

$$\|H'_1\| = \|H_i|_{\bar{x}_V = \bar{a}_V}\| = \|H_1\|_{\mathrm{var}(P) \setminus \{\ell\}} = \|H_1\|_{\mathrm{var}(P)}.$$

By the induction hypothesis applied on the simplification $\mathrm{sim}(C'')$ of C'':

$$\|H'_1 \cdot Q'_1 / \gcd(C'')\| \leq \varphi(t,s).$$

While by Observation 2.8

$$\|\gcd(C'')\| \leq \|\gcd(H'_1, \ldots, H'_t)\| \cdot \|Q'_1\| \cdot \ldots \cdot \|Q'_t\| \leq s^{t-1} \cdot \|Q'_1\|.$$

Putting the above together we obtain:

$$\|F_1\|_{\mathrm{var}(P)} = \|H_1\|_{\mathrm{var}(P)} \cdot \|Q_1\|_{\mathrm{var}(P)} \leq \|H_1'\| \cdot s \leq \varphi(t,s) \cdot \frac{\|\gcd C''\|}{\|Q_1'\|} \cdot s \leq \varphi(t,s) \cdot s^t.$$

□

We now return to the proof Theorem 4.1.
Assume for a contradiction (and w.l.o.g) that $\|F_k\| > s^{5k^2}$. We will show that this implies $\|F_i\| \leq s^{5k^2-1}$ for $1 \leq i \leq k-1$. As $\sum_{i=1}^{k} F_i \equiv 0$ we would obtain that

$$\|F_k\| = \left\|\sum_{i=1}^{k-1} F_i\right\| \leq \sum_{i=1}^{k-1} \|F_i\| < (k-1) \cdot s^{5k^2-1} < s^{5k^2}$$

thus leading us to a contradiction.

For the sake of simplicity, we show that the claim holds for $i = 1$ (i.e. $\|F_1\| \leq s^{5k^2-1}$). Note however, the same proof can be repeated for every $1 \leq i \leq k-1$ due to the symmetry of the circuit. We first show that there are "many" (irreducible) P_{kj}-s such that $P_{kj} \not| F_1$. Recall that we can assume w.l.o.g. that P_{ij}-s are irreducible. For that purpose we define $F_k' \triangleq F_k / \gcd(F_1, F_k)$. Let w.l.o.g. $P_{k1}, P_{k2}, \ldots, P_{kd}$ be its irreducible factors, that is, $F_k' = P_{k1} \cdot P_{k2} \cdot \ldots \cdot P_{kd}$. By definition, each such P_{kj} divides F_k and does **not** divide F_1. By Lemma 4.2 $\|\gcd(F_1, F_k)\| \leq s^{5(k-1)^2}$ and hence:

$$s^d \geq \|P_{k1} \cdot P_{k2} \cdot \ldots \cdot P_{kd}\| = \|F_k'\| = \frac{\|F_k\|}{\|\gcd(F_1, F_k)\|} \geq s^{5k^2 - 5(k-1)^2}$$

implying that $d \geq 5k^2 - 5(k-1)^2 = 10k - 5$. Fix some some $j \in [d]$ and consider P_{kj}. By Lemma 4.3 there exists a set $A \subseteq [k]$ of size $2 \leq |A| \leq k-1$ such that $1 \in A$, the sub-circuit $C_A(\bar{x})$ is P_{kj}-minimal, and $P_{kj} \mid C_A$. Assume w.l.o.g that $A = \{1, 2, \ldots, t\}$ when $t = |A|$. Let $G \triangleq \gcd(F_1, F_2, \ldots, F_t)$. By Lemma 4.2.

$$\|G\| \leq \varphi(k-t+1, s) \leq s^{5(k-t+1)^2}.$$

From the P_{kj}-minimality of C_A we have that for each $i \in [t]$ $P_{kj} \not| F_i$ and hence $P_{kj} \not| G$. Consequently, we obtain that $\mathrm{sim}(C_A) \triangleq \sum_{i=1}^{t} F_i/G$ is a simple and P_{kj}-minimal, multilinear $\Sigma\Pi\Sigma\Pi(t)$ circuit such that $P_{kj} \mid \mathrm{sim}(C_A)$ (as the simplification does not affect the P_{kj}-minimality). Thus, by Lemma 4.4:

$$\|F_1/G\|_{\mathrm{var}(P_{kj})} \leq \varphi(t,s) \cdot s^t \leq s^{5t^2+t}.$$

Putting together:

$$\|F_1\|_{\mathrm{var}(P_{kj})} \leq \|F_1/G\|_{\mathrm{var}(P_{kj})} \cdot \|G\| \leq s^{5(k-t+1)^2+t+5t^2} \leq s^{5(k-1)^2+k+19}.$$ [3]

[3] For $k \geq 3$ and $2 \leq t \leq k-1$ it holds that: $5(k-t+1)^2 + t + 5t^2 \leq 5(k-1)^2 + k + 19$.

Recall that this inequality holds for every P_{kj} when $1 \leq j \leq 10k-5 \leq d$. In addition, recall that P_{kj}-s are variable-disjoint polynomials (as factors of a multilinear polynomial). Hence, we can upper bound $\|F_1\|$ using Corollary 2.6.

$$\|F_1\| \leq \left(\prod_{j=1}^{10k-5} \|F_1\|_{\mathrm{var}(P_{kj})} \right)^{\frac{1}{10k-6}} \leq s^{\left(5(k-1)^2+k+19\right)\cdot\frac{10k-5}{10k-6}} < s^{5k^2-1}.^{4}$$

As it was previously stated, the above inequality holds, in fact, for every F_i when $1 \leq i \leq k-1$. Therefore, since $F_k = -\sum\limits_{i=1}^{k-1} F_i$ it holds that $\|F_k\| < s^{5k^2}$. In conclusion we obtain that $\|F_i\| \leq s^{5k^2}$ for each $i \in [k]$. □

4.1 Lower Bound

To give a complete picture, we show that lower bound of $\varphi(k,s) = s^{\Omega(k)}$ on the sparsity of the multiplication gates in a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit, in terms of the size of the circuit. Our lower bound is over sufficiently large fields and it suggests that our result is near optimal. More specifically, for every $\ell \geq 2$ we construct a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit of size $s = \mathrm{poly}(k)$, computing the zero polynomial with $s^{\Omega(k)}$-dense multiplication gates, when $k = \Omega(\sqrt{\ell})$. In fact, our $\Sigma\Pi\Sigma\Pi(k)$ circuit is a $\Sigma\Pi\Sigma(k)$. Our construction is carried out in two steps. At the first step, we construct a multilinear $\Sigma\Pi\Sigma(\ell)$ circuit C with distinct linear functions and show that it computes the zero polynomial. The distinct linear functions imply that every subcircuit of C is simple. At the second step, we consider one of C's minimal subcircuits - C' and use our upper bound to show that C' has a "large" fan-in. A similar example was given in [SV09].

Lemma 4.5. *For every $\ell \geq 2$ there exist $k = \Omega(\sqrt{\ell})$ and a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit $C = \sum\limits_{i=1}^{k} F_i$ of size $s = \mathrm{poly}(k)$, computing the zero polynomial such that $\|F_i\| = s^{\Omega(k)}$ for every $i \in [k]$. (Assuming that $|\mathbb{F}| \geq \ell + 2$).*

Proof. Let $A = \{\alpha_1, \alpha_2, \ldots, \alpha_{\ell+1}\} \subseteq \mathbb{F} \setminus \{0\}$ be a subset of $\ell+1$ distinct non-zero elements. For every $i \in [\ell+1]$ let $u_i(w)$ be the i-th Lagrange Interpolation Polynomial over A. Let $R : \mathbb{F}^{\ell^2+1} \to \mathbb{F}$ be defined as: $R(\bar{x}, y) = \prod\limits_{i=1}^{\ell} (x_{i,1} + x_{i,2} + \ldots + x_{i,\ell} + y)$. Since the degree of y in $R(\bar{x}, y)$ is ℓ we get that $R(\bar{x}, y) = \sum\limits_{i=1}^{\ell+1} u_i(y) \cdot R(\bar{x}, \alpha_i)$, by interpolating $R(\bar{x}, y)$ as a degree ℓ polynomial in y. Now consider:

$$C \triangleq \sum_{i=1}^{\ell+1} u_i(0) \cdot R(\bar{x}, \alpha_i) - R(\bar{x}, 0)$$

By definition, C is a multilinear $\Sigma\Pi\Sigma\Pi(\ell+2)$ circuit computing the zero polynomial. Let $C' = \sum\limits_{i=1}^{k} F_i$ be a minimal subcircuit of C computing the zero polynomial. Note that for distinct α_i-s the multiplication gates (F_i-s) contain distinct linear functions and that $u_i(0) \neq 0$ for every i. Consequently, C' is a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit computing the zero

[4] For $k \geq 3$ it holds that: $\left(5(k-1)^2+k+19\right) \cdot \frac{10k-5}{10k-6} < 5k^2 - 1$.

polynomial, with $2 \le k \le \ell+2$. Let s denote the size of C. Clearly, $s = \mathrm{poly}(k, \ell) = \ell^{\Theta(1)}$. Now, let $i \in [k]$. On one hand, we have that $\|F_i\| \ge \ell^\ell = s^{\Omega(\ell)}$. On the other hand, by Theorem 1 $\|F_i\| = s^{\mathcal{O}(k^2)}$. This implies that $\|F_i\| = s^{\Omega(k)}$ and that $k = \Omega(\sqrt{\ell})$, which in turn implies that $s = \mathrm{poly}(k)$, as required. □

5 Black-Box PIT for Multilinear $\Sigma\Pi\Sigma\Pi(k)$ Circuits

In this section we give an efficient black-box PIT algorithm for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits. We do so by constructing a generator for such circuits, which gives us a small hitting set. We start by describing the construction. Intuitively, we construct a family of mappings $H_{\ell,m}$ such that the image of $H_{\ell,m}$ contains all vectors which are obtained as a concatenation of a prefix of a vector from $\mathrm{Im}\left(H_{\ell-1,m}\right)$ and a suffix of a vector from $\mathcal{S}_m$ (a generator for m-sparse multilinear polynomials. See Lemma 2.13). The correctness of our construction relies on the Sparsity Bound $\varphi(k, s)$, as defined in Theorem 4.1.

We assume that $|\mathbb{F}| > n$ as we are allowed to use elements from an appropriate extension field. Throughout the entire section we fix a set $A = \{\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n\} \subseteq \mathbb{F}$ of $n+1$ distinct elements.

Definition 5.1. *For every $i \in [n]$ let $U_i(y) : \mathbb{F} \to \mathbb{F}$ be defined as the degree n polynomial satisfying: $U_i(\alpha_j) = 1$ if $j \ge i$ and 0 otherwise. For every $\ell \ge 1$ and $m \ge 1$ we define: for $i \in [n]$ $H^i_{\ell,m}(\bar{w}_1, \dots, \bar{w}_\ell, y_1, \dots, y_\ell) : \mathbb{F}^{q\cdot\ell+\ell} \to \mathbb{F}$ as*

$$H^i_{\ell,m}(\bar{w}_1, \dots, \bar{w}_\ell, y_1, \dots, y_\ell) \stackrel{\Delta}{=} H^i_{\ell-1,m}(\bar{w}_1, \dots, \bar{w}_{\ell-1}, y_1, \dots, y_{\ell-1}) \cdot U_i(y_\ell) + \mathcal{S}^i_m(\bar{w}_\ell) \cdot (1 - U_i(y_\ell)).$$

and $H_{\ell,m}(\bar{w}_1, \dots, \bar{w}_\ell, y_1, \dots, y_\ell) : \mathbb{F}^{q\cdot\ell+\ell} \to \mathbb{F}^n$ as $H_{\ell,m} \stackrel{\Delta}{=} \left(H^1_{\ell,m}, H^2_{\ell,m}, \dots, H^n_{\ell,m}\right)$.

For the sake of completeness we set $H^i_{0,m} \equiv 0$. We will use the following immediate but crucial observation:

Observation 5.2. *For every $0 \le t \le n$, it holds that*

$$H_{\ell,m}|_{y_\ell = \alpha_t} = \left(H^1_{\ell-1,m}, \dots, H^t_{\ell-1,m}, \mathcal{S}^{t+1}_m, \dots \mathcal{S}^n_m\right)$$

and hence, for every $\bar{a} \in \mathrm{Im}\left(H_{\ell-1,m}\right)$ and $\bar{b} \in \mathrm{Im}\left(\mathcal{S}_m\right)$ it holds that

$$(a_1, \dots, a_t, b_{t+1}, \dots, b_n) \in \mathrm{Im}\left(H_{\ell,m}\right).$$

In particular, $\mathrm{Im}\left(H_{\ell-1,m}\right) \cup \mathrm{Im}\left(\mathcal{S}_m\right) \subseteq \mathrm{Im}\left(H_{\ell,m}\right)$.

To use the construction, we first show that every "non sparse" circuit can be "shrunk" into a "somewhat sparse" circuit.

Lemma 5.3. *Let $M \ge 1$ and let $C(\bar{x}) = \sum_{i=1}^{k} F_i(\bar{x}) = \sum_{i=1}^{k} \prod_{j=1}^{d_i} P_{ij}(\bar{x})$ be a multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit of size s such that $\max_i \|F_i\| > M$. Let $\bar{a} \in \mathbb{F}^n$ such that $F_i(\bar{a}) \neq 0$ for each $i \in [k]$. Then there exists $0 \le t \le n-1$ such that $M < \max_i \|F_i|_{\bar{x}_{[t]} = \bar{a}_{[t]}}\| \le M \cdot s$*

Proof. We apply the hybrid argument since $\max_i \|F_i\| > M$ and $\max_i \|F_i|_{\bar{x}_{[n]}=\bar{a}_{[n]}}\| \leq 1 \leq M$. Let $0 \leq t \leq n-1$ be the maximal index such that $\max_i \|F_i|_{\bar{x}_{[t]}=\bar{a}_{[t]}}\| > M$. From the choice of t we have that $\max_i \|F_i|_{\bar{x}_{[t+1]}=\bar{a}_{[t+1]}}\| \leq M$. For the remaining condition, note that for each $i \in [k]$ the polynomial $F_i|_{\bar{x}_{[t+1]}=\bar{a}_{[t+1]}}$ is obtained from $F_i|_{\bar{x}_{[t]}=\bar{a}_{[t]}}$ by fixing the value of x_{t+1} to a_{t+1}. As F_i is multilinear and $F_i(\bar{a}) \neq 0$ this fixation can affect at most one P_{ij} in it, and hence, reduce the maximal sparsity by a factor of at most $\|P_{ij}\| \leq s$. □

Next, we use our structure theorem to guarantees that in the process of shrinking, we do not inadvertently end up making a non-zero circuit into a zero circuit, thus allowing an inductive step.

Lemma 5.4. *Let $k \geq 2$ and $Q \not\equiv 0 \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be a polynomial computed by a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit $C(\bar{x}) = \sum_{i=1}^{k} F_i(\bar{x})$ of size s. In addition, let $\mathcal{G}_{k-1}$ be a generator for $\Sigma\Pi\Sigma\Pi(k-1)$ circuits of size s and $(2s^2)$-sparse polynomials. Then there exists $\bar{a} \in \operatorname{Im}(\mathcal{G}_{k-1})$ and $0 \leq t \leq n-1$, such that $Q' \triangleq Q|_{\bar{x}_{[t]}=\bar{a}_{[t]}}$ is a non-zero, $\big(\varphi(k,s) \cdot s^2\big)$-sparse polynomial.*

Proof. If $\max_i \|F_i\| \leq \varphi(k,s)$, then clearly $\|Q\| \leq \varphi(k,s) \cdot k \leq \varphi(k,s) \cdot s^2$. Suppose $\max_i \|F_i\| > \varphi(k,s)$. We define the following polynomial:

$$\Phi = \prod_{\emptyset \subsetneq A \subsetneq [k]} C_A \cdot \prod_{\ell, i \,:\, D_\ell(F_i, F_1) \not\equiv 0} D_\ell(F_i, F_1)$$

From the properties of D_ℓ (Lemma 2.10) we get that all multiplicands of Φ are either $(2s^2)$-sparse polynomials or $\Sigma\Pi\Sigma\Pi(k-1)$ circuits. Therefore, by Observation 2.12 we have that $\Phi(\mathcal{G}_{k-1}) \not\equiv 0$ and consequently, there exists $\bar{a} \in \operatorname{Im}(\mathcal{G}_{k-1})$ such that $\Phi(\bar{a}) \neq 0$. By definition, $F_i(\bar{a}) \neq 0$ for each $i \in [k]$. Thus, by Lemma 5.3 there exists $0 \leq t \leq n-1$ such that

$$\varphi(k,s) < \max_i \|F_i|_{\bar{x}_{[t]}=\bar{a}_{[t]}}\| \leq \varphi(k,s) \cdot s$$

Consider the circuit $C' \triangleq C|_{\bar{x}_{[t]}=\bar{a}_{[t]}} = \sum_{i=1}^{k} F_i|_{\bar{x}_{[t]}=\bar{a}_{[t]}}$. By a reasoning similar to Lemma 4.2 we obtain that C' is simple and minimal. We now argue that $C' \not\equiv 0$. Assume the contrary. By Theorem 4.1 we get that $\max_i \|F_i|_{\bar{x}_{[t]}=\bar{a}_{[t]}}\| \leq \varphi(k,s)$, which leads us to a contradiction. Finally, note that $\|Q'\| \leq \varphi(k,s) \cdot s \cdot k \leq \varphi(k,s) \cdot s^2$. □

Having the above, we use a generator to work around the problems raised by lack of simplicity or minimality. This idea has been previously used in [KMSV10].

Theorem 5.5 ($H_{k,m}$ is a Generator)**.** *Let $k \geq 1$, $s \geq 2$. Let $Q \not\equiv 0 \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be a polynomial computed by a multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit $C(\bar{x}) = \sum_{i=1}^{k} F_i(\bar{x}) = \sum_{i=1}^{k} \prod_{j=1}^{d_i} P_{ij}(\bar{x})$ of size s. Then for every $m \geq \varphi(k,s) \cdot s^2$ it holds that $Q(H_{k,m}) \not\equiv 0$.*

Proof. We apply induction on k. For $k = 1$ note that Q is a product of s-sparse polynomials. Thus, the claim follows from Observations 5.2 and 2.12, and properties of $\mathcal{S}_m$ (Lemma 2.13). Assume that $k \geq 2$. We will argue that we can assume w.l.o.g that C is simple and minimal. Suppose C is not minimal. Then there exists a $\Sigma\Pi\Sigma\Pi(k')$ circuit C' with $k' < k$ computing Q. Therefore, by the induction hypothesis it holds that $Q(H_{k',m}) \not\equiv 0$; and hence, by Observation 5.2 we get that $Q(H_{k,m}) \not\equiv 0$. Now, suppose C is not simple. As previously, we assume w.l.o.g that P_{ij} are irreducible and write: $Q \equiv G \cdot C'$, when $G = \gcd(C)$ and $C' = \mathrm{sim}(C)$. By definition, G is a product of s-sparse polynomials and thus, as previously, $G(H_{k,m}) \not\equiv 0$. Therefore, it is sufficient to show that $C'(H_{k',m}) \equiv 0$. Recall that $\mathrm{sim}(C)$ is a simple $\Sigma\Pi\Sigma\Pi(k)$ circuit. Due to the above, we can assume w.l.o.g that C is simple and minimal. By the induction hypothesis $H_{k-1,m}$ is a generator for $\Sigma\Pi\Sigma\Pi(k-1)$ circuits of size s. In addition, it is a generator for $(2s^2)$-sparse polynomials (Observation 5.2). By Lemma 5.4 there exists $\bar{a} \in \mathrm{Im}\left(H_{k-1,m}\right)$ and $0 \leq t \leq n-1$, such that $Q' \triangleq Q|_{\bar{x}_{[t]}=\bar{a}_{[t]}}$ is a non-zero, $\left(\varphi(k,s) \cdot s^2\right)$-sparse polynomial. Being such, there exists $\bar{b} \in \mathrm{Im}\left(\mathcal{S}_m\right)$ for which $Q'(\bar{b}) \neq 0$, or equivalently $Q(a_1, \ldots, a_t, b_{t+1}, \ldots, b_n) = Q'(\bar{b}) \neq 0$. Observation 5.2 completes the proof. □

We conclude by giving an explicit construction a hitting set for polynomials Q computed by $\Sigma\Pi\Sigma\Pi(k)$ circuits. The idea is that $Q(H_{k,m})$ is a non-zero polynomial depending on a small number of variables $r = q \cdot k + k$, with individual degrees less than $n^2 + 1$. Consequently, it is sufficient to evaluate Q on the set $(V^q)^k \times V^k$.

```
Input: $n, k, s \geq 1$.
Output: A set $\mathcal{H}$
Let $V \subseteq \mathbb{F}$ be of size $|V| = n^2 + 1$;
Set $\mathcal{H} \triangleq H_{k\ ,\ \varphi(k,s)\cdot s^2}\left((V^q)^k \times V^k\right)$ ;
```

Algorithm 1: Construction of a hitting set for $\Sigma\Pi\Sigma\Pi(k)$ circuits of size s.

Theorem 5.6. *Given n, s, k as input, Algorithm 1 runs in time $n^{\mathcal{O}(k)} \cdot s^{\mathcal{O}(k^3)}$ and outputs $\mathcal{H}$ of size $n^{\mathcal{O}(k)} \cdot s^{\mathcal{O}(k^3)}$, which is a hitting set for n-variate polynomials that can be computed by multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits of size s.*

Proof. Let $Q \not\equiv 0 \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be a polynomial computed by a multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit of size s. Let $\mathcal{H}$ be the set given by Algorithm 1. By Theorem 5.5 we get that $Q\left(H_{k\ ,\ \varphi(k,s)\cdot s^2}\right)$ is a non-zero polynomial depending on $r = q \cdot k + k$ variables, with individual degrees less than $n^2 + 1$ (the degrees of y's and w are at most n and Q is multilinear). Consequently, Lemma 2.14 implies that $Q|_{\mathcal{H}} \not\equiv 0$. For the size of $\mathcal{H}$. Recall Lemma 2.13 and that $\varphi(k,s) \leq s^{5k^2}$. We obtain: $|\mathcal{H}| \leq |V|^{k+q\cdot k} = n^{\mathcal{O}(k)} \cdot \left(n^{\mathcal{O}(k^2 \cdot \log_n s)}\right)^k = n^{\mathcal{O}(k)} \cdot s^{\mathcal{O}(k^3)}$. □

6 Non Black-Box PIT for Multilinear $\Sigma\Pi\Sigma\Pi(k)$ Circuits

In this section we give a non black-box PIT algorithm for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits. This running time of the algorithm is slightly better then the running of the black-box one.

6.1 Overview

Given Theorem 1 it is fairly easy to check if a simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit C computes the zero polynomial. More specifically, let $C = \sum_{i=1}^{k} F_i = \sum_{i=1}^{k} \prod_{j=1}^{d_i} P_{ij}$ and let s

denote the size of C. First, compute the sparsity of each F_i by noting that $\|F_i\| = \prod_{j=1}^{d_i} \|P_{ij}\|$. Now, if there exists $i \in [k]$ for which $\|F_i\| > s^{5k^2}$ then by Theorem 1 $C \not\equiv 0$. Otherwise, C computes a $(k \cdot s^{5k^2})$-sparse polynomial and hence we can check if $C \equiv 0$ by computing its monomial expansion. However, an arbitrary multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit C may be neither minimal nor simple. The idea is first to transform the circuit C into a simple and minimal circuit C' such that $C \equiv 0$ iff $C' \equiv 0$, and afterwards apply the procedure described above. In fact, we are going to "simplify" and "minimize" C. The "minimization" is carried out by (recursively) checking if all the proper subcircuits of C compute the zero polynomial. For the "simplification" of C, suppose that all P_{ij} were irreducible. Under this assumption, each multiplication gate F_i contains among its P_{ij}-s the exact same list of irreducible polynomials (up to multiplication by a field element) forming the gcd of C (if $\gcd(C) \neq 1$). Therefore, we can erase those polynomials and obtain a simple circuit. Although the irreducibility of the P_{ij}-s is a valid assumption for the analysis, this might not be the case in the given circuit. We handle this scenario by factorizing each P_{ij} to its irreducible factors. The factorization is carried out using a recent result of [SV10] (Lemma 2.7). Note that factorization does not affect the sparsity of F_i-s and can only decrease the size of the circuit.

6.2 The Algorithm

We now present the algorithm. First, we start with an algorithm that simplifies a given $\Sigma\Pi\Sigma\Pi(k)$ circuit.

Lemma 6.1. *There is a deterministic algorithm that when given as input a multilinear* $\Sigma\Pi\Sigma\Pi(k)$ *circuit* C *of size* s *on* n *variables runs in time* $\mathrm{poly}(n, s)$ *and outputs a simple* $\Sigma\Pi\Sigma\Pi(k)$ *circuit* C' *of size* s *such that* $C \equiv 0$ *iff* $C' \equiv 0$.

Proof. Given $C = \sum_{i=1}^{k} \prod_{j=1}^{d_i} P_{ij}$ we factorize each P_{ij} to its irreducible factors in time $\mathrm{poly}(n, s)$ using the algorithm in Lemma 2.7 (recall that $\|P_{ij}\| \leq s$). We then replace each P_{ij} by the product of its irreducible factors, obtaining the circuit by $C_{red} = \sum_{i=1}^{k} \prod_{j=1}^{d'_i} P'_{ij}$. Observe that the size of C_{red} is at most s. As there are at most s different P_{ij}-s, the total running time of this step is $\mathrm{poly}(n, s)$. By definition, all P'_{ij}-s are irreducible. Therefore each multiplication gate of C_{red} contains among its P'_{ij}-s the exact same set of irreducible polynomials (up to multiplication by a field element) forming the gcd of C_{red}. Let C' be the circuit resulting from erasing the mentioned set of polynomials. Clearly, C' is a simple $\Sigma\Pi\Sigma\Pi(k)$ circuit of size s. The polynomials computed by C and C' differ by a multiplicative factor of $\gcd(C_{red})$. Consequently, $C \equiv 0$ if and only if $C' \equiv 0$. □

Finally, we present our PIT algorithm for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits, thus proving Theorem 2.

Lemma 6.2. *Given a multilinear* $\Sigma\Pi\Sigma\Pi(k)$ *circuit* C *of size* s *on* n *variables Algorithm 2 runs in time* $\mathrm{poly}(n) \cdot s^{\mathcal{O}(k^2)}$ *and outputs "true" if and only if* $C \equiv 0$.

Proof. We begin the correctness analysis by induction on k. For $k = 1$ the claim is clear. Let $k \geq 2$. Assume the correctness for smaller values of k. By Lemma 6.1 $C \equiv 0$ iff $C' \equiv 0$. If there exists $\emptyset \subsetneq A \subsetneq [k]$ such that $C'_A \equiv 0$ then, clearly $C' \equiv 0$ iff $C'_{[k]\setminus A} \equiv 0$. Otherwise, C' is simple and minimal, multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuit of size s with $k \geq 2$. By Theorem 4.1 if $C' \equiv 0$ then $\|F'_i\| \leq s^{5k^2}$ for every $i \in [k]$. Therefore, if there exists $i \in [k]$ such that $\|F'_i\| > s^{5k^2}$, then $C \not\equiv 0$. The last step is correct by its definition.

```
Input: $\Sigma\Pi\Sigma\Pi(k)$ circuit $C$ of size $s$
Output: "true" iff $C \equiv 0$

if $k = 1$ then
    return "true" iff $F_1 \equiv 0$

Compute $C' = \sum_{i=1}^{k} F'_i$  /* using Lemma 6.1.                */
foreach $\emptyset \subsetneq A \subsetneq [k]$ do
    if $C'_A \equiv 0$ then
        return "true" iff $C'_{[k]\setminus A} \equiv 0$

for $i = 1$ to $k$ do
    Compute $\|F'_i\|$
    if $\|F'_i\| > s^{5k^2}$ then
        return "false"

return "true" iff $C' \equiv 0$, by computing the monomial expansion of $C'$
```

Algorithm 2: Non-Black PIT algorithm for multilinear $\Sigma\Pi\Sigma\Pi(k)$ circuits

Time complexity: By Lemma 6.1 C' can be computed in time $\mathrm{poly}(n, s)$. Next, as $\|F'_i\| \leq 2^n$ for each $i \in [k]$ the computation of $\|F'_i\|$ can be done in time $\mathrm{poly}(n, s)$. Finally, note that the last line is reached only in the case that every F'_i is s^{5k^2}-sparse. Therefore, the polynomial computed by C' in this case has at most $k \cdot s^{5k^2}$ monomials. Consequently, this step takes $\mathrm{poly}(n) \cdot s^{\mathcal{O}(k^2)}$ time. Putting all together we obtain the following recurrent relation: $T(k, n, s) \leq 2^{k-1} \cdot T(k-1, n, s) + \mathrm{poly}(n) \cdot s^{\mathcal{O}(k^2)}$ and hence $T(k, n, s) = \mathrm{poly}(n) \cdot s^{\mathcal{O}(k^2)}$. □

Acknowledgments

The authors would like to thank Amir Shpilka for many useful discussions and lots of helpful advice.

References

[Agr05] M. Agrawal. Proving lower bounds via pseudo-random generators. In *Proceedings of the 25th FSTTCS*, volume 3821 of *LNCS*, pages 92–105, 2005.

[AKS04] M. Agrawal, N. Kayal, and N. Saxena. Primes is in P. *Annals of Mathematics*, 160(2):781–793, 2004.

[ALM$^+$98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *JACM*, 45(3):501–555, 1998.

[Alo99] N. Alon. Combinatorial nullstellensatz. *Combinatorics, Probability and Computing*, 8:7–29, 1999.

[AM10] V. Arvind and P. Mukhopadhyay. The monomial ideal membership problem and polynomial identity testing. *Information and Computation*, 208(4):351–363, 2010.

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *JACM*, 45(1):70–122, 1998.

[AV08] M. Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *Proceedings of the 49th Annual FOCS*, pages 67–75, 2008.

[AvMV11] M. Anderson, D. van Melkebeek, and I. Volkovich. Derandomizing polynomial identity testing for multilinear constant-read formulae. In *Proceedings of the 26rd Annual IEEE Conference on Computational Complexity, CCC*, 2011.

[BOT88] M. Ben-Or and P. Tiwari. A deterministic algorithm for sparse multivariate polynominal interpolation. In *Proceedings of the 20th Annual STOC*, pages 301–309, 1988.

[CGFS86] F R Chung, R L Graham, P Frankl, and J B Shearer. Some intersection theorems for ordered sets and graphs. *J. Comb. Theory Ser. A*, 43(1):23–37, 1986.

[DS06] Z. Dvir and A. Shpilka. Locally decodable codes with 2 queries and polynomial identity testing for depth 3 circuits. *SIAM J. on Computing*, 36(5):1404–1434, 2006.

[DSY09] Z. Dvir, A. Shpilka, and A. Yehudayoff. Hardness-randomness tradeoffs for bounded depth arithmetic circuits. *SIAM J. on Computing*, 39(4):1279–1293, 2009.

[HS80] J. Heintz and C. P. Schnorr. Testing polynomials which are easy to compute (extended abstract). In *Proceedings of the 12th annual STOC*, pages 262–272, 1980.

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

[KMSV10] Z. S. Karnin, P. Mukhopadhyay, A. Shpilka, and I. Volkovich. Deterministic identity testing of depth 4 multilinear circuits with bounded top fan-in. In *Proceedings of the 42nd Annual STOC*, pages 649–658, 2010.

[KS01] A. Klivans and D. Spielman. Randomness efficient identity testing of multivariate polynomials. In *Proceedings of the 33rd Annual STOC*, pages 216–223, 2001.

[KS07] N. Kayal and N. Saxena. Polynomial identity testing for depth 3 circuits. *Computational Complexity*, 16(2):115–138, 2007.

[KS08] Z. S. Karnin and A. Shpilka. Deterministic black box polynomial identity testing of depth-3 arithmetic circuits with bounded top fan-in. In *Proceedings of the 23rd Annual CCC*, pages 280–291, 2008.

[KS09] N. Kayal and S. Saraf. Blackbox polynomial identity testing for depth 3 circuits. In *Proceedings of the 50th Annual FOCS*, pages 198–207, 2009.

[LFKN92] C. Lund, L. Fortnow, H. Karloff, and N. Nisan. Algebraic methods for interactive proof systems. *JACM*, 39(4):859–868, 1992.

[Lov79] L. Lovasz. On determinants, matchings, and random algorithms. In L. Budach, editor, *Fundamentals of Computing Theory*. Akademia-Verlag, 1979.

[LV03] R. J. Lipton and N. K. Vishnoi. Deterministic identity testing for multivariate polynomials. In *Proceedings of the 14th annual SODA*, pages 756–760, 2003.

[MVV87] K. Mulmuley, U. Vazirani, and V. Vazirani. Matching is as easy as matrix inversion. *Combinatorica*, 7(1):105–113, 1987.

[Raz06] R. Raz. Separation of multilinear circuit and formula size. *Theory of Computing*, 2(1):121–135, 2006.

[Raz09] R. Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size. *J. ACM*, 56(2), 2009.

[RSY08] R. Raz, A. Shpilka, and A. Yehudayoff. A lower bound for the size of syntactically multilinear arithmetic circuits. *SIAM J. on Computing*, 38(4):1624–1647, 2008.

[RY09] R. Raz and A. Yehudayoff. Lower bounds and separations for constant depth multilinear circuits. *Computational Complexity*, 18(2):171–207, 2009.

[Sax08] N. Saxena. Diagonal circuit identity testing and lower bounds. In *ICALP*, pages 60–71, 2008.

[Sch80] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, 1980.

[Sha90] Adi Shamir. IP=PSPACE. In *Proceedings of the Thirty First Annual Symposium on Foundations of Computer Science*, pages 11–15, 1990.

[SS09] N. Saxena and C. Seshadhri. An almost optimal rank bound for depth-3 identities. In *Proceedings of the 24th annual CCC*, pages 137–148, 2009.

[SS10] N. Saxena and C. Seshadhri. From sylvester-gallai configurations to rank bounds: Improved black-box identity test for deph-3 circuits. In *Proceedings of the 51st Annual FOCS*, pages 21–30, 2010.

[SS11] N. Saxena and C. Seshadhri. Blackbox identity testing for bounded top fanin depth-3 circuits: the field doesn't matter. In *Proceedings of the 43th Annual STOC*, 2011.

[SV09] A. Shpilka and I. Volkovich. Improved polynomial identity testing for read-once formulas. In *APPROX-RANDOM*, pages 700–713, 2009.

[SV10] A. Shpilka and I. Volkovich. On the relation between polynomial identity testing and finding variable disjoint factors. In *ICALP*, pages 408–419, 2010.

[SY10] A. Shpilka and A. Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3-4):207–388, 2010.

[Zip79] R. Zippel. Probabilistic algorithms for sparse polynomials. In *Symbolic and algebraic computation*, pages 216–226. 1979.

A Missing Proofs

For the sake of completeness we give here some of the missing proofs.

Proof of Lemma 2.11. For the first direction, assume $\gcd(F_1, F_2, \ldots, F_k) \neq 1$. Then, by definition, there exists an irreducible polynomial P such that $P \mid F_i$ for every $i \in [k]$. Equivalently, we can write $F_i = F_i' \cdot P$. Pick $\ell \in \mathrm{var}(P)$. Clearly, $\ell \in \mathrm{var}(F_1)$ and hence $\ell \not\in \mathrm{var}(F_i')$ since the factors of a multilinear polynomial are variable-disjoint. From Lemma 2.10 we get that $D_\ell(F_i, F_1) = D_\ell(F_i' \cdot P, F_1' \cdot P) = F_i' \cdot F_1' \cdot D_\ell(P, P) \equiv 0$. For the second direction, let $\ell \in \mathrm{var}(F_1)$. We can write $F_1 = F_1' \cdot P$, when P denotes the irreducible factor of F_1 that depends on ℓ. From the statement $F_1' \cdot D_\ell(F_i, P) = D_\ell(F_i, F_1) \equiv 0$. As $F_1' \not\equiv 0$, we get that $D_\ell(F_i, P) \equiv 0$ for every $i \in [k]$. It follows from Lemma 2.10 that $P \mid F_i$ for every $i \in [k]$, or equivalently $P \mid \gcd(F_1, F_2, \ldots, F_k)$. In particular this implies $\gcd(F_1, F_2, \ldots, F_k) \neq 1$ and completes the proof. □

Proof Observation 2.8. Denote $R = \gcd(F_1 \cdot G_1, F_2 \cdot G_2, \ldots, F_k \cdot G_k)$, $G_0 = \gcd(F_1, F_2, \ldots, F_k)$ and consider $P \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ an irreducible factor of R. By definition $P \mid F_i \cdot G_i$ for each $i \in [k]$. Therefore, for each $i \in [k]$ it must be the case that either $P \mid F_i$ or $P \mid G_i$ holds. Now, if $P \mid F_i$ for all i-s, then by definition $P \mid G_0$. Otherwise, there must exist $j \in [k]$ such that $P \mid G_j$. Consequently, $P \mid G_0 \cdot G_1 \cdot G_2 \cdot \ldots \cdot G_k$. Since the above holds for every irreducible factor of R we obtain that $R \mid G_0 \cdot G_1 \cdot G_2 \cdot \ldots \cdot G_k$. Equivalently, there exists Q such that $R \cdot Q = G_0 \cdot G_1 \cdot G_2 \cdot \ldots \cdot G_k$. and hence $\|R\| \leq \|G_0\| \cdot \|G_1\| \cdot \ldots \cdot \|G_k\|$. □

We also show that is inequality is tight. Let $\Phi \overset{\Delta}{=} \prod_{i=1}^{k}(x_i + 1)$. Take: $G_i = (x_i + 1), F_i = \Phi / G_i$. Then $\gcd(F_1 \cdot G_1, F_2 \cdot G_2, \ldots, F_k \cdot G_k) = \Phi = G_1 \cdot G_2 \cdot \ldots \cdot G_k \cdot \gcd(F_1, F_2, \ldots, F_k)$. Note that $\gcd(F_1, F_2, \ldots, F_k) = 1$.