FROM SYLVESTER-GALLAI CONFIGURATIONS TO RANK BOUNDS: IMPROVED BLACK-BOX IDENTITY TEST FOR DEPTH-3 CIRCUITS

NITIN SAXENA AND C. SESHADHRI

ABSTRACT. We study the problem of identity testing for depth-3 circuits of top fanin k and degree d (called $\Sigma\Pi\Sigma(k,d)$ identities). We give a new structure theorem for such identities. A direct application of our theorem improves the known deterministic $d^{k^{O(k)}}$-time black-box identity test over rationals (Kayal & Saraf, FOCS 2009) to one that takes $d^{O(k^2)}$-time. Our structure theorem essentially says that the number of independent variables in a real depth-3 identity is very small. This theorem settles affirmatively the stronger rank conjectures posed by Dvir & Shpilka (STOC 2005) and Kayal & Saraf (FOCS 2009). Our techniques provide a unified framework that actually beats all known rank bounds and hence gives the best running time (for *every field*) for black-box identity tests.

Our main theorem (almost optimally) pins down the relation between higher dimensional Sylvester-Gallai theorems and the rank of depth-3 identities in a very transparent manner. The existence of this was hinted at by Dvir & Shpilka (STOC 2005), but first proven, for reals, by Kayal & Saraf (FOCS 2009). We introduce the concept of *Sylvester-Gallai rank bounds* for any field, and show the intimate connection between this and depth-3 identity rank bounds. We also prove the first ever theorem about high dimensional Sylvester-Gallai configurations over *any field*. Our proofs and techniques are very different from previous results and devise a very interesting ensemble of combinatorics and algebra. The latter concepts are ideal theoretic and involve a new Chinese remainder theorem. Our proof methods explain the structure of *any* depth-3 identity C: there is a *nucleus* of C that forms a low rank identity, while the remainder is a high dimensional Sylvester-Gallai configuration.

1. INTRODUCTION

Polynomial identity testing (PIT) ranks as one of the most important open problems in the intersection of algebra and computer science. We are provided an arithmetic circuit that computes a polynomial $p(x_1, x_2, \cdots, x_n)$ over a field $\mathbb{F}$, and we wish to test if p is identically zero (in other words, if p is the zero polynomial). In the black-box setting, the circuit is provided as a black-box and we are only allowed to evaluate the polynomial p at various domain points. The main goal is to devise a deterministic polynomial time algorithm for PIT. Kabanets & Impagliazzo [KI04] and Agrawal [Agr05, Agr06] have shown connections between deterministic algorithms for identity testing and circuit lower bounds, emphasizing the importance of this problem. To know more about the current state of the general identity testing problem see the surveys [Sax09, AS09].

The first randomized polynomial time PIT algorithm, which was a black-box algorithm, was given (independently) by Schwartz [Sch80] and Zippel [Zip79]. Randomized algorithms that use less randomness were given by Chen & Kao [CK00], Lewin & Vadhan [LV98], and Agrawal & Biswas [AB03]. Klivans & Spielman [KS01] observed that even for depth-3 circuits for bounded top fanin, deterministic identity testing was open. Progress towards

Hausdorff Center for Mathematics, Bonn - 53115, Germany. ns@hcm.uni-bonn.de.
IBM Almaden, San Jose - 95126, USA. csesha@gmail.com.

this was first made by Dvir & Shpilka [DS06], who gave a quasi-polynomial time algorithm, although with a doubly-exponential dependence on the top fanin. The problem was resolved by a polynomial time algorithm given by Kayal and Saxena [KS07], with a running time exponential in the top fanin. As expected, the current understanding of depth-4 circuits is even more sparse. Identity tests are known only for rather special depth-4 circuits [AM07, Sax08, SV09, KMSV09]. Why is progress restricted to such small depth circuits? Agrawal and Vinay [AV08] showed that an efficient black-box identity test for depth-4 circuits will actually give a quasi-polynomial black-box test, and subexponential lower bounds, for circuits of *all depths* (that compute *low degree* polynomials). Thus, understanding depth-3 identities seems to be a natural first step towards the goal of proving more general lower bounds.

For deterministic black-box testing, the first results were given by Karnin & Shpilka [KS08]. Based on results in [DS06], they gave an algorithm for bounded top fanin depth-3 circuits having a quasi-polynomial running time (with a doubly-exponential dependence on the top fanin). The dependence on the top fanin was later improved (to singly-exponential) by the rank bound results of Saxena & Seshadhri [SS09] (for *any* $\mathbb{F}$). But the time complexity also had a quasi-polynomial dependence on the degree of the circuit. This dependence is inevitable in rank-based methods over finite fields (as shown by [KS07]). However, over the field of rationals, Kayal & Saraf [KS09b] showed how to remove this quasi-polynomial dependence on the degree at the cost of doubly-exponential dependence on the top fanin, thus giving a polynomial time complexity for bounded top fanin. In this work we achieve the best of the two works [SS09] and [KS09b], i.e. we prove (for rationals) a time complexity that depends only *polynomially* on the degree and "only" *singly*-exponentially on the fanin.

In a quite striking result, Kayal & Saraf [KS09b] proved how Sylvester-Gallai theorems can get better rank bounds over the reals. We introduce the concept of *Sylvester-Gallai rank bounds* that deals with the rank of vectors (over some given field) that have some special incidence properties. This is a very convenient way to express known Sylvester-Gallai results. These are inspired by the famous Sylvester-Gallai theorem about point-line incidences. We show how this very interesting quantity is tightly connected to depth-3 identities. Sylvester-Gallai rank bounds over high dimensions were known over the reals, and are used to prove depth-3 rank bounds over reals. We prove the first ever theorem for high dimensional Sylvester-Gallai configurations over *any field.*

1.1. **Definitions and Previous Work.** This work focuses on depth-3 circuits. A structural study of depth-3 identities was initiated in [DS06] by defining a notion of *rank* of *simple* and *minimal* identities. A depth-3 circuit C over a field $\mathbb{F}$ is:

$$C(x_1, \ldots, x_n) = \sum_{i=1}^{k} T_i$$

where, T_i (*a multiplication term*) is a product of d_i linear polynomials $\ell_{i,j}$ over $\mathbb{F}$. Note that for the purposes of studying identities we can assume wlog (by *homogenization*) that $\ell_{i,j}$'s are linear *forms* (i.e. linear polynomials with a zero constant coefficient) and that $d_1 = \cdots = d_k =: d$. Such a circuit is referred to as a $\Sigma\Pi\Sigma(k, d)$ circuit (or $\Sigma\Pi\Sigma(k, d, n)$ depending on the context), where k is the *top fanin* of C and d is the *degree* of C. We give a few definitions from [DS06].

Definition 1. [Simple Circuit] C *is a* simple *circuit if there is no nonzero linear form dividing all the T_i's.*

[**Minimal Circuit**] *C is a* minimal *circuit if for every proper subset* $S \subset [k]$, $\sum_{i \in S} T_i$ *is nonzero.*

[**Rank of a circuit**] *Every* $\ell_{i,j}$ *can be seen as an n-dimensional vector over* $\mathbb{F}$. *The* rank *of the circuit, rk(C), is defined as the rank of the set of all linear forms* $\ell_{i,j}$*'s viewed as n-dimensional vectors.*

Can all the forms $\ell_{i,j}$ be independent, or must there be relations between them? The rank can be interpreted as the minimum number of variables that are required to express C. There exists a linear transformation converting the n variables of the circuit into $rank(C)$ independent variables. A trivial upper bound on the rank (for any $\Sigma\Pi\Sigma$-circuit) is kd, since that is the total number of linear forms involved in C. The rank is a fundamental property of a $\Sigma\Pi\Sigma(k,d)$ circuit and it is crucial to understand how large this can be for identities. A substantially smaller rank bound than kd shows that identities do not have as many "degrees of freedom" as general circuits, and leads to deterministic identity tests. Furthermore, the techniques used to prove rank bounds show us structural properties of identities that may suggest directions to resolve PIT for $\Sigma\Pi\Sigma(k,d)$ circuits.

The rank bounds, in addition to being a natural property of identities, have found applications in black-box identity testing [KS08] and learning $\Sigma\Pi\Sigma$ circuits [Shp09, KS09a]. The result of [KS08] showed rank bounds imply black-box testers: if $R(\mathbb{F},k,d)$ is a rank bound for simple minimal $\Sigma\Pi\Sigma(k,d,n)$ identities over field $\mathbb{F}$, then there is a deterministic black-box identity tester for such circuits, that runs in $\text{poly}(n, d^{R(\mathbb{F},k,d)})$ $\mathbb{F}$-operations. (For the time complexity over $\mathbb{Q}$, we actually count the *bit operations.*)

Dvir & Shpilka [DS06] proved that the rank of a simple, minimal $\Sigma\Pi\Sigma(k,d)$ identity is bounded by $2^{O(k^2)}(\log d)^{k-2}$. This rank bound was improved to $O(k^3 \log d)$ by Saxena & Seshadhri [SS09]. Fairly basic identity constructions show that the rank is $\Omega(k)$ over the reals and $\Omega(k \log d)$ for finite fields [DS06, KS07, SS09]. Dvir & Shpilka [DS06] conjectured that $\text{rk}(C)$ should be some $\text{poly}(k)$ over the reals. Through a very insightful use of Sylvester-Gallai theorems, Kayal & Saraf [KS09b] subsequently bounded the rank of identities, over reals, by $k^{O(k)}$. This means that for a constant top fanin circuit, the rank of identities is constant, independent of the degree. This also leads to the first truly polynomial-time deterministic black-box identity testers for this case.

Unfortunately, as soon as k becomes even $\Omega(\log n)$, this bound becomes trivial. We improve this rank bound exponentially, to $O(k^2)$, which is almost optimal. This gives a major improvement in the running time of the black-box testers. We also improve the rank bounds for general fields from $O(k^3 \log d)$ to $O(k^2 \log d)$. We emphasize that we give a unified framework to prove all these results. Table 1 should make it easier to compare the various bounds.

Kayal & Saraf [KS09b] connect Sylvester-Gallai theorems to rank bounds. They need advanced versions of these theorems that deal with colored points and have to prove certain *hyperplane decomposition theorems*. We make the connection much more transparent (at the loss of some color from the theorems). We reiterate that our techniques are completely different, and employ a very powerful algebraic framework to dissect identities. This allows us to use as a "black-box" the most basic form of the higher dimensional Sylvester-Gallai theorems.

1.2. **Our Results.** Before we state our results, it will be helpful to understand Sylvester-Gallai configurations. A set of points S with the property that every line through two points of S passes through a third point in S is called a *Sylvester-Gallai configuration*. The famous Sylvester-Gallai theorem states: for a set S of points in $\mathbb{R}^2$, not all collinear,

TABLE 1. Known rank bounds and black-box PIT

Paper	Result	Asymptotics	over field
[DS06, KS08]	rank bound	$2^{k^2} \log^{k-2} d$	any
	time complexity	$nd^{(2^{k^2} \log^{k-2} d)}$	any
[SS09]	rank bound	$k^3 \log d$	any
	time complexity	$nd^{k^3 \log d}$	any
[KS09b]	rank bound	k^k	$\mathbb{R}$
	time complexity	$nd^{(k^k)}$	$\mathbb{Q}$
Ours	rank bound	k^2	$\mathbb{R}$
	time complexity	nd^{k^2}	$\mathbb{Q}$
	rank bound	$k^2 \log d$	any
	time complexity	$nd^{k^2 \log d}$	any

there exists a line passing through exactly two points of S. In other words, the only Sylvester-Gallai configuration in $\mathbb{R}^2$ is a set of collinear points. This basic theorem about point-line incidences was extended to higher dimensions [Han65, BE67]. We introduce the notion of *Sylvester-Gallai rank bounds*. This is a clean and convenient way of expressing these theorems.

Definition 2. *Let S be a finite subset of the* projective space $\mathbb{FP}^n$. *Alternately, S is a subset of vectors in $\mathbb{F}^{n+1}$ without* multiples: *no two vectors in S are scalar multiples of each other[1]. Suppose, for every set $V \subset S$ of k linearly independent vectors, the linear span of V contains at least $k+1$ vectors of S. Then, the set S is said to be* SG_k-*closed.*
The largest possible rank of an SG_k-closed set of at most m vectors in $\mathbb{F}^n$ (for any n) is denoted by $SG_k(\mathbb{F}, m)$.

The classic Sylvester-Gallai theorem essentially states[2] that for all m, $\mathrm{SG}_2(\mathbb{R}, m) \leqslant 2$. Higher dimensional analogues [Han65, BE67] prove that $\mathrm{SG}_k(\mathbb{R}, m) \leq 2(k-1)$. One of our auxiliary theorems is such a statement for *all fields*.

Theorem 3 (SG_k for all fields). *For any field $\mathbb{F}$ and $k, m \in \mathbb{N}^{>1}$, $SG_k(\mathbb{F}, m) \leq 9k \lg m$.*

Our main theorem is a simple, clean expression of how Sylvester-Gallai influences identities.

Theorem 4 (From SG_k to Rank). *Let $|\mathbb{F}| > d$. The rank of a simple and minimal $\Sigma\Pi\Sigma(k, d)$ identity over $\mathbb{F}$ is at most $2k^2 + k \cdot SG_k(\mathbb{F}, d)$.*

Remark. If $\mathbb{F}$ is small, then we choose an extension $\mathbb{F}' \supset \mathbb{F}$ of size $> d$ and get a rank bound with $\mathrm{SG}_k(\mathbb{F}', d)$.

---

[1] When $|\mathbb{F}| > |S|$, such an S is, wlog, a subset of distinct vectors with first coordinate 1.

[2] To see this, take an SG_2-closed set S of vectors. Think of each vector being represented by an infinite line through the origin, hence giving a set S in the projective space. Take a 2-dimensional plane P not passing through the origin and take the set of intersection points I of the lines in S with P. Observe that the coplanar points I have the property that a line passing through two points of I passes through a third point of I.

Plugging in SG_k-rank bounds gives us the desired theorem for depth-3 identities. We have a slightly stronger version of the above theorem that we use to get better constants (refer to Theorem 18).

Theorem 5 (Depth-3 Rank Bounds)**.** *Let C be a $\Sigma\Pi\Sigma(k,d)$ circuit, over field $\mathbb{F}$, that is simple, minimal and zero. Then,*

- *For $\mathbb{F} = \mathbb{R}$, $rk(C) < 3k^2$.*
- *For any $\mathbb{F}$, $rk(C) < 3k^2(\lg 2d)$.*

As discussed before, a direct application of this result to Lemma 4.10 of [KS08] gives a deterministic black-box identity test for $\Sigma\Pi\Sigma(k,d,n)$ circuits (we will only discuss $\mathbb{Q}$ here as the other statement is analogous). Formally, we get the following *hitting set generator* for $\Sigma\Pi\Sigma$ circuits with real coefficients.

Corollary 6 (Black-box PIT over $\mathbb{Q}$)**.** *There is a deterministic algorithm that takes as input a triple (k,d,n) of natural numbers and in time $poly(nd^{k^2})$, outputs a hitting set $\mathcal{H} \subset \mathbb{Z}^n$ with the following properties:*

1) *Any $\Sigma\Pi\Sigma(k,d,n)$ circuit C over $\mathbb{R}$ computes the zero polynomial iff $\forall a \in \mathcal{H}$, $C(a) = 0$.*
2) *$\mathcal{H}$ has at most $poly(nd^{k^2})$ points.*
3) *The total bit-length of each point in $\mathcal{H}$ is $poly(kn \log d)$.*

Remark.

1) Our black-box test has *quasi-polynomial* in n time complexity (with polynomial-dependence on d) for top fanin as large as $k = \operatorname{poly} \log(n)$, and *sub-exponential* in n time complexity (with polynomial-dependence on d) even for top fanin as large as $k = o(\sqrt{n})$. This is the first tester to achieve such bounds.
2) The fact that the points in $\mathcal{H}$ are integral and have "small" bit-length is important to estimate the time complexity of our algorithm in terms of *bit operations*. Thus, the hitting set generator takes at most $poly(nd^{k^2})$ *bit* operations to compute $\mathcal{H}$.

2. Proof Outline, Ideas, and Organization

Our proof of the rank bound comprises of several new ideas, both at the conceptual and the technical levels. In this section we will give the basic intuition of the proof. The three notions that are crucially used (or developed) in the proof are: ideal Chinese remaindering, matchings and Sylvester-Gallai rank bounds. These have appeared (in some form) before in the works of Kayal & Saxena [KS07], Saxena & Seshadhri [SS09] and Kayal & Saraf [KS09b] respectively, to prove different kinds of results. Here we use all three of them together to show quite a strong structure in $\Sigma\Pi\Sigma$ identities. We will talk about them one by one in the following three subsections outlining the three steps of the proof. Each step proves a new property of identities which is interesting in its own right. The first two steps set up the algebraic framework and prove theorems that hold for all fields. The third step is where the Sylvester-Gallai theorems are brought in. Some (new and crucial) algebraic lemmas and their proofs have been moved to the Appendix. The flow of the actual proof will be identical to the overview that we now provide.

2.1. **Step 1: Matching the Gates in an Identity.** We will denote the set $\{1, \ldots, n\}$ by $[n]$. We fix the base field to be $\mathbb{F}$, so the circuits compute multivariate polynomials in the *polynomial ring* $R := \mathbb{F}[x_1, \ldots, x_n]$.

A *linear form* is a linear polynomial in R with zero constant term. We will denote the set of all linear forms by $L(R) := \{\sum_{i=1}^{n} a_i x_i \mid a_1, \ldots, a_n \in \mathbb{F}\}$. Clearly, $L(R)$ is a vector (or linear) space over $\mathbb{F}$ and that will be quite useful. Much of what we do shall deal with *multi*-sets of linear forms (sometimes polynomials in R too), equivalence classes inside them, and various maps across them. A *list* of linear forms is a multi-set of forms with an arbitrary order associated with them. The actual ordering is unimportant: we will heavily use maps between lists, and the ordering allows us to define these maps unambiguously. The object, list, comes with all the usual set operations naturally defined.

Definition 7. *We collect some important definitions from* [SS09]:

[**Multiplication term,** $L(\cdot)$ **&** $M(\cdot)$] *A multiplication term f is an expression in R given as (the product may have repeated ℓ's), $f := c \cdot \prod_{\ell \in S} \ell$, where $c \in \mathbb{F}^*$ and S is a list of nonzero linear forms. The* list of linear forms in f, $L(f)$, *is just the list S of forms occurring in the product above. For a list S of linear forms we define the* multiplication term of S, $M(S)$, *as $\prod_{\ell \in S} \ell$ or 1 if $S = \phi$.*

[**Forms in a Circuit**] *We will represent a $\Sigma\Pi\Sigma(k,d)$ circuit C as a sum of k multiplication terms of degree d, $C = \sum_{i=1}^{k} T_i$. The list of* linear forms occurring in C is $L(C) := \bigcup_{i \in [k]} L(T_i)$. *Note that $L(C)$ is a list of size exactly kd. The* rank of C, $rk(C)$, *is just the number of linearly independent linear forms in $L(C)$. (Remark: for the purposes of this paper T_i's are given in circuit representation and thus the list $L(T_i)$ is unambiguously defined from C)*

[**Similar forms**] *For any two polynomials $f, g \in R$ we call f* similar *to g if there exists $c \in \mathbb{F}^*$ such that $f = cg$. We say f is* similar *to g mod I, for some ideal I of R, if there is some $c \in \mathbb{F}^*$ such that $f \equiv cg \pmod{I}$. Note that "similarity mod I" is an equivalence relation (reflexive, symmetric and transitive) and partitions any list of polynomials into equivalence classes.*

[**Span** $\mathbf{sp}(\cdot)$] *For any $S \subseteq L(R)$ we let $sp(S) \subseteq L(R)$ be the* linear span *of the linear forms in S over the field $\mathbb{F}$. (Conventionally, $sp(\emptyset) = \{0\}$.)*

[**Matchings**] *Let U, V be lists of linear forms and I be a subspace of $L(R)$. An I-matching π between U, V is a bijection π between lists U, V such that: for all $\ell \in U$, $\pi(\ell) \in \mathbb{F}^* \ell + I$.*

When f, g are multiplication terms, an I-matching between f, g would mean an I-matching between $L(f), L(g)$.
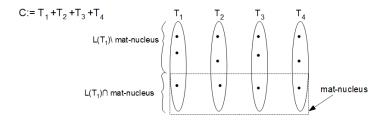
We will show that all the multiplication terms of a minimal $\Sigma\Pi\Sigma$ identity can be matched by a "low" rank space.

Theorem 8 (Matching-Nucleus)**.** *Let $C = T_1 + \cdots + T_k$ be a $\Sigma\Pi\Sigma(k,d)$ circuit that is minimal and zero. Then there exists a linear subspace K of $L(R)$ such that:*
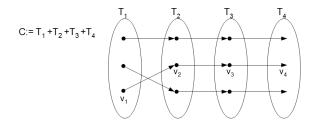
1) $rk(K) < k^2$.

2) $\forall i \in [k]$, there is a K-matching π_i between T_1, T_i.

The idea of matchings within identities was first introduced in [SS09], but nothing as powerful as this theorem has been proven. This theorem gives us a space of small rank, *independent of d*, that contains most of the "complexity" of C. All forms in C outside K are just mirrored in the various terms. This starts connecting the algebra of depth-3 identities to a combinatorial structure. Indeed, the graphical picture (explained in detail below) that this theorem provides, really gives an intuitive grasp on these identities. The proof of this involves some interesting generalizations of the Chinese Remainder Theorem to some special ideals.

Definition 9 (mat-nucleus). *Let C be a minimal $\Sigma\Pi\Sigma(k,d)$ identity. The linear subspace K given by Theorem 8 is called* mat-nucleus *of C.*

The notion of mat-nucleus is easier to see in the following unusual representation of the $\Sigma\Pi\Sigma(4,d)$ circuit $C = \sum_{i\in[4]} T_i$. The four bubbles refer to the four multiplication terms of C and the points inside the bubbles refer to the linear forms in the terms. The proof of Theorem 8 gives mat-nucleus as the space generated by the linear forms in the dotted box. The linear forms that are not in mat-nucleus lie "above" the mat-nucleus and are all (mat-nucleus)-matched, i.e. $\forall \ell \in (L(T_1) \setminus \text{mat-nucleus})$, there is a form similar to ℓ modulo mat-nucleus in each $(L(T_i) \setminus \text{mat-nucleus})$. Thus the essence of Theorem 8 is: the mat-nucleus part of the terms of C has low rank k^2, while the part of the terms above mat-nucleus all look "similar".

Proof Idea for Theorem 8. The key insight in the construction of mat-nucleus is a reinterpretation of the identity test of Kayal & Saxena [KS07] as a structural result for $\Sigma\Pi\Sigma$ identities. Again, refer to the following figure depicting a $\Sigma\Pi\Sigma(4,d)$ circuit C and think of each bubble having d linear forms. Roughly, [KS07] showed that $C=0$ iff for every *path* (v_1,v_2,v_3) (where $v_i \in L(T_i)$): $T_4 \equiv 0 \pmod{v_1,v_2,v_3}$ or in ideal terms, $T_4 \in \langle v_1, v_2, v_3 \rangle$. Thus, roughly, it is enough to go through all the d^3 paths to certify the zeroness of C. This is why the time complexity of the identity test of [KS07] is dominated by d^k.

Now if we are given a $\Sigma\Pi\Sigma(4,d)$ identity C which is *minimal*, then we know that $T_1 + T_2 + T_3 \neq 0$. Thus, by applying the above interpretation of [KS07] to $T_1 + T_2 + T_3$ we will get a path (v_1, v_2) such that $T_3 \notin \langle v_1, v_2 \rangle$. Since $C=0$ this means that $T_3 + T_4 \equiv 0 \pmod{v_1, v_2}$ but $T_3, T_4 \not\equiv 0 \pmod{v_1, v_2}$ (if T_4 is in $\langle v_1, v_2 \rangle$ then so will be T_3). Thus, $T_3 \equiv -T_4 \pmod{v_1, v_2}$ is a nontrivial congruence and it immediately gives us a $\langle v_1, v_2 \rangle$-matching between T_3, T_4 (see Lemma 44). By repeating this argument with a different permutation of the terms we could match different terms (by a different ideal), and finally we expect to match all the terms (by the union of the various ideals).

This fantastic argument has numerous technical problems, but they can all be taken care of by suitable algebraic generalizations. The main stumbling block is the presence of *repeating* forms. It could happen that $(\bmod\ v_1)$, v_2 occurs in many terms, or in the same term with a higher power. The most important tool developed is an ideal version of Chinese

remaindering that forces us to consider not just linear forms v_1, v_2, but *multiplication terms* v_1, v_2 dividing T_1, T_2 respectively. We give the full proof in Section 3. (Interestingly, the *non*-blackbox identity test of [KS07] guides in devising a blackbox test of "similar" complexity over rationals.)

2.2. **Step 2: Certificate for Linear Independence of Gates.** Theorem 8 gives us a space K, of rank $< k^2$, that matches T_1 to each term T_i. In particular, this means that the list $L_K(T_i) := L(T_i) \cap K$ has the same cardinality d' for each $i \in [k]$. In fact, if we look at the corresponding multiplication terms $K_i := M(L_K(T_i))$, $i \in [k]$, then they again form a $\Sigma\Pi\Sigma(k, d')$ identity! Precisely, $C' = \sum_{i \in [k]} \alpha_i K_i$ for some α_i's in $\mathbb{F}^*$ (see Lemma 46) is an identity. We would like C' to somehow mimic the structure of C. Of course C' is simple but is it again minimal? Unfortunately, it may not be. For reasons that will be clear later, minimality of C' would have allowed us to go directly to Step 3. Now step 2 will involve increasing the space K (but not by too much) that gives us a C' that "behaves" like C. Specifically, if $T_1, \ldots, T_{k'}$ are *linearly independent* (i.e. $\nexists \overline{\beta} \in \mathbb{F}^{k'} \setminus \{\overline{0}\}$ s.t. $\sum_{i \in [k']} \beta_i T_i = 0$), then so are $K_1, \ldots, K_{k'}$.

Theorem 10 (Nucleus). *Let $C = \sum_{i \in [k]} T_i$ be a minimal $\Sigma\Pi\Sigma(k, d)$ identity and let $\{T_i | i \in \mathcal{I}\}$ be a maximal set of linearly independent terms $(1 \leqslant k' := |\mathcal{I}| < k)$. Then there exists a linear subspace K of $L(R)$ such that:*

1) *$rk(K) < 2k^2$.*
2) *$\forall i \in [k]$, there is a K-matching π_i between T_1, T_i.*
3) *(Define $\forall i \in \mathcal{I}$, $K_i := M(L_K(T_i))$.) The terms $\{K_i | i \in \mathcal{I}\}$ are linearly independent.*

Definition 11 (nucleus). *Let C be a minimal $\Sigma\Pi\Sigma(k, d)$ identity. The linear subspace K given by Theorem 10 is called the nucleus of C. By Lemma 46, the subspace K induces an identity $C' = \sum_{i \in [k]} \alpha_i K_i$ which we call the nucleus identity.*

The notion of the nucleus is easier to grasp when C is a $\Sigma\Pi\Sigma(k, d)$ identity that is *strongly minimal*, i.e. $T_1, \ldots, T_{k-1}$ are linearly independent. Clearly, such a C is also minimal[3]. For such a C, Theorem 10 gives a nucleus K such that the corresponding nucleus identity is strongly minimal. The structure of C is very strongly represented by C'. As a bonus, we actually end up greatly simplifying the polynomial-time PIT algorithm of Kayal & Saxena [KS07] (although we will not discuss this point in detail in this paper).

Proof Idea for Theorem 10. The first two properties in the theorem statement are already satisfied by mat-nucleus of C. So we incrementally add linear forms to the space mat-nucleus till it satisfies property (3) and becomes the nucleus. The addition of linear forms is guided by the ideal version of Chinese remaindering. For convenience assume T_1, T_2, T_3 to be linearly independent. Then, by homogeneity and equal degree, we have an equivalent ideal statement: $T_2 \notin \langle T_1 \rangle$ and $T_3 \notin \langle T_1, T_2 \rangle$ (see Lemma 42). Even in this general setting the path analogy (used in the last subsection) works and we essentially get linear forms $v_1 \in L(T_1)$ and $v_2 \in L(T_2)$ such that: $T_2 \notin \langle v_1 \rangle$ and $T_3 \notin \langle v_1, v_2 \rangle$. We now add these forms v_1, v_2 to the space mat-nucleus, and call the new space K. It is expected that the new K_1, K_2, K_3 are now linearly independent.

Not surprisingly, the above argument has numerous technical problems. But it can be made to work by careful applications of the ideal version of Chinese remaindering. We give the full proof in Section 4.

---

[3]If for some proper $S \subset [k]$, $\sum_{i \in S} T_i = \sum_{i \in \overline{S}} T_i = 0$ then linear independence of $T_1, \ldots, T_{k-1}$ is violated.

2.3. **Step 3: Invoking Sylvester-Gallai Theorems.** We make a slight, but hopefully interesting, detour and leave depth-3 circuits behind. We rephrase the standard Sylvester-Gallai theorems in terms of *Sylvester-Gallai closure* (or configuration) and *rank bounds*. This is far more appropriate for our application, and seems to be very natural in itself.

Definition 12 (SG$_k$-closed). *Let $k \in \mathbb{N}^{>1}$. Let S be a subset of non-zero vectors in $\mathbb{F}^n$ without* multiples*: no two vectors in S are scalar multiples of each other[4]. Suppose that for every set V of k linearly independent vectors in S, the linear span of V contains at least $(k+1)$ vectors of S. Then, the set S is said to be* SG$_k$-closed.

We would expect that if S is finite then it will get harder to keep S SG$_k$-closed as rk(S) is gradually increased. This intuition holds up when $\mathbb{F} = \mathbb{R}$. As we mentioned earlier, the famous Sylvester-Gallai Theorem states: if a finite $S \subset \mathbb{R}^n$ is SG$_2$-closed, then rk$(S) \leqslant 2$. It is optimal as the line $S := \{(1,0), (1,1), (1,2)\}$ has rank 2 and is SG$_2$-closed.

In fact, there is also a generalization of the Sylvester-Gallai theorem known (as stated in Theorem 2.1 of [BE67]) : *Let S be a finite set in $\mathbb{RP}^{2t}$ spanning that projective space. Then, there exists a t-flat H such that $|H \cap S| = t + 1$, and H is spanned by those points $H \cap S$.*

Let $S \subset \mathbb{R}^n$ be a finite set of points with first coordinate being 1 and let $k \geq 2$. We claim that if S is SG$_k$-closed, then rk$(S) \leqslant 2(k-1)$. Otherwise the above theorem guarantees k vectors V in S whose $(k-1)$-flat H has only k points of S. If sp(V) has a point $s \in S \setminus V$ then as S has first coordinates 1, it would mean that a *convex* linear combination of V (i.e. sum of coefficients in the combination is 1) is s. In other words, $s \in H$, which contradicts H having only k points of S. Thus, sp(V) also has no point in $S \setminus V$, but this contradicts SG$_k$-closure of S. This shows that higher dimensional Sylvester-Gallai theorem implies that if S is SG$_k$-closed then rk$(S) \leqslant 2(k-1)$. We prefer using this rephrasal of the higher dimensional Sylvester-Gallai Theorem. This motivates the following definitions.

Definition 13 (SG operator). *Let $k, m \in \mathbb{N}^{>1}$.*

[**SG**$_k(\cdot, \cdot)$] *The largest possible rank of an SG$_k$-closed set of at most m points in $\mathbb{F}^n$ is denoted by $SG_k(\mathbb{F}, m)$. For example, the above discussion entails $SG_k(\mathbb{R}, m) \leqslant 2(k-1)$ which is, interestingly, independent of m. (Also verify that $SG_k(\mathbb{F}, m) \leq SG_{k'}(\mathbb{F}', m')$ for $k \leqslant k'$, $m \leqslant m'$ and $\mathbb{F} \subseteq \mathbb{F}'$.)*

[**SG**$_k(\cdot)$] *Suppose a set $S \subseteq \mathbb{F}^n$ has rank greater than $SG_k(\mathbb{F}, m)$ (where $\#S \leqslant m$). Then, by definition, S is not SG$_k$-closed. In this situation we say the k-dimensional Sylvester-Gallai operator $SG_k(S)$ (applied on S) returns a set of k linearly independent vectors V in S whose span has no point in $S \setminus V$.*

The Sylvester-Gallai theorem in higher dimensions can now be expressed succinctly.

Theorem 14 (High dimension Sylvester-Gallai for $\mathbb{R}$)**.** [Han65, BE67] $SG_k(\mathbb{R}, m) \leqslant 2(k-1)$.

Remark. This theorem is also optimal, for if we set S to be a union of $(k-1)$ "skew lines" then S has rank $2(k-1)$ and is SG$_k$-closed. For example, when $k = 3$ define $S := \{(1,1,0,0), (1,1,1,0), (1,1,2,0)\} \cup \{(1,0,1,0), (1,0,1,1), (1,0,1,2)\}$. It is easy to verify that rk$(S) = 4$ and the span of every three linearly independent vectors in S contains a fourth vector!

---

[4]This is just a set of elements in the projective space $\mathbb{FP}^{n-1}$, but this formulation in terms of vectors is more convenient for our applications.

Using some linear algebra and combinatorial tricks, we prove the first ever Sylvester-Gallai bound for all fields. The proof is in Section 6, where there is a more detailed discussion of this (and the connection with LDCs).

Theorem 3 (SG_k for all fields). *For any field $\mathbb{F}$ and $k, m \in \mathbb{N}^{>1}$, $SG_k(\mathbb{F}, m) \leq 9k \lg m$.*

2.3.1. *Back to identities.* Let C be a simple and strongly minimal $\Sigma\Pi\Sigma(k, d)$ identity. Theorem 10 gives us a nucleus K, of rank $< 2k^2$, that matches T_1 to each term T_i. As seen in Step 2, if we look at the corresponding multiplication terms $K_i := M(L_K(T_i))$, $i \in [k]$, then they again form a $\Sigma\Pi\Sigma(k, d')$ "nucleus identity" $C' = \sum_{i \in [k]} \alpha_i K_i$, for some α_i's in $\mathbb{F}^*$, which is simple and strongly minimal. Define the *non-nucleus part* of T_i as $L_K^c(T_i) := L(T_i) \setminus K$, for all $i \in [k]$ (c in the exponent annotates "complement", since $L(T_i) = L_K(T_i) \sqcup L_K^c(T_i)$). What can we say about the rank of $L_K^c(T_i)$?

Define the *non-nucleus part of C* as $L_K^c(C) := \bigcup_{i \in [k]} L_K^c(T_i)$. Our goal in Step 3 is to bound $\text{rk}(L_K^c(C) \bmod K)$ by $2k$ when the field is $\mathbb{R}$. This will give us a rank bound of $\text{rk}(K) + \text{rk}(L_K^c(C) \bmod K) < (2k^2 + 2k)$ for simple and strongly minimal $\Sigma\Pi\Sigma(k, d)$ identities over $\mathbb{R}$. The proof is mainly combinatorial, based on higher dimensional Sylvester-Gallai theorems and a property of set partitions, with a sprinkling of algebra.

We will finally apply SG_k operator not directly on the forms in $L(C)$ but on a suitable truncation of those forms. So we need another definition.

Definition 15 (Non-K rank). *Let K be a linear subspace of $L(R)$. Then $L(R)/K$ is again a linear space (the* quotient space*). Let S be a list of forms in $L(R)$. The* non-K rank of S is defined to be $rk(S \bmod K)$ (i.e. the rank of S when viewed as a subset of $L(R)/K$).*

Let C be a $\Sigma\Pi\Sigma(k, d)$ identity with nucleus K. The non-K rank of the non-nucleus part $L_K^c(T_i)$ is called the non-nucleus rank of T_i*. Similarly, the non-K rank of the non-nucleus part $L_K^c(C) := \bigcup_{i \in [k]} L_K^c(T_i)$ is called the* non-nucleus rank of C*.*

We give an example to explain the non-K rank. Let $R = \mathbb{F}[z_1, \cdots, z_n, y_1, \cdots, y_m]$. Suppose $K = \text{sp}(z_1, \cdots, z_n)$ and $S \subset L(R)$. We can take any element ℓ in S and simply drop all the z_i terms, i.e. 'truncate' z-part of ℓ. This gives a set of linear forms over the y variables. The rank of these is the non-K rank of S.

We are now ready to state the theorem that is proved in Step 3. It basically shows a neat relationship between the non-nucleus part and Sylvester-Gallai.

Theorem 16 (Bound for simple, strongly minimal identities). *Let $|\mathbb{F}| > d$. The non-nucleus rank of a simple and strongly minimal $\Sigma\Pi\Sigma(k, d)$ identity over $\mathbb{F}$ is at most $SG_{k-1}(\mathbb{F}, d)$.*

Given a simple, minimal $\Sigma\Pi\Sigma(k, d)$ identity C that is not strongly minimal. Let $T_1, \ldots, T_{k'}$ be linearly independent and form a basis of $\{T_i | i \in [k]\}$. Then it is clear that $\exists \bar{a} \in \mathbb{F}^{k'} \setminus \{\bar{0}\}$ such that $\sum_{i \in [k']} a_i T_i + T_{k'+1}$ is a strongly minimal $\Sigma\Pi\Sigma(k'', d)$ identity (for some $1 < k'' \leq k' + 1$). Hence, we could apply the above theorem on this identity and get a rank bound for the non-nucleus part. The only problem is this fanin-k'' identity may not be simple. Our solution for this is to replace $T_{k'+1}$ by the suitable linear combination of $\{T_i | i \in [k']\}$ in C and repeat the above argument on the new identity. In Section 5.2 we show this takes care of the whole non-nucleus part and bounds its rank by $k \cdot SG_k(\mathbb{F}, d)$. To state the theorem formally, we need a more refined notion than the fanin of a $\Sigma\Pi\Sigma$ circuit.

Definition 17 (Independent-fanin). *Let $C = \sum_{i \in [k]} T_i$ be a $\Sigma\Pi\Sigma(k, d)$ circuit. The independent-fanin of C, ind-fanin(C), is defined to be the size of the maximal $\mathcal{I} \subseteq [k]$*

such that $\{T_i | i \in \mathcal{I}\}$ are linearly independent polynomials. (Remark: If ind-fanin$(C) = k$ then $C \neq 0$. Also, for an identity C, C is strongly minimal iff ind-fanin$(C) = k - 1$.)

We now state the following stronger version of the main theorem.

Theorem 18 (Final bound). *Let $|\mathbb{F}| > d$. The rank of a simple, minimal $\Sigma\Pi\Sigma(k, d)$, independent-fanin k', identity is at most $2k^2 + (k - k') \cdot SG_{k'}(\mathbb{F}, d)$.*

Remark: In particular, the rank of a simple, minimal $\Sigma\Pi\Sigma(k, d)$ identity over reals is at most $2k^2 + (k - k') \cdot SG_{k'}(\mathbb{R}, d) \leq 2k^2 + (k - k')2(k' - 1) < 3k^2$, proving the main theorem over reals. Likewise, for any $\mathbb{F}$, we get the rank bound of $2k^2 + (k - k') \cdot SG_{k'}(\mathbb{F}, d) \leq 2k^2 + (k - k')9k' \lg d \leqslant 2k^2 + \frac{9k^2}{4} \lg d < 3k^2 \lg 2d$, proving the main theorem.

Proof Idea for Theorem 16. Basically, we apply the $SG_k(\cdot)$ operator on the non-nucleus part of the term T_1, i.e. we treat a linear form $\sum_i a_i x_i$ as the point $(1, \frac{a_2}{a_1}, \ldots, \frac{a_n}{a_1}) \in \mathbb{F}^n$ for the purposes of Sylvester-Gallai and then we consider $SG_k(L_K^c(T_1))$ assuming that the non-nucleus rank of T_1 is more than $SG_k(\mathbb{F}, d)$. This application of Sylvester-Gallai is much more direct compared to the methods used in [KS09b]. There, they needed versions of Sylvester-Gallai that dealt with colored points and had to prove a *hyperplane decomposition* property after applying essentially a $SG_{k^{O(k)}}(\cdot)$ operator on $L(C)$. Since, modulo the nucleus, all multiplication terms look essentially the same, it suffices to focus attention on just one of them. Hence, we apply the SG_k-operator on a single multiplication term.

To continue with the proof idea, assume C is a simple, strongly minimal $\Sigma\Pi\Sigma(k, d)$ identity with terms $\{T_i | i \in [k]\}$ and let K be its nucleus given by Step 2. It will be convenient for us to fix a linear form $y_0 \in L(R)^*$ and a subspace U of $L(R)$ such that we have the following *orthogonal* vector space decomposition $L(R) = \mathbb{F}y_0 \oplus U \oplus K$ (i.e. $\ell \in \mathbb{F}y_0 \cap U$ implies $\ell = 0$ and $\ell \in (\mathbb{F}y_0 \oplus U) \cap K$ implies $\ell = 0$). This means for any form $\ell \in L(R)$, there is a unique way to express $\ell = \alpha y_0 + u + v$, where $\alpha \in \mathbb{F}$, $u \in U$ and $v \in K$. Furthermore, we will assume wlog that for every form $\ell \in L_K^c(T_1)$ the corresponding α is nonzero, i.e. each form in $L_K^c(T_1)$ is *monic* wrt y_0 (see Lemma 40).

Definition 19 (trun$(\cdot)$). *Fix a decomposition $L(R) = \mathbb{F}y_0 \oplus U \oplus K$. For any form $\ell \in L_K^c(T_1)$, there is a unique way to express $\ell = \alpha y_0 + u + v$, where $\alpha \in \mathbb{F}^*$, $u \in U$ and $v \in K$.*
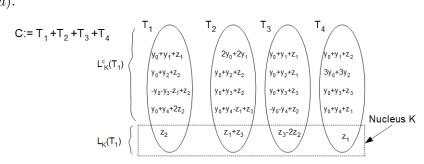
The truncated form trun(ℓ) is the linear form obtained by dropping the K part and normalizing, i.e. trun$(\ell) := y_0 + \alpha^{-1}u$.

Given a list of forms S we define trun(S) to be the corresponding set (thus no repetitions) of truncated forms.

To be precise, we fix a basis $\{y_1, \ldots, y_{rk(U)}\}$ of U so that each form in trun$(L_K^c(T_1))$ has representation $y_0 + \sum_{i \geq 1} a_i y_i$ (a_i's $\in \mathbb{F}$). We view each such form as the *point* $(1, a_1, \ldots, a_{rk(U)})$ while applying Sylvester-Gallai on trun$(L_K^c(T_1))$. Assume, for the sake of contradiction, that the non-nucleus rank of T_1, rk(trun$(L_K^c(T_1))) > SG_k(\mathbb{F}, d)$ then (by definition) $SG_k(\text{trun}(L_K^c(T_1)))$ gives k linearly independent forms $\ell_1, \ldots, \ell_k \in (y_0 + U)$ whose span contains no *other* linear form of trun$(L_K^c(T_1))$.

For simplicity of exposition, let us fix $k = 4$, K spanned by z's, U spanned by y's and $\ell_i = y_0 + y_i$ ($i \in [4]$). Note that (by definition) trun$(\alpha y_0 + \sum_i \alpha_i z_i + \sum_i \beta_i y_i) = y_0 + \sum_i \frac{\beta_i}{\alpha} y_i$. We want to derive a contradiction by using the SG_4-tuple $(y_0 + y_1, y_0 + y_2, y_0 + y_3, y_0 + y_4)$ *and* the fact that C is a simple, strongly minimal $\Sigma\Pi\Sigma(4, d)$ identity. The contradiction is easy to see in the following configuration: Suppose the linear forms in C that are similar

to a form in $\bigcup_{i\in[4]}(y_0 + y_i + K)$ are *exactly* those depicted in the figure. Let us consider C modulo the ideal $I := \langle y_0 + y_1 + z_1, y_0 + y_2 + z_2, -y_0 - y_4 + z_2 \rangle$. It is easy to see that these forms (call them $\ell_1', \ell_2', \ell_4'$) "kill" the first three gates, leaving $C \equiv T_4 \pmod{I}$. As C is an identity this means $T_4 \in I$, thus there is a form $\ell \in L(T_4)$ such that $\ell \in \mathrm{sp}(\ell_1', \ell_2', \ell_4')$. Now none of the forms $\ell_1', \ell_2', \ell_4'$ divide T_4. Also, their non-trivial combination, say $\alpha\ell_1' + \beta\ell_2'$ for $\alpha\beta \neq 0$, cannot occur in $L(T_4)$. Otherwise, by the matching property $\mathrm{trun}(\alpha\ell_1' + \beta\ell_2') = (\alpha+\beta)^{-1}(\alpha\ell_1 + \beta\ell_2)$ will be in $\mathrm{trun}(L_K^c(T_1))$. This contradicts the ℓ_i's being a SG_4-tuple. Thus, T_4 cannot be in I, a contradiction. This means that the non-nucleus rank of T_1 is $\leq \mathrm{SG}_4(\mathbb{F}, d)$, which by matching properties implies the non-nucleus rank of C is $\leq \mathrm{SG}_4(\mathbb{F}, d)$.

We were able to force a contradiction because we used a set of forms in an SG-tuple that killed three terms and "preserved" the last term. Can we always do this? This is not at all obvious, and that is because of repeating forms. Suppose, after going modulo form ℓ, the circuit looks like $x^3y + 2x^2y^2 + xy^3 = 0$. This is not simple, but *it does not have to be*. We are only guaranteed that the original circuit is simple. Once we go modulo ℓ, that property is lost. Now, the choice of *any* form kills all terms. In the figure above, $\langle y_0 + y_1 + z_1, y_0 + y_2 + z_2, y_0 + y_3 + z_3 \rangle$ does not yield a contradiction. We will use our more powerful Chinese remaindering tools and the nucleus properties to deal with this. We have to prove a special theorem about partitions of $[k]$ and use strong minimality (which we did not use in the above sketch). The full proof is given in Section 5.1.

3. Matching the Terms in an Identity: Construction of mat-nucleus

3.1. **Chinese Remaindering for Multiplication Terms.** Traditionally, Chinese remaindering is the fact: if two coprime polynomials (resp. integers) f, g divide a polynomial (resp. integer) h then fg divides h. The key tool in constructing mat-nucleus is a version of Chinese remaindering specialized for multiplication terms but generalized to ideals. Similar methods appeared first in [KS07] but we turn those on their head and give a "simpler" proof. In particular, we avoid the use of local rings and Hensel lifting.

Definition 20 (Radical-span). *Let $S := \{f_1, \dots, f_m\}$ be multiplication terms generating an ideal I. Define linear space $\mathrm{radsp}(S) := \mathrm{sp}(L(f_1) \cup \dots \cup L(f_m))$.*

When the set of generators S are clear from the context we will also use the notation $\mathrm{radsp}(I)$. Similarly, $\mathrm{radsp}(I, f)$ would be a shorthand for $\mathrm{radsp}(S \cup \{f\})$.

Remark. Radical-span is motivated by the *radical* of an ideal but it is not quite that, for example, $\mathrm{radical}(x_1^2, x_1x_2) = \langle x_1 \rangle$ but $\mathrm{radsp}(x_1^2, x_1x_2) = \mathrm{sp}(x_1, x_2)$. It is easy to see that the ideal generated by radsp always contains the radical ideal.

Now we can neatly state Chinese remaindering as an *ideal decomposition* statement.

Theorem 21 (Ideal Chinese remaindering). *Let $f_1, \ldots, f_m, z, f, g$ be multiplication terms. Define the ideal $I := \langle f_1, \ldots, f_m \rangle$. Assume $L(z) \subseteq radsp(I)$ while, $L(f) \cap radsp(I) = \emptyset$ and $L(g) \cap radsp(I, f) = \emptyset$. Then, $\langle I, zfg \rangle = \langle I, z \rangle \cap \langle I, f \rangle \cap \langle I, g \rangle$.*

Proof. If h is a polynomial in $\langle I, zfg \rangle$ then clearly it is in each of the ideals $\langle I, z \rangle$, $\langle I, f \rangle$ and $\langle I, g \rangle$.

Suppose h is a polynomial in $\langle I, z \rangle \cap \langle I, f \rangle \cap \langle I, g \rangle$. Then by definition there exist $i_1, i_2, i_3 \in I$ and $a, b, c \in R$ such that,

$$h = i_1 + az = i_2 + bf = i_3 + cg.$$

The second equation gives $bf \in \langle I, z \rangle$. Since $L(f) \cap \mathrm{radsp}(I, z) = L(f) \cap \mathrm{radsp}(I) = \emptyset$, repeated applications of Lemma 41 give us, $b \in \langle I, z \rangle$. Implying $bf \in \langle I, z \rangle f \subseteq \langle I, zf \rangle$, hence $h = i_2 + bf \in \langle I, zf \rangle$. This ensures the existence of $i_2' \in I$ and a polynomial b' such that,

$$h = i_2' + b'zf = i_3 + cg.$$

Again this system says that $cg \in \langle I, zf \rangle$. Since $L(g) \cap \mathrm{radsp}(I, zf) = L(g) \cap \mathrm{radsp}(I, f) = \emptyset$, repeated applications of Lemma 41 give us $c \in \langle I, zf \rangle$. Implying $cg \in \langle I, zf \rangle g \subseteq \langle I, zfg \rangle$, hence $h = i_3 + cg \in \langle I, zfg \rangle$. This finishes the proof. $\square$

The conditions in this theorem suggest that factoring a multiplication term f into parts corresponding to the equivalence classes of "similarity mod radsp(I)" would be useful.

Definition 22 (Nodes). *Let f be a multiplication term and let I be an ideal generated by some multiplication terms. As the relation "similarity mod radsp(I)" is an equivalence relation on $L(R)$, it partitions, in particular, the list $L(f)$ into equivalence classes.*

$[\mathbf{rep}_I(f)]$ For each such class pick a representative ℓ_i and define $rep_I(f) := \{\ell_1, \ldots, \ell_r\}$. (Note that form 0 can also appear in this set, it represents the class $L(f) \cap radsp(I)$.)

$[\mathbf{nod}_I(f)]$ For each $\ell_i \in rep_I(f)$, we multiply the forms in f that are similar to ℓ_i mod radsp(I). We define nodes of f mod I as the set of polynomials $nod_I(f) := \{M(L(f) \cap (\mathbb{F}^* \ell + radsp(I))) \mid \ell \in rep_I(f)\}$. (Remark: When $I = \{0\}$, nodes of f are just the coprime powers-of-forms dividing f.)

$[\ldots \mathbf{wrt\ a\ subspace}]$ Let K be a linear subspace of $L(R)$. Clearly, the relation "similarity mod K" is an equivalence relation on $L(R)$. It will be convenient for us to also use notations $rep_K(f)$ and $nod_K(f)$. They are defined by replacing radsp(I) in the above definitions by K.

Observe that the product of polynomials in $\mathrm{nod}_I(f)$ just gives f. Also, modulo radsp(I), each node is just a form-power ℓ^r. In other words, modulo radsp(I), a node is rank-one term. The choice of the word "node" might seem a bit mysterious, but we will eventually construct paths through these. To pictorially see what is going on, think of each term T_i as a set of its constituent nodes.

We prove a corollary of the ideal Chinese remaindering theorem that will be very helpful in both Steps 1 and 2.

Corollary 23. *Let $h \in R$, f be a multiplication term, and let I be an ideal generated by some multiplication terms. Then, $h \notin \langle I, f \rangle$ iff $\exists g \in nod_I(f)$ such that $h \notin \langle I, g \rangle$.*

Proof. If $h \notin \langle I, g \rangle$, for some $g \in \mathrm{nod}_I(f)$, then clearly $h \notin \langle I, f \rangle$.

Conversely, assume $h \notin \langle I, f \rangle$. Let $\mathrm{rep}_I(f) = \{\ell_1, \ldots, \ell_r\}$ and correspondingly, $\mathrm{nod}_I(f) = \{g_1, \ldots, g_r\}$. If $r = 1$ then f is similar to g_1, hence $h \notin \langle I, g_1 \rangle$ and we are done. So assume $r \geq 2$. Also, in case $L(f)$ has a form in radsp(I), assume wlog ℓ_1 is the representative of the class $L(f) \cap \mathrm{radsp}(I)$. Define $G_i := \prod_{i < j \leqslant r} g_i$, for all $i \in [r-1]$.

We claim that for all $i \in [r-1]$, $L(G_i) \cap \mathrm{radsp}(I, g_i) = \emptyset$. Otherwise $\exists \ell \in L(G_i)$ such that either $\ell \in \mathrm{radsp}(I)$ or $\ell \in (\mathbb{F}^* \ell_i + \mathrm{radsp}(I))$. Former case contradicts ℓ_1 being the representative of the class $L(f) \cap \mathrm{radsp}(I)$, while the latter case contradicts $\ell_{i+1}, \ldots, \ell_r$ being *non*-similar to ℓ_i mod $\mathrm{radsp}(I)$. Thus, for all $i \in [r-1]$, $L(G_i) \cap \mathrm{radsp}(I, g_i) = \emptyset$, and by applying Theorem 21 on $\langle I, g_i G_i \rangle$ for each $i \in [r-1]$, we deduce:

$$\langle I, f \rangle = \left\langle I, \prod_{i \in [r]} g_i \right\rangle = \bigcap_{i \in [r]} \langle I, g_i \rangle.$$

Thus, $h \notin \langle I, f \rangle$ implies the existence of some $i \in [r]$ such that $h \notin \langle I, g_i \rangle$. $\qquad\square$

3.2. **Applying Chinese Remaindering to $\Sigma\Pi\Sigma$ Circuits.** We showed the effect of ideal Chinese remaindering on a single multiplication term f in Corollary 23. Now we show the effect on a *tuple* of multiplication terms, for example, appearing in a $\Sigma\Pi\Sigma$ circuit. We then need, quite naturally, a notion of *path* of nodes.

Definition 24 (Paths)**.** *Let I be an ideal generated by some multiplication terms. Let $C = \sum_{i \in [k]} T_i$ be a $\Sigma\Pi\Sigma(k, d)$ circuit. Let v_i be a sub-term of T_i (i.e. $L(v_i) \subseteq L(T_i)$), for all $i \in [k]$. We call the tuple $(I, v_1, \ldots, v_k)$ a path of C mod I if, for all $i \in [k]$, $v_i \in nod_{\langle I, v_1, \ldots, v_{i-1}\rangle}(T_i)$. It is of length k. (Remark: We have defined path $\overline{p}$ as a tuple but, for convenience, we will sometimes treat it as a set of multiplication terms, eg. when operated upon by $sp(\cdot)$, $\langle \cdot \rangle$, $radsp(\cdot)$, etc.)*

Conventionally, when $k = 0$ the circuit C has just "one" gate: 0. In that case, the only path C has is (I), which is of length 0.

We also define, for any subset $S \subseteq [k]$, the sub-circuit $C_S := \sum_{s \in S} T_s$.

For an $i \in \{0, \ldots, k-1\}$, define $[i]' := [k] \setminus [i]$. We set $[0] := \emptyset$ and $C_\emptyset := 0$.

We now show that if C is a nonzero $\Sigma\Pi\Sigma(k, d)$ circuit then $\exists i \in \{0, \ldots, k-1\}$, such that $C_{[i]}$ has a path $\overline{p}$ for which, $C \pmod{\langle \overline{p} \rangle}$ is nonzero *and* similar to some multiplication term. This rather special path inside C can be seen as a certificate for the nonzeroness. The rank of the linear forms appearing in this path can be at most $i + \mathrm{rk}(\mathrm{radsp}(I))$, since the rank of each node is one, modulo the radical-span of the previous nodes in the path. Hence, it is a *low-rank certificate for the nonzeroness of C*.

Theorem 25 (Certificate for a Non-identity)**.** *Let I be an ideal generated by some multiplication terms. Let $C = \sum_{i \in [k]} T_i$ be a $\Sigma\Pi\Sigma(k, d)$ circuit that is nonzero modulo I. Then $\exists i \in \{0, \ldots, k-1\}$ such that $C_{[i]}$ mod I has a path $\overline{p}$ satisfying: $C_{[i]'} \equiv \alpha \cdot T_{i+1} \not\equiv 0 \pmod{\overline{p}}$ for some $\alpha \in \mathbb{F}^*$.*

Before we prove the theorem, we make an aside observation. If the reader has kept the mental picture of the terms as consisting of rank-one (modulo $\mathrm{radsp}(I)$) nodes, then the notion of a path has some meaning. A path $\overline{p}$ kills the terms that is passed through, and collapses remaining circuit into a single term. This is very reminiscent of the poly-time algorithm of Kayal & Saxena [KS07]. Indeed, this theorem is a (shorter) *proof* of the correctness of the algorithm. Why? Consider the path $\overline{p}$ given by the theorem when I is the zero ideal. The path $\overline{p}$ can be represented by a list of at most k 'forms' in $L(C)$. This path comes from some $C_{[i]}$, which means that $C_{[i]} = 0 \pmod{\overline{p}}$. So, we get that $C \equiv \alpha \cdot T_{i+1} \not\equiv 0 \pmod{\overline{p}}$. Since T_{i+1} is a product of linear forms, it is easy to algorithmically check if $C \equiv 0 \pmod{\overline{p}}$. If C is identically zero, such a path cannot exist. Since there are at most d^k different paths, we can exhaustively check all of them. That yields an alternative view of [KS07]'s test.

Proof. Fix an $i \in \{0, \ldots, k-1\}$ and a path $\overline{p}$ of $C_{[i]}$ mod I such that:

1) $C_{[i]'} \notin \langle \overline{p} \rangle$ and,
2) the set $J_i := \{j \in [i]' \mid T_j \notin \langle \overline{p} \rangle\} \neq \emptyset$ is the smallest possible (over all i).

Note that for values $i = 0$, $\overline{p} = (I)$, the condition (1) is satisfied and the corresponding $J_i \neq \emptyset$. Thus, there also exist i and $\overline{p}$ satisfying both the conditions (1) and (2).

Let j^* be the smallest element in J_i. This means that for every m, $i < m < j^*$, $T_m \in \langle \overline{p} \rangle$. This means, by repeated applications of Lemma 41, $v_m := M(L_{\mathrm{radsp}(\overline{p})}(T_m)) \in \langle \overline{p} \rangle$. Thus, $\langle \overline{p} \rangle = \langle \overline{p} \cup \{v_m | i < m < j^*\} \rangle$. This makes $\overline{q} := (\overline{p}, (v_m | i < m < j^*))$ also a path of $C_{[j^*-1]}$ mod I. We now claim that $\overline{q}$ is the path promised in the theorem statement.

Note that $C_{[j^*-1]'} \equiv C_{[i]'}$ (mod $\overline{p}$) and $C_{[i]'} \notin \langle \overline{p} \rangle = \langle \overline{q} \rangle$, in other words, path $\overline{q}$ also satisfies:

1) $C_{[j^*-1]'} \notin \langle \overline{q} \rangle$ and,
2) the set $J_{j^*-1} = \{j \in [j^*-1]' \mid T_j \notin \langle \overline{q} \rangle\} = J_i$ is still the smallest possible.

If $C_{[j^*-1]'} \notin \langle \overline{q}, T_{j^*} \rangle$ then, by Corollary 23, there exists $v_{j^*} \in \mathrm{nod}_{\langle \overline{q} \rangle}(T_{j^*})$ such that $C_{[j^*-1]'} \notin \langle \overline{q}, v_{j^*} \rangle$, hence $C_{[j^*]'} = C_{[j^*-1]'} - T_{j^*} \notin \langle \overline{q}, v_{j^*} \rangle$. Define $\overline{q}' := (\overline{q}, v_{j^*})$, clearly it is a path of $C_{[j^*]}$ mod I. Wrt this path $\overline{q}'$, $J_{j^*} \subseteq J_i \setminus \{j^*\} \subsetneq J_i$ together with $C_{[j^*]'} \notin \langle \overline{q}' \rangle$, contradicting the minimality assumption on i. Thus, we assume $C_{[j^*-1]'} \in \langle \overline{q}, T_{j^*} \rangle$. By Lemma 42, this guarantees the existence of an $\alpha \in \mathbb{F}$ such that,

$$(C_{[j^*-1]'} - \alpha T_{j^*}) \in \langle \overline{q} \rangle = \langle \overline{p} \rangle.$$

Since $C_{[j^*-1]'} \equiv C_{[i]'} \not\equiv 0$ (mod $\overline{p}$), the above equation can be rewritten as:

$$C_{[j^*-1]'} \equiv \alpha T_{j^*} \not\equiv 0 \ (\mathrm{mod} \ \langle \overline{q} \rangle).$$

Thus, finishing the proof (α nonzero is implied). $\qquad \square$

Remark. The above theorem is quite powerful, for instance, it only needs the *non-zeroness* of C mod I without referring to any simplicity or minimality requirements.

3.3. **Using Minimality to get mat-nucleus.** If we are given a circuit that is zero & minimal (may not be simple) then a repeated application of Theorem 25 gives us a space *mat-nucleus* that matches all the multiplication terms of C.

Theorem 8 (Matching-Nucleus). *Let $C = T_1 + \cdots + T_k$ be a $\Sigma\Pi\Sigma(k, d)$ circuit that is minimal and zero. Then there exists a linear subspace K of $L(R)$ such that:*

1) $\mathrm{rk}(K) < k^2$.
2) $\forall i \in [k]$, there is a K-matching π_i between T_1, T_i.

Proof. The proof is an iterative process with at most k rounds. We maintain a set $\mathcal{P}$, containing paths of some sub-circuits of C, and an undirected graph $G = ([k], E)$. For convenience, define $U := \mathrm{radsp}(p | p \in \mathcal{P})$ (i.e. consider each path p as a set of multiplication terms, take the union of all these sets, and compute its radical-span). The *invariant* at the end of each round is: $(i, j) \in E$ iff T_i, T_j are U-matched. At the end of round 0 we assume, $\mathcal{P} := \{(0)\}$ and $E := \{(i, j) \in [k]^2 \mid T_i, T_j \text{ are similar}\}$. We want to eventually make G a connected graph (infact a k-*clique*) by keeping $\mathrm{rk}(U)$ as small as possible.

Suppose the invariant holds till the end of some round $(r-1) \geq 0$. If G is connected then the process stops at round $(r-1)$. Otherwise, we will show how to *decrease* the number of connected components of G in round r. Say, G has a maximal connected component on vertices $S \subsetneq [k]$. Since $C_S \neq 0$ (by minimality), we can apply Theorem 25 on C_S mod $\langle 0 \rangle$ to get a path p_S inside C_S mod $\langle 0 \rangle$ such that $\exists i \in S$, $C_S \equiv \alpha T_i \not\equiv 0$ (mod p_S) for some $\alpha \in \mathbb{F}^*$.

Define $S' := [k] \setminus S$. Now,

$$(1) \qquad\qquad C \equiv C_{S'} + \alpha T_i \equiv 0 \ (\mathrm{mod} \ p_S).$$

This means $C_{S'} \notin \langle p_S \rangle$ (otherwise $\alpha T_i \in \langle p_S \rangle$, a contradiction). Thus, we can apply Theorem 25 on $C_{S'} \bmod \langle p_S \rangle$ to get a path $p_{S'}$ inside $C_{S'} \bmod \langle p_S \rangle$ such that, $C_{S'} \equiv \beta T_j \not\equiv 0 (\mathrm{mod} \ p_{S'})$, for some $\beta \in \mathbb{F}^*$. This allows us to rewrite Equation (1) as:

$$\alpha T_i \equiv -\beta T_j \not\equiv 0 \ (\mathrm{mod} \ p_{S'})$$

Define $K' := \mathrm{radsp}(p_{S'})$. As $p_{S'}$ is, after all, a path of some sub-circuit of $C \bmod \langle 0 \rangle$, of length at most $|S| - 1 + |S'| - 1 = k - 2$, we deduce that $\mathrm{rk}(K') < (k-1)$. Also, by Lemma 44, the above congruence implies a K'-matching between T_i and T_j. We append the path $p_{S'}$ to $\mathcal{P}$ and update U. Note that for any edge (i, i') in the connected component S, and for any edge (j, j') in the connected component $\tilde{S}$ (of vertex j): since $T_i, T_{i'}$ are still U-matched; $T_j, T_{j'}$ are still U-matched; T_i, T_j are newly K'-matched; gives us that $T_{i'}, T_{j'}$ are newly U-matched. In other words, the two different connected components S and $\tilde{S}$ of G will now form a bigger connected component (infact a *clique*) when we update the graph as, $E := \{(a, b) \in [k]^2 \mid T_a, T_b \text{ are } U\text{-matched}\}$.

So in every round we are increasing $\mathrm{rk}(U)$ by at most $(k-1)$, maintaining the invariant, and decreasing the number of connected components in G by at least one. Thus, after at most $(k-1)$ repetitions we get a U that matches T_1, T_i, for all $i \in [k]$, and $\mathrm{rk}(U) < k^2$. We define this U as K, finishing the proof. $\qquad\square$

4. Certificate for Linear Independence of terms: Constructing nucleus

Suppose we have multiplication gates $T_1, \ldots, T_{k'}$ and a space K' of $L(R)$ such that T_1, T_i is K'-matched, for all $i \in [k']$. We show in this section that if $T_1, \ldots, T_{k'}$ are linearly independent (i.e. $\nexists \bar{\beta} \in \mathbb{F}^{k'} \setminus \{\bar{0}\}$ s.t. $\sum_{i \in [k']} \beta_i T_i = 0$) then K' can be extended to a linear space K of rank at most $(\mathrm{rk}(K') + k'^2)$ such that: $M(L_K(T_1)), \ldots, M(L_K(T_{k'}))$ are also linearly independent. This will prove Theorem 10.

Theorem 10 (Nucleus). Let $C = \sum_{i \in [k]} T_i$ be a minimal $\Sigma\Pi\Sigma(k, d)$ identity and let $\{T_i \mid i \in \mathcal{I}\}$ be a maximal set of linearly independent terms $(1 \leqslant k' := |\mathcal{I}| < k)$. Then there exists a linear subspace K of $L(R)$ such that:

1) $\mathrm{rk}(K) < 2k^2$.
2) $\forall i \in [k]$, there is a K-matching π_i between T_1, T_i.
3) (Define $\forall i \in \mathcal{I}, K_i := M(L_K(T_i))$.) The terms $\{K_i \mid i \in \mathcal{I}\}$ are linearly independent.

Proof. For convenience, and wlog, assume $\mathcal{I} = [k']$. The proof is an iterative process with at most k'^2 iterations, and gradually builds the promised space K. Each iteration of the process maintains a space U of $L(R)$ which is intended to grow at each step and bring us closer to K. For convenience, define $U_i := M(L_U(T_i))$, for all $i \in [k']$. Also for each $i \in \{2, \ldots, k'\}$, define ideal $\mathcal{I}_i := \langle U_1, \ldots, U_{i-1} \rangle$.

The process has two nested iterations, or phrased differently, a double induction. We will call the outer "loop" a *phase*, and the inner loop a *round*. In each round the rank of U increases by at most 1, and the i-th phase has at most i rounds. At the end of the i-th phase $(i \geq 2)$, we will ensure $T_i \notin \mathcal{I}_i$. (Remark: By Lemma 41 this is equivalent to ensuring $U_i \notin \mathcal{I}_i$, which by Lemma 42 means that U_i is linearly independent of $U_1, \ldots, U_{i-1}$.)

In the first phase we set $U := K'$, where K' is the matching-nucleus obtained by applying Theorem 8 on C. This immediately gives us property (2) promised in the theorem statement, i.e. the matching property. Also, $\mathrm{rk}(U) < k^2$ at the end of the first phase.

Now the second phase. As T_1, T_2 are linearly independent, we get, by Lemma 42, that $T_2 \notin \langle T_1 \rangle$. By an application of Corollary 23, $\exists v \in \mathrm{nod}_{\langle 0 \rangle}(T_1)$ such that $T_2 \notin \langle v \rangle$. We update $U \leftarrow (U + \mathrm{radsp}(v))$. Note that after updation $T_2 \notin \langle U_1 \rangle = \mathcal{I}_2$ (otherwise $T_2 \in \langle U_1 \rangle \subseteq \langle v \rangle$, since $v|U_1$).

Now, for the $i > 2$ phase. Inductively, we assume that $\forall r < i$, $T_r \notin \mathcal{I}_r$ (remember that all these ideals are wrt the *current U*). The phase consists of various rounds. At the end of the j-th round ($1 \leqslant j < i$), we just want to ensure $T_i \notin \langle U_1, \ldots, U_j, T_{j+1}, \cdots, T_{i-1} \rangle$. So we do nothing in the j-th round unless this is violated. What do we do when it is violated?

Claim 26. *Let $i > 2$ and $1 \leqslant j < i$. Suppose $\forall r < i$, $T_r \notin \langle U_1, \ldots, U_{r-1} \rangle$. Suppose $T_i \in \langle U_1, \cdots, U_j, T_{j+1}, \cdots, T_{i-1} \rangle$ but $T_i \notin \langle U_1, \cdots, U_{j-1}, T_j, \cdots, T_{i-1} \rangle$. There exists a $v \in \mathrm{nod}_{\langle U_1, \cdots, U_{j-1} \rangle}(T_j)$ such that for the updated $U' \leftarrow (U + \mathrm{radsp}(v))$ we have $T_i \notin \langle U'_1, \cdots, U'_j, T_{j+1}, \cdots, T_{i-1} \rangle$.*

Proof of Claim 26. Since $T_i \in \langle U_1, \cdots, U_j, T_{j+1}, \cdots, T_{i-1} \rangle$, by Lemma 42, we get $T_i + \sum_{r=j+1}^{i-1} \alpha_r T_r \in \langle U_1, \cdots, U_j \rangle$ for some α_r-s in $\mathbb{F}$. Suppose there are two distinct choices for α_r-s (we will call them α_r and α'_r). Then,

$$\left(T_i + \sum_{r=j+1}^{i-1} \alpha_r T_r \right), \ \left(T_i + \sum_{r=j+1}^{i-1} \alpha'_r T_r \right) \in \langle U_1, \cdots, U_j \rangle.$$

Subtracting, we get $\sum_{r=j+1}^{i-1}(\alpha - \alpha'_r)T_r \in \langle U_1, \cdots, U_j \rangle$. Let s be the largest index such that $\alpha_s - \alpha'_s \neq 0$. (By the distinctness of the sequences, such an index exists.) We get that $T_s \in \langle U_1, \cdots, U_j, T_{j+1}, \cdots, T_{s-1} \rangle \subseteq \langle U_1, \cdots, U_{s-1} \rangle$. Since $s \leq i-1$, this contradicts the hypothesis. Hence, the sequence $\{\alpha_r\}$ is unique.

The claim hypothesis says that $T_i \notin \langle U_1, \cdots, U_{j-1}, T_j, \cdots, T_{i-1} \rangle$. That implies $T_i + \sum_{r=j+1}^{i-1} \alpha_r T_r \notin \langle U_1, \cdots, U_{j-1}, T_j \rangle$. Thus, by Corollary 23, $\exists v \in \mathrm{nod}_{\langle U_1, \cdots, U_{j-1} \rangle}(T_j)$ such that $T_i + \sum_{r=j+1}^{i-1} \alpha_r T_r \notin \langle U_1, \cdots, U_{j-1}, v \rangle$. Let us update U to $U' \leftarrow (U + \mathrm{radsp}(v))$. (This updates U_r-s to U'_r-s.)

We now argue that $T_i \notin \langle U'_1, \ldots, U'_j, T_{j+1}, \cdots, T_{i-1} \rangle$. Suppose not. Then, by Lemma 42, for some sequence β_r, $T_i + \sum_{r=j+1}^{i-1} \beta_r T_r \in \langle U'_1, \ldots, U'_j \rangle \subseteq \langle U_1, \ldots, U_j \rangle$ (since for all r, $U_r | U'_r$). By the uniqueness of $\{\alpha_r\}$, we have $\beta_r = \alpha_r$, for all r. But that implies $T_i + \sum_{r=j+1}^{i-1} \alpha_r T_r \in \langle U'_1, \ldots, U'_j \rangle \subseteq \langle U_1, \cdots, U_{j-1}, v \rangle$. This is a contradiction and hence completes the proof. $\square$

Let us look at the first round (i.e. $j = 1$). Suppose $T_i \notin \langle U_1, T_2, \cdots, T_{i-1} \rangle$. Then, we move directly to the second round, since we have already satisfied the round invariant. Otherwise, $T_i \in \langle U_1, T_2, \cdots, T_{i-1} \rangle$. Furthermore, by linear independence and Lemma 42, we have $T_i \notin \langle T_1, \cdots, T_{i-1} \rangle$, so we can invoke Claim 26 to get a $v \in \mathrm{nod}_{\langle 0 \rangle}(T_1)$. This allows us to update $U \leftarrow (U + \mathrm{radsp}(v))$ such that $T_i \notin \langle U_1, T_2, \cdots, T_{i-1} \rangle$.

Now for the induction step. We assume that, by the end of the $(j-1)$th round, $T_i \notin \langle U_1, \cdots, U_{j-1}, T_j, \cdots, T_{i-1} \rangle$. For the j-th round, either we would have to do nothing or have to apply Claim 26 and update U. In either case, $\mathrm{rk}(U)$ increases by at most 1. At the end of the round, $T_i \notin \langle U_1, \cdots, U_j, T_{j+1}, \cdots, T_{i-1} \rangle$.

This continues till $j = i - 1$. We finally have $T_i \notin \langle U_1, \cdots, U_{i-1} \rangle = \mathcal{I}_i$, giving us the required invariant for the i-th phase. This completes the proof. $\square$

5. INVOKING SYLVESTER-GALLAI THEOREMS: THE FINAL RANK BOUND

In this section we will bound the non-nucleus rank of a simple, minimal $\Sigma\Pi\Sigma(k,d)$, independent-fanin k', identity C by $(k-k') \cdot \mathrm{SG}_{k'}(\mathbb{F}, d)$. Thus, proving Theorem 18. We divide the proof into two subsections. First, we bound the non-nucleus rank of a simple, *strongly* minimal $\Sigma\Pi\Sigma(k,d)$ identity C by $\mathrm{SG}_{k-1}(\mathbb{F}, d)$, finishing the proof of Theorem 16. Second, we show how to repeatedly use this result on a simple, minimal but not strongly-minimal identity.

5.1. **The strongly minimal case.** Assume that $C := \sum_{i\in[k]} T_i$ is a simple, strongly minimal $\Sigma\Pi\Sigma(k,d)$ identity (recall: then $T_1, \ldots, T_{k-1}$ are linearly independent polynomials). Let K be its nucleus given by Theorem 10. There are two important properties of this nucleus that we restate (and elaborate upon) for emphasis.

The first is the *matching property*. For any $i \in [k]$, $L_K^c(T_1)$ $(= L(T_1) \setminus K)$ is K-matched to $L_K^c(T_i)$ $(= L(T_i) \setminus K)$. In other words for any $\ell \in L_K^c(T_1)$, the degrees of $M(L_K^c(T_1) \cap (\mathbb{F}^*\ell + K))$ and $M(L_K^c(T_i) \cap (\mathbb{F}^*\ell + K))$ are equal (remark: they are polynomials in $\mathrm{nod}_K(T_1)$ and $\mathrm{nod}_K(T_i)$ respectively). This observation motivates the following definition.

Definition 27 (Family). *Let C be a $\Sigma\Pi\Sigma(k,d)$ identity and K be its nucleus. Let $\ell \in L_K^c(C)$. The* family *of ℓ is defined to be the list, $\mathrm{fam}(\ell) := \{M(L_K^c(T_i) \cap (\mathbb{F}^*\ell + K)) \mid i \in [k]\}$. Note that $\mathrm{fam}(\ell)$ is a multiset of size exactly k, having equal degree polynomials corresponding to each term T_i, we fix this ordering on the list (i.e. i-th element in $\mathrm{fam}(\ell)$ corresponds & divides the multiplication term T_i).*

Verify that any two forms in $L_K^c(C)$ that are "similar mod K" have the same families.

[**Partition, Class, Split & Preserve**] Let us focus on a list $\mathrm{fam}(\ell)$. The equivalence relation of similarity (i.e. mod $\langle 0 \rangle$) on $\mathrm{fam}(\ell)$, induces a *partition* of $[k]$ (i.e. if $f_i, f_j \in \mathrm{fam}(\ell)$ are similar then place i and j in the same partition-class). Denote this partition induced on $[k]$, by $\mathrm{Part}(\ell)$. Observe that $\mathrm{Part}(\ell)$ must contain at least 2 classes (otherwise simplicity of C is violated).

Each set in this partition is called a *class*, and we naturally have a class $\mathrm{cl}(f)$ associated with each member of $f \in \mathrm{fam}(\ell)$.

We say that $\mathrm{Part}(\ell)$ *splits* a subset $S \subseteq [k]$ if there is some class $X \in \mathrm{Part}(\ell)$ such that $X \cap S \neq \emptyset, S$. Otherwise, we say that $\mathrm{Part}(\ell)$ *preserves* S. Note that a singleton is always preserved.

For classes $A_1 \in \mathrm{Part}(\ell_1)$ and $A_2 \in \mathrm{Part}(\ell_2)$, the *complement* $\overline{A_1 \cup A_2}$ is just the set $[k] \setminus (A_1 \cup A_2)$. We will be later interested in the properties of this complement set wrt the two partitions.

The second property of the nucleus, the *linear independence*, says something technical about the nucleus identity. By definition $K_i = M(L_K(T_i))$, for all $i \in [k]$, and by Lemma 46 : $\sum_{i\in[k]} \alpha_i K_i = 0$ for some α_i-s $\in \mathbb{F}^*$. Furthermore,

Claim 28. *For $1 < r < k$, let $\{s_1, \cdots, s_r\}$ be a subset $S \subsetneq [k]$, where $s_1 < s_2 < \cdots < s_r$. Then $K_{s_r} \notin \langle K_{s_1}, \cdots, K_{s_{r-1}} \rangle$.*

Proof. If $s_r < k$, then this just holds from the linear independence of $\{K_1, \ldots, K_{k-1}\}$ and Lemma 42. So, we can assume $s_r = k$ and $K_k \in \langle K_{s_1}, \cdots, K_{s_{r-1}} \rangle$. By Lemma 42, this means $K_k = \sum_{i\in[r-1]} \beta_{s_i} K_{s_i}$ for some β-s $\in \mathbb{F}$. The nucleus identity gives us

$K_k = -\sum_{i \in [k-1]} \frac{\alpha_i}{\alpha_k} K_i = \sum_{i \in [r-1]} \beta_{s_i} K_{s_i}$. Since $r < k$, this implies that for some γ-s in $\mathbb{F}$, not all zero, $\sum_{i \in [k-1]} \gamma_i K_i = 0$. This contradicts the linear independence of $\{K_1, \ldots, K_{k-1}\}$, finishing the proof. $\qquad\square$

Before applying Sylvester-Gallai-type theorems (i.e. the SG_{k-1} operator) we emphasize that, as discussed in Section 2.3, there is a distinguished linear form $y_0 \in L(R)^*$ and a subspace U of $L(R)$ such that $L(R) = \mathbb{F}y_0 \oplus U \oplus K$ and every form in $L_K^c(C)$ is monic wrt y_0. Thus, for every $\ell \in L_K^c(C)$ there exists a unique way to express : $\ell = \alpha y_0 + u + v$ ($\alpha \in \mathbb{F}^*$, $u \in U$ and $v \in K$). This allows us to define the truncation operator : $\mathrm{trun}(\ell) = y_0 + \alpha^{-1}u$.

Lemma 29 (Partitions from SG_{k-1}-tuple). *Suppose $rk(\mathrm{trun}(L_K^c(T_1))) > SG_{k-1}(\mathbb{F}, d)$, and $SG_{k-1}(\mathrm{trun}(L_K^c(T_1)))$ gives the set $\{\ell_1, \ell_2, \cdots, \ell_{k-1}\}$. For all $i \in [k-1]$, let $\ell_i' \in L_K^c(T_1)$ be a form satisfying $\mathrm{trun}(\ell_i') = \ell_i$.*

Let $\mathcal{I} \subseteq [k-1]$ be nonempty, and A_i be any class in $\mathrm{Part}(\ell_i')$ for all $i \in \mathcal{I}$. Suppose $S := \overline{\bigcup_{i \in \mathcal{I}} A_i} \neq \emptyset$. Then S is split by $\mathrm{Part}(\ell_c')$, for some $c \in \mathcal{I}$.

Proof. We prove by contradiction. Suppose S is preserved by $\mathrm{Part}(\ell_i')$, for all $i \in \mathcal{I}$. Since for all $i \in \mathcal{I}$, $A_i \in \mathrm{Part}(\ell_i')$, by definition there exists an $f_i \in \mathrm{fam}(\ell_i')$ such that $A_i = \mathrm{cl}(f_i)$. Similarly, for all $i \in \mathcal{I}$, there exists a $g_i \in \mathrm{fam}(\ell_i')$ such that $S \subseteq \mathrm{cl}(g_i)$. Note that, by definition, sets A_i and S are disjoint, hence the classes $\mathrm{cl}(f_i)$ and $\mathrm{cl}(g_i)$ are different, implying f_i, g_i are *not* similar, for all $i \in \mathcal{I}$.

Define ideal $I := \langle f_i | i \in \mathcal{I} \rangle$. Let us focus on the sub-circuit $C_S = \sum_{j \in S} T_j$. Since $C = 0$ and $S = \overline{\bigcup_{i \in \mathcal{I}} \mathrm{cl}(f_i)}$, we deduce $C_S \in I$ (as f_i "kills" the term T_r for all $r \in \mathrm{cl}(f_i)$, and "spares" the other terms). For all $i \in \mathcal{I}$, $S \subseteq \mathrm{cl}(g_i)$ we deduce that : $\prod_{i \in \mathcal{I}} g_i$ divides T_j, for all $j \in S$. So $T_j' := T_j / (\prod_{i \in \mathcal{I}} g_i)$ is again a multiplication term with none of its form in $\bigcup_{i \in \mathcal{I}} (\mathbb{F}^* \ell_i' + K) = \bigcup_{i \in \mathcal{I}} (\mathbb{F}^* \ell_i + K)$. Thus, we get an important equation:

$$C_S = \left(\prod_{i \in \mathcal{I}} g_i\right) \cdot \left(\sum_{j \in S} T_j'\right) \in \langle f_i \mid i \in \mathcal{I} \rangle.$$

By a repeated application of Lemma 47 on the above system, we get :

$$(2) \qquad \sum_{j \in S} T_j' \in \langle f_i' \mid i \in \mathcal{I} \rangle =: I', \text{ where, } f_i' := \frac{f_i}{\gcd(f_i, g_i)}, \forall i \in \mathcal{I}.$$

Since f_i, g_i are not similar, f_i' has degree ≥ 1, for all $i \in \mathcal{I}$. Let the elements of S be $s_1 < s_2 < \cdots < s_r$, for some $r \in [k-1]$. Since we have only changed the non-nucleus part of T_j to get T_j', we deduce $K_{s_i} | T_{s_i}'$, for all $i \in [r]$. Thus, modulo the ideal $I'' := \langle I', K_{s_1}, \cdots, K_{s_{r-1}} \rangle$, Equation (2) becomes : $T_{s_r}' \in I''$. We have $\mathrm{radsp}(I'') \subseteq \mathrm{sp}(\ell_i \mid i \in \mathcal{I}) + K$. Let us factor $T_{s_r}' = B_0 B_1$, where B_0 is the product of all forms in $\mathrm{radsp}(I'')$ and B_1 is the remaining product. Thus, $B_0 B_1 \in I''$. By Lemma 41, B_1 can be cancelled out and we get $B_0 \in I''$.

Suppose all forms of B_0 are in K, so $B_0 = K_{s_r}$. This means $K_{s_r} \in I''$ implying,

$$(3) \qquad\qquad K_{s_r} \in \langle K_{s_1}, \cdots, K_{s_{r-1}}, \{f_i' \mid i \in \mathcal{I}\} \rangle.$$

Recall that each form in f_i' is similar to some form in $(\mathbb{F}^* \ell_i + K)$, for all $i \in \mathcal{I}$. Suppose form $(\beta_i \ell_i + u_i) | f_i'$, for all $i \in \mathcal{I}$, for some β-s in $\mathbb{F}^*$ and u-s in K. In Equation (3) make the *evaluation* : $\ell_i \leftarrow -\beta_i^{-1} u_i$, for all $i \in \mathcal{I}$. This is a valid evaluation since $\{\ell_i \mid i \in \mathcal{I}\}$ are linearly independent mod K, and values substituted are from K. Clearly, this evaluation leaves the polynomial K_s ($s \in S$) unchanged. Thus, we get $K_{s_r} \in \langle K_{s_1}, \cdots, K_{s_{r-1}} \rangle$, contradicting Claim 28.

As a result, we have a form $\ell | B_0$ such that $\ell \notin K$. We have $\ell \in \text{radsp}(I'') \subseteq \text{sp}(\ell_i \mid i \in \mathcal{I}) + K$, and by the way T'_{s_r} was defined, $\ell \notin \bigcup_{i \in \mathcal{I}}(\mathbb{F}^* \ell_i + K)$. By the matching property of the nucleus, this gives us an $\ell' \in L^c_K(T_1)$ such that : $\ell' \in (\text{sp}(\ell_i | i \in \mathcal{I}) + K) \setminus K$ and $\ell' \notin \bigcup_{i \in \mathcal{I}}(\mathbb{F}^* \ell_i + K)$. This means that there exist constants β_i-s in $\mathbb{F}$, not all zero, such that $\ell' \in \sum_{i \in \mathcal{I}} \beta_i \ell_i + K$. As the coefficient of y_0 in ℓ' is nonzero while that in ℓ_i $(i \in \mathcal{I})$ is 1, we deduce : $\text{trun}(\ell') = (\sum_{i \in \mathcal{I}} \beta_i)^{-1}(\sum_{i \in \mathcal{I}} \beta_i \ell_i)$. If exactly one β_i is nonzero, then $\ell' \in (\mathbb{F}^* \ell_i + K)$, which is a contradiction. So at least two β_i-s are nonzero, implying that $\text{trun}(\ell') \in \text{trun}(L^c_K(T_1))$ is a non-trivial combination of the ℓ_i-s, contradicting the fact that $\{\ell_1, \cdots, \ell_{k-1}\}$ were obtained from $\text{SG}_{k-1}(\text{trun}(L^c_K(T_1)))$.

This contradiction proves that S is split by $\text{Part}(\ell'_i)$, for some $i \in \mathcal{I}$. $\square$

To prove Theorem 16, we need a combinatorial lemma about general partitions. It is helpful to abstract out some of the details specific to identities and frame this as a purely combinatorial problem. Since the proof is fairly involved, we present that in the next subsection. For now, we give the necessary definitions and claims. We have a universe $\mathcal{U} := [k]$ of elements. We deal with set systems with special properties.

Definition 30 (Unbroken chain). *A partition of $\mathcal{U}$ is* trivial *if it contains the single set $\mathcal{U}$.*

Let $\mathfrak{P}$ be a collection of non-trivial partitions of $\mathcal{U}$ (here a collection refers to a multiset, *i.e. $\mathfrak{P}$ can have partitions repeated). A* chain *in $\mathfrak{P}$ is a sequence of sets $A_1, A_2, \cdots, A_s$ (for some s) such that each set comes from a different element of $\mathfrak{P}$ (say $A_i \in \mathcal{P}_i \in \mathfrak{P}$).*

The chain $A_1, A_2, \cdots, A_s$ is an unbroken chain, *if $\overline{\bigcup_{i \in [s]} A_i}$ is non-empty and preserved in $\mathcal{P}_i$, for each $i \in [s]$.*

Note that if $\overline{\bigcup_{i \leq s} A_i}$ is a singleton then it is trivially preserved in any partition, therefore, such a chain would be unbroken. By Lemma 29, the collection $\{\text{Part}(\ell'_i) | i \in [k-1]\}$ has no unbroken chain. By purely studying partitions, we will show that such a phenomenon is absurd. The following combinatorial lemma implies Theorem 16.

Lemma 31 (Partitions have unbroken chain). *Let $\mathfrak{P}$ be a collection of non-trivial partitions of $\mathcal{U}$. If $\mathfrak{P}$ contains at least $|\mathcal{U}| - 1$ partitions then $\mathfrak{P}$ contains an unbroken chain.*

Theorem 16. (Bound for simple, strongly minimal identities). *Let $|\mathbb{F}| > d$. The non-nucleus rank of a simple and strongly minimal $\Sigma\Pi\Sigma(k, d)$ identity over $\mathbb{F}$ is at most $\text{SG}_{k-1}(\mathbb{F}, d)$.*

Proof. (of Theorem 16) Let $C = \sum_{i \in [k]} T_i$ be a simple and strongly minimal $\Sigma\Pi\Sigma(k, d)$ identity over $\mathbb{F}$, and let K be the nucleus provided by Theorem 10. As $|\mathbb{F}| > d$ we can assume (wlog by Lemma 40) the existence of a truncation operator on $L^c_K(T_1)$. We will show that the rank of $\text{trun}(L^c_K(T_1))$ is at most $\text{SG}_{k-1}(\mathbb{F}, d)$. By the matching property of the nucleus, $\text{trun}(L^c_K(T_1))$ together with K span $L(C)$. Therefore, a non-nucleus rank bound of the former suffices to bound the non-nucleus rank of $L(C)$.

Assuming that the rank of $\text{trun}(L^c_K(T_1))$ is greater than $\text{SG}_{k-1}(\mathbb{F}, d)$, as in Lemma 29, we invoke $\text{SG}_{k-1}(\text{trun}(L^c_K(T_1)))$ to get $\{\ell_1, \ell_2, \cdots, \ell_{k-1}\}$. Associated with each of these, we have the partition $\text{Part}(\ell'_i)$. There are $k - 1$ partitions in the collection $\mathfrak{P} := \{\text{Part}(\ell'_i) | i \in [k-1]\}$, which are all non-trivial by the simplicity of C. Lemma 31 tells us that $\mathfrak{P}$ has an unbroken chain, while Lemma 29 says that $\mathfrak{P}$ has none. This contradiction implies the rank of $\text{trun}(L^c_K(T_1))$ is at most $\text{SG}_{k-1}(\mathbb{F}, d)$, thus finishing the proof. $\square$

5.1.1. *The combinatorial proof of Lemma 31.* Intuitively, when the partitions in $\mathfrak{P}$ have many classes then an unbroken chain should be easy to find, for example, when $(k-1)$ partitions in $\mathfrak{P}$ are all equal to $\{\{1\}, \ldots, \{k\}\}$ then there is an easy unbroken chain, namely $\{1\}, \ldots, \{k-1\}$. On the other hand, when the partitions in $\mathfrak{P}$ contain few classes then we can effectively decrease the universe and apply induction. Most of this subsection would deal with the former case. Let us first define the splitting property.

Definition 32 (Splitting property)**.** *Let $\mathfrak{P}$ be a collection of partitions of $\mathcal{U}$. Suppose for all non-empty $S \subset \mathcal{U}$, S is split by at least $(|S|-1)$ partitions of $\mathfrak{P}$. Then $\mathfrak{P}$ is said to have the* splitting property*.*

Claim 33. *Let $\mathfrak{P}$ be a collection of at least $(k-1)$ non-trivial partitions of $[k]$. If $\mathfrak{P}$ has the splitting property then there is a chain $A_1, \cdots, A_{k-1}$ in $\mathfrak{P}$ such that $\overline{\bigcup_{i \leq k-1} A_i} = \{k\}$. (In particular, $\mathfrak{P}$ has an unbroken chain.)*

We defer its proof and, instead, first show why this claim would suffice.

Proof. (of Lemma 31) We will prove this by induction on the universe size k. For the base case, suppose $k = 3$ and $\mathfrak{P} = \{\mathcal{P}_1, \mathcal{P}_2, \ldots\}$. So we have at least two partitions. If any partition (say $\mathcal{P}_1$) contains exactly 2 sets, it must be a pair and a singleton (say $\mathcal{P}_1 = \{\{1, 2\}, \{3\}\}$). But then $\{1, 2\}$ is itself an unbroken chain in $\mathfrak{P}$. So, all the partitions can be assumed to consist only of singletons. But then we can take the set, say, $\{1\}$ from $\mathcal{P}_1$ and, say, $\{2\}$ from $\mathcal{P}_2$ to get an unbroken chain.

Now for the induction step. Suppose $\mathfrak{P}$ has at least $(k-1)$ partitions. We assume that the claim is true for universes of size upto $(k-1)$. If $\mathfrak{P}$ has the splitting property, then we are done by Claim 33. If not, then for some subset $S \subset \mathcal{U}$ of size at least 2, S is split in at most $(|S|-2)$ partitions. Let the collection of partitions in $\mathfrak{P}$ that preserve S be $\mathfrak{P}'$. So $\mathfrak{P}'$ contains at least $(k-1) - (|S|-2) = (k-|S|+1)$ partitions. Merge the elements of S into a new element, to get a new universe $\mathcal{U}'$ of size $(k-|S|+1)$. The partitions in $\mathfrak{P}'$ are valid partitions of $\mathcal{U}'$, and still maintain their structure. We now have a universe of size $k - |S| + 1 < k$, and at least $k - |S| + 1$ partitions. By the induction hypothesis, there is an unbroken chain in $\mathfrak{P}'$. Observe that it is (under the natural correspondence) still an unbroken chain in the original collection $\mathfrak{P}$, and we are done. $\square$

Proof. (of Claim 33) We will label the partitions in $\mathfrak{P}$ in such a way that its first $(k-1)$ elements, $\mathcal{P}_1, \cdots, \mathcal{P}_{k-1}$ satisfy : $\mathcal{P}_i$ splits $\{i, k\}$, for all $i \in [k-1]$. Thus, there is a set $A_i \in \mathcal{P}_i$ that contains i but not k. Naturally, $\overline{\bigcup_{i \leq k} A_i} = \{k\}$.

We will construct this labelling through an iterative process. In the ith phase, we will find $\mathcal{P}_i$. At the end of this phase, we will have $\mathcal{P}_1, \cdots, \mathcal{P}_i$ with the desired property and the remaining *pool* $\mathfrak{P}$ of remaining partitions. We warn the reader that this labelling is very dynamic, so during the ith phase, we may change the labels of $\mathcal{P}_1, \cdots, \mathcal{P}_{i-1}$ by moving them to $\mathfrak{P}$ and labelling new partitions with older labels. At any stage, we have the *labelled* partitions and the *unlabelled* partitions $\mathfrak{P}$. Before the beginning of the first phase, $\mathfrak{P}$ is just the given collection of all permutations.

[Phase 1] The first phase is easy to understand. By the splitting property, there is some partition that splits $\{1, k\}$. We set this to $\mathcal{P}_1$.

[Phase i] The ith phase, $i \geq 2$, is a rather involved process. We describe the various sets associated with it and explain them. By the beginning of this phase, we have already determined $\mathcal{P}_1, \cdots, \mathcal{P}_{i-1}$. The *covered* elements are just $[i-1]$. We maintain a partition $E_1, \cdots, E_{i-1}$ of the covered elements. We set $E_0 = \{i, k\}$. Corresponding to each set E_j, we have a set of partitions $\mathcal{C}_j$ $(:= \{\mathcal{P}_b | b \in E_j\})$. We fix $\mathcal{C}_0 = \emptyset$. Note that the $\mathcal{C}_j$'s form

a partition of the labelled partitions. We set $E_{\leq j} = \bigcup_{0 \leq l \leq j} E_l$. We get a similar set of partitions $\mathcal{C}_{\leq j} = \bigcup_{0 \leq l \leq j} \mathcal{C}_l$. We will *always* maintain that $\mathcal{P}_j$ splits $\{j, k\}$.

There will be various rounds in a phase. To aid understanding, we will describe the first and second round in detail.

[**Phase** i**, Round** 1] We now explain the first round. Initially, we set $E_1 = [i-1]$. This is because, as of now, we know nothing about the elements in the set containing k in the various $\mathcal{P}_j$'s. Note that *at any stage*, if we have a partition in $\mathfrak{P}$ that splits $\{i, k\}$, we can set this to $\mathcal{P}_i$ and we are done.

Now, for every element b in E_1 (currently, it is $[i-1]$) check if there is a partition in $\mathfrak{P}$ that splits $\{i, k, b\}$. If so, call this a *success* for b. Note that $\{i, k\}$ is not split in any partition $\mathfrak{P}$. Now, we can label *this* partition as $\mathcal{P}_b$ and move the old one to the pool $\mathfrak{P}$. All the labelled partitions still have their desired property. If the old partition splits $\{i, k\}$, then we are done (since this is now in $\mathfrak{P}$). So, we can assume that (even after this switching) that $\{i, k\}$ is preserved in all of $\mathfrak{P}$. For our (new) $\mathcal{P}_b$, we know that $\{i, k\}$ is preserved. So we have some extra information about it. This is represented by "promoting" b from E_1 to E_2. This just involves removing b from E_1 and putting it in E_2. Let us repeat this for all elements in E_1 until we have a maximal set E_2, containing all successful elements. Note that when $\mathfrak{P}$ changes because of the switching, we check all elements in E_1 again for successes.

We are now at the end of this round and have the following information. $\{E_1, E_2\}$ is a partition of $[i-1]$. For any successful element $b \in E_2$, $\mathcal{P}_b$ preserves $\{i, k\}$. So, the labelled partitions $\mathcal{C}_2$ preserve $\{i, k\} = E_0$. For every failure $b \in E_1$, *every* partition in $\mathfrak{P}$ preserves $\{i, k, b\}$. In other words, every partition in $\mathfrak{P}$ preserves $E_1 \cup \{i, k\} = E_{\leq 1}$. Successes create the new E_2, while failures increase the size of the set preserved by $\mathfrak{P}$.

Let us understand this a little more. Suppose *all* elements are eventually successful, so $E_2 = [i-1]$. Therefore, all labelled partitions preserve $\{i, k\}$. But so do all partitions in $\mathfrak{P}$. So $\{i, k\}$ is preserved in *all* partitions, contradicting the splitting property. There must be some failures. Suppose everything is a failure, so E_1 is still $[i-1]$. The set $E_{\leq 1}$ has size $i+1$. But the only partitions that split $E_{\leq 1}$ are the labelled ones since $\mathfrak{P}$ preserves $E_{\leq 1}$. There are only $i-1$ labelled partitions so this contradicts the splitting property. So there are some successes and some failures and E_1, E_2 form a non-trivial partition. In some sense, we made "progress".

[**Phase** i**, Round** 2] We move to the next round. For every $b \in E_2$, we check if $E_{\leq 1} \cup \{b\}$ is split in any partition of $\mathfrak{P}$. If we get a success, then we set this partition to be the new $\mathcal{P}_b$. We "promote" b from E_2 to a new set E_3. We need to shift this old partition (call it $\mathcal{P}$) to our pool $\mathfrak{P}$. But we want to ensure that $E_{\leq 1}$ is preserved in all of $\mathfrak{P}$, and this may not happen for $\mathcal{P}$. So, first we check if $\{i, k\}$ is preserved in $\mathcal{P}$. If not, we are done. Assume otherwise. We start checking if $\{i, k, c\}$ is preserved, for all $c \in E_1$. If it is so for all c, then we know that $E_{\leq 1}$ is preserved in $\mathcal{P}$. So, we maintain our condition about $\mathfrak{P}$, and we continue to the next b. If not (this is the interesting part!), then we have found a partition that separates c from $\{i, k\}$. Note that the reason why c belongs to E_1, is, because we were unable (in the previous round) to find such a partition. So, we label $\mathcal{P}$ as $\mathcal{P}_c$. We "promote" c from E_1 to E_2. The old $\mathcal{P}_c$ is moved to the pool $\mathfrak{P}$, so we repeat the above procedure for this partition as well. So, either we maintain the invariant that $E_{\leq 1}$ is preserved in all of $\mathfrak{P}$, or we promote elements from E_1 to E_2. If, at some stage, there are no elements in E_1, then we are done. (Why? Because every labelled partition now preserves $\{i, k\}$, by the splitting property, there must be a partition in $\mathfrak{P}$ splitting this.) For all the failures $b \in E_1$, we know that $E_{\leq 1} \cup \{b\}$ is preserved in $\mathfrak{P}$. All

successes are promoted to E_2. So at the end we have the partition E_1, E_2, E_3 of $[i-1]$. All of $\mathfrak{P}$ preserves $E_{\leq 2}$. The partitions $\mathcal{C}_3$ preserve $E_{\leq 1}$ and those in $\mathcal{C}_2$ preserve $E_{\leq 0}$. If E_3 is empty, then $E_{\leq 2}$ is of size $i+1$. All of $\mathfrak{P}$ preserves $E_{\leq 2}$ so the splitting property is violated. If E_1 is empty, then also we are done.

Let us give a formal proof by describing the invariant at the end of a round.

Claim 34. *By the end of the (ith phase-) jth round, suppose we do not find the right $\mathcal{P}_i$. Then we can construct a partition of $[i-1]$, $E_1, \cdots, E_{j+1}$, where E_{j+1} is non-empty and the following hold: the partitions $\mathcal{C}_l$ preserve $E_{\leq l-2}$, for all $2 \leqslant l \leqslant (j+1)$, and the unlabelled partitions $\mathfrak{P}$ preserve $E_{\leq j}$.*

Proof. We prove by induction on j. We have already proven this for $j = 1, 2$. Assuming this is true upto j, we will show this for $j+1$. The round repeatedly "processes" elements of E_{j+1}. Processing $b \in E_{j+1}$ involves checking if all partitions in $\mathfrak{P}$ preserve $E_{\leq j} \cup \{b\}$. If they do, then b is a failure. If $\mathcal{P} \in \mathfrak{P}$ splits $E_{\leq j} \cup \{b\}$, then we "swap" it with $\mathcal{P}_b$, i.e. $\mathcal{P}$ is now the old $\mathcal{P}_b$ and is denoted a *hanging partition*. The element b is promoted from E_{j+1} to the new set E_{j+2}. How to deal with the hanging partition $\mathcal{P}$? We first check if it splits E_0. If so, we have found $\mathcal{P}_i$. Otherwise, we check if it preserves $E_0 \cup \{c\}$, for all $c \in E_1$. If it splits $\{i, k, c\}$, then we swap $\mathcal{P}$ with $\mathcal{P}_c$. We promote c from E_1 to E_2. The old $\mathcal{P}_c$ becomes the new hanging partition $\mathcal{P}$. If $\mathcal{P}$ preserves $E_{\leq 1}$, then we move on to E_2. In general, if $\mathcal{P}$ preserves $E_{\leq l}$, then we check if $\mathcal{P}$ preserves all $E_{\leq l} \cup \{c\}$, for $c \in E_{l+1}$. If $\mathcal{P}$ splits $E_{\leq l} \cup \{c\}$, we swap $\mathcal{P}$ with $\mathcal{P}_c$ and promote c from E_{l+1} to E_{l+2}. Note that the sets $E_{\leq p}$ (for any p) can only decrease on such a promotion. So still the partitions in $\mathcal{C}_p$ preserve $E_{\leq p-2}$. The old $\mathcal{P}_c$ becomes the new hanging partition $\mathcal{P}$ and we repeat this process. If, on the other hand, $\mathcal{P}$ preserves all $E_{\leq l} \cup \{c\}$, then $\mathcal{P}$ preserves $E_{\leq l+1}$. So we repeat this process with E_{l+2}, and so on. If we end up with $\mathcal{P}$ preserving $E_{\leq j}$, then we can safely move $\mathcal{P}$ into $\mathfrak{P}$. Otherwise, we have made a promotion and we deal with a new hanging partition. Note that when $\mathfrak{P}$ changes, we again process all elements in E_{j+1}. There can only be a finite number of promotions, so this round must end. We end up with $E_0, \cdots, E_{j+2}$, with $\mathcal{C}_l$ preserving $E_{\leq l-2}$. All the failures are still in E_{j+1}, and $\mathfrak{P}$ preserves all $E_{\leq j} \cup \{c\}$, $\forall c \in E_{j+1}$. So $\mathfrak{P}$ now preserves $E_{\leq j} \cup E_{j+1} = E_{\leq j+1}$. Note that if E_{j+2} is empty, we have a contradiction. This is because $E_{\leq j+1}$ is of size $i+1$ and there are at most $i-1$ partitions splitting it. $\qquad \square$

Now we show that in this phase $i \geq 2$ there can be at most i rounds before we get the desired $\mathcal{P}_i$.

Claim 35. *Suppose $E_1, \cdots, E_j$ is a partition of $[i-1]$ such that C_l preserves $E_{\leq l-2}$ and $\mathfrak{P}$ preserves $E_{\leq j}$. Then all E_l's are non-empty.*

Proof. Suppose E_l is empty, for some $l \in [j]$. So C_l is also empty. Any partition that is not in $C_{\leq l-1}$ is either in $\mathfrak{P}$ or in C_p, for some $p \geq l+1$ (if it exists). All these partitions preserve $E_{\leq l-1}$. Thus, the only partitions splitting $E_{\leq l-1}$ are those in $C_{\leq l-1}$. Since $|C_{\leq l-1}| = |E_{\leq l-1}| - 2$, we contradict the splitting property. $\qquad \square$

The sets $E_1, \cdots, E_j$ form a partition of $[i-1]$. The above claim tells us that we can run at most $i-1$ rounds to completion. Hence, if we do not find $\mathcal{P}_i$ by $i-1$ rounds, then, by Claim 34, we will find it in the ith round. This completes the proof. $\qquad \square$

5.2. **The general case.** Now, we deal with simple, minimal identities and remove the strong minimality condition. This will come at a cost of an extra k factor in the rank

bound. First, we recall the definition of gcd and simple parts of a general $\Sigma\Pi\Sigma$ circuit, as given in older works [DS06, SS09].

Definition 36 (Gcd & Simple part). *Let $C = \sum_{i\in[k]} T_i$ be a $\Sigma\Pi\Sigma(k,d)$ circuit over a field $\mathbb{F}$. The gcd of C is defined to be the usual gcd of the polynomials T_i-s, i.e. $\gcd(C) := \gcd(T_i | i \in [k])$.*

The simple part of C is the $\Sigma\Pi\Sigma(k,d')$ circuit, $\mathrm{sim}(C) := C/\gcd(C)$, where $d' := d - \deg(\gcd(C))$.

The following will be shown to be a consequence of Theorem 16.

Theorem 18. (Final bound). *Let $|\mathbb{F}| > d$. The rank of a simple, minimal $\Sigma\Pi\Sigma(k,d)$, independent-fanin k', identity is at most $2k^2 + (k - k') \cdot \mathrm{SG}_{k'}(\mathbb{F}, d)$.*

Proof. Let circuit C be $T_1 + \cdots + T_k = 0$. Wlog let $T_1, \cdots, T_{k'}$ be a linear basis for $T_1, \cdots, T_k$. Obviously, we have $1 < k' < k$ (first by simplicity and second by zeroness). By Theorem 10, there exists a nucleus K wrt the set $\mathcal{I} := [k']$. The rank of K is at most $2k^2$. So, it remains to bound the non-nucleus rank of C by $(k - k') \cdot \mathrm{SG}_{k'}(\mathbb{F}, d)$.

As $T_1, \cdots, T_{k'}$ form a basis, for each $i \in [k'+1, k]$, there exists $\alpha_{i,j}$-s in $\mathbb{F}$ such that we have a zero circuit $D_i := \sum_{j\in[k']} \alpha_{i,j}T_j + T_i = 0$. Define N_i to be the set of j-s for which $\alpha_{i,j} \neq 0$. Thus,

$$(4) \qquad \forall i \in [k'+1, k], \ D_i = \sum_{j\in N_i} \alpha_{i,j}T_j + T_i = 0$$

Since $\{\alpha_{i,j}T_j \mid j \in N_i\}$ are $|N_i|$ linearly independent terms, we get that D_i is a strongly minimal $\Sigma\Pi\Sigma(|N_i|+1, d)$ identity, for all $i \in [k'+1, k]$. By nucleus properties, $\{K_j | j \in N_i\}$ are linearly independent polynomials, implying that the polynomials $\{K_j/g_i | j \in N_i\}$ are also linearly independent, where $g_i := M(L_K(\gcd(D_i)))$. Thus, the linear space K remains a nucleus of the new identity $\mathrm{sim}(D_i)$, showing at the same time that it is strongly minimal. We conclude that $\mathrm{sim}(D_i)$ is a simple, strongly minimal $\Sigma\Pi\Sigma(k_i, d_i)$ identity with nucleus K (although of $\mathrm{rk} < 2k^2$), $k_i \leqslant (k'+1), d_i \leqslant d$, for all $i \in [k'+1, k]$. Theorem 16 bounds the non-nucleus (non-K to be precise) rank of each of these identities by $\mathrm{SG}_{k'}(\mathbb{F}, d)$.

Suppose a linear form $\ell | \gcd(D_i)$ for all $i \in [k'+1, k]$. Then ℓ divides T_j for all $j \in \bigcup_{i\in[k'+1,k]} N_i \cup [k'+1, k]$. Consider the case $\bigcup_{i\in[k'+1,k]} N_i = [k']$, it means that ℓ divides every term in C, contradicting simplicity. Thus, in that case every linear form ℓ of C appears in at least one of the circuits $\{\mathrm{sim}(D_i) | i \in [k'+1, k]\}$, whose total non-nucleus rank we have already bounded by $(k - k') \cdot \mathrm{SG}_{k'}(\mathbb{F}, d)$, so we will be done.

The case, left to handle, is when : $S := \bigcup_{i\in[k'+1,k]} N_i \subsetneq [k']$. This means, by summing over i in Equation (4), $\sum_{i\in[k'+1,k]} T_i = \sum_{s\in S} \beta_s T_s$, for some β-s in $\mathbb{F}$. Substituting this in the equation $C = 0$ we get,

$$C = C_{[k']} + C_{[k'+1,k]} \ = \ \sum_{i\in[k']} T_i + \sum_{s\in S} \beta_s T_s \ = \ 0.$$

As S is a proper subset of $[k']$, the above equation could only mean that a nontrivial combination of T_i ($i \in [k']$) is vanishing, contradicting the linear independence of those polynomials. Thus, $S = [k']$. This completes the proof. $\qquad\square$

6. Sylvester-Gallai Rank Bounds for any $\mathbb{F}$

We wish to bound $SG_k(\mathbb{F}, m)$, for *any* field $\mathbb{F}$. We will prove the following theorem, which can be seen as the first attempt ever to give a Sylvester-Gallai Theorem for *all*

fields. It is convenient to think of a set of vectors S in $\mathbb{F}^n$ as *multiple-free*: this means that no two vectors in S are scalar multiples of each other.

Theorem 37 (High dimension Sylvester-Gallai for any field). *Suppose $k \in \mathbb{N}^{>1}$ and S is an SG_k-closed set of vectors in $\mathbb{F}^n$ of rank $r \geqslant 9k$. Then, $|S| \geqslant 2^{r/9k}$. In other words, for every $m \in \mathbb{N}^{>1}$, $SG_k(\mathbb{F}, m) \leq 9k \lg m$.*

Remark. This bound is not tight. Over $\mathbb{F}_p$, the best construction we can come up with is an SG_k-closed set with rank around $k + \log_p m$. Consider $\mathbb{F}_p^{k-1+r}$, for some $r > 0$ and $p \nmid (k-1)$. Take the set $S_1 \subset \mathbb{F}_p^{k-1+r}$ of vectors $e_1, e_2, \cdots, e_{k-1}, \frac{1}{k-1} \cdot \sum_{i \in [k-1]} e_i$, where e_i is just the unit vector in the ith direction. Then take the set $S_2 \subset \mathbb{F}_p^{k-1+r}$ of all (non-zero) vectors which have zeroes in the first $(k-1)$ coordinates. Extend this to $\mathbb{F}_p^{k+r}$ by putting a 1 in the new coordinate. This gives a set of non-multiple distinct vectors in $\mathbb{F}_p^{k+r}$. Observe that S_1 is SG_{k-1}-closed and S_2 is SG_2-closed. Hence $S_1 \cup S_2$ is SG_k-closed. The size of S is $(k + p^r - 1)$ and the rank is just $k + r$.

In some sense, bounds for $SG_2(\mathbb{F}, m)$ are already implicit in known theorems (used to prove lower bounds for LDCs). Concretely, Corollary 2.9 of [DS06] can be interpreted as a proof that $SG_2(\mathbb{F}, m) = O(\log m)$. This is an extension of theorems in [GKST02] that prove this for $\mathbb{F}_2$. In the context of SG_2, these proofs can be interpreted as a "doubling trick". In essence, each time we want to increase the rank of an SG_2-closed set by 1, we are forced to double the number of vectors. A naive attempt to implement this for SG_k does not work.

Roughly speaking, we want to argue that if we want to increase the rank of an SG_k-closed set by $(k-1)$, then the size of the set must double. But, when $k \geq 3$, this is not true! It is possible to increase the rank by $(k-1)$ by adding a very small number of vectors. So we have a sort of two-pronged approach. If the size does not increase much, even though the rank increases, then we show that the set has some very special SG properties. Namely, *many* small subspaces of the set are SG_{k-1}-closed. Even though these subspaces can intersect very heavily, we are still able to argue that the set must now be very large.

We will require two auxiliary claims. The first claim is probably of independent interest, but the second is tied to our current approach.

Claim 38. *Suppose $k \in \mathbb{N}^{>1}$ and S is SG_k-closed. Let vectors $e_1, e_2, \cdots, e_r$ be elements of S that form a basis for S. Hence, every element in S is represented by an r-tuple of coordinates in $\mathbb{F}$. There exists some element in S whose representation has at least $r/(k-1)$ non-zero coordinates.*

Proof. Consider any vector $v \in S$. Let $v = \sum_{i=1}^{r} \alpha_i e_i$, for $\alpha_i \in \mathbb{F}$. We denote by $N(v)$ the index set $\{i | \alpha_i \neq 0\}$. In other words, $N(v)$ is the set of indices for which the corresponding coordinates of v are non-zero. Let $v_1 \in S$ be some vector that maximizes $|N(v_1)|$. Choose some $v_2 \in S$ such that $N(v_2) \cap N(v_1) = \emptyset$ and $|N(v_2)|$ is maximized. Such a v_2 exists as long as $|N(v_1)| < r$, as in this case we can get v_2 from $\mathrm{sp}(e_i | i \notin N(v_1))$. Iteratively, choose $v_j \in S$ such that $N(v_j)$ is disjoint to $\bigcup_{l < j} N(v_l)$ and $|N(v_j)|$ is maximized. As long as $\bigcup_{l \leq j} N(v_l) \neq [r]$, we can always choose the next v_{j+1}, again, from $\mathrm{sp}(e_i | i \notin \bigcup_{l \leq j} N(v_l))$. We keep choosing v_j's until we cover all coordinates. At the end, $\bigcup_{l \leq j} N(v_j) = [r]$. Note that the sets $N(v_1), \cdots, N(v_j)$ form a partition of $[r]$.

Suppose $j \geq k$. Then take the vectors $v_1, \cdots, v_k$. They are certainly linearly independent, since they are defined on a disjoint set of coordinates. By the SG_k-closure of S, some

non-trivial linear combination of these vectors exists in S. Suppose some non-zero combination of $v_{i_1}, v_{i_2}, \cdots$, denoted by v, is in S (where $i_1 < i_2 \cdots$). Note that $N(v) \supset N(v_{i_1})$ and $N(v)$ is disjoint to $\bigcup_{l < i_1} N(v_l)$. That contradicts that choice of v_{i_1}. Hence $j < k$. Because the sets $N(v_l)$ form a partition of $[r]$, $|N(v_1)| \geq r/(k-1)$. That completes the proof. $\qquad\square$

Claim 39. *Suppose $k \in \mathbb{N}^{>2}$. Consider a set of linearly independent vectors $e_1, \cdots, e_{r'}$ in S. Let $I \subseteq [r']$ and $|I| \geq r'/4$. Let the set E_I be $\{e_i | i \in I\}$ and S_I be the set $S \cap sp(E_I)$. If for all such I, S_I is SG_{k-1}-closed, then $|S_{[r']}| \geq 2^{r'/8k}$.*

Proof. By Claim 38, for every such I, there is a $v_I \in S$ such that v_I has at least $|I|/(k-2)$ non-zero coordinates (wrt basis E_I). Fix these v_I, for each such I. As before, we denote the set of indices corresponding to non-zero coordinates of v by $N(v)$. We describe a random process to generate a subset of $[r']$. Simply pick each element in $[r']$ independently with probability $1/2$. Let I and J be two sets generated independently this way. A "good" event occurs if $|I| \geq r'/4$ and $v_I \notin S_J$. We will call this the *good event for (I, J)*. How can a good event not happen? Either $|I| < r'/4$, or, if $|I| \geq r'/4$, then $v_I \in S_J$. What is the probability of a "bad" event (call this $\mathcal{E}(I, J)$) happening? This is:

$$\Pr[|I| < r'/4] \; + \; \Pr[|I| \geq r'/4] \cdot \Pr[v_I \in S_J \mid |I| \geq r'/4]$$

Since $\mathbf{E}[|I|] = r'/2$, the probability that $|I| < r'/4$ is at most $e^{-r'/8}$ (by a Chernoff bound, refer to notes [O'D09] for the exact form used). Obviously, $\Pr[|I| \geq r'/4] < 1$. Now assume that $|I| \geq r'/4$. So $|N(v_I)| \geq r'/4(k-2)$. For v_I to be in S_J, J must contain $N(v_I)$. By the random construction of J, the probability of this is at most $2^{-r'/4(k-2)}$. Thus, the probability of a bad event is at most $e^{-r'/8} + 2^{-r'/4(k-2)} \leq 2 \cdot 2^{-r'/4(k-1)}$.

Now, let us choose $q = 2^{r'/8k}$ subsets of $[r']$ independently through this random distribution. Call these $I_1, I_2, \cdots, I_q$. For indices $1 \leqslant a < b \leq q$, let X_{ab} be the indicator random variable for the event $\mathcal{E}(I_a, I_b)$. What is the expected total number of bad events?

$$\mathbf{E}\left[\sum_{1 \leqslant a < b \leq q} X_{ab}\right] = \sum_{1 \leqslant a < b \leq q} \mathbf{E}[X_{ab}] < \frac{q^2}{2} \Pr[\mathcal{E}(I_1, I_2)] \leq 2^{r'/4k} \cdot 2^{-r'/4(k-1)} < 1$$

By the probabilistic method, there exist subsets of $[r']$, $I_1, I_2, \cdots, I_q$ such that no event $\mathcal{E}(I_a, I_b)$ happens. This means all sets I_a are of size at least $r'/4$. Furthermore, for $1 \leqslant a < b \leq q$, $v_{I_a} \notin S_{I_b}$. This means that $v_{I_a} \neq v_{I_b}$. Therefore, there must be at least q distinct vectors in $S_{[r']}$. $\qquad\square$

Proof. (of Theorem 37) It will be convenient to assume that $|S| < 2^{r/9k}$, and arrive at a contradiction. Let $T \subset S$ be a subset of rank $t \leq r/9$. We can construct a basis of S, $e_1, \cdots, e_{r'}, e_{r'+1}, \cdots, e_r$, using elements of S such that : $r' \geq 8r/9$, $\{e_1, \cdots, e_{r'}\}$ are orthogonal to T, and $\{e_{r'+1}, \cdots, e_r\}$ spans T. If $k > 2$ then by applying Claim 39 to $e_1, \cdots, e_{r'}$ we get some $I \subseteq [r']$ such that $|I| \geq r'/4 \geq 2r/9 \geq 2k$, and S_I is *not* SG_{k-1}-closed (recall, $S_I := S \cap \text{sp}(e_i | i \in I)$). There exist linearly independent $v_1, v_2, \cdots, v_{k-1} \in S_I$ such that no non-trivial combination of these are in S_I (and hence in S). On the other hand, if $k = 2$ then define $v_1 := e_1$. Note that $v_1, v_2, \cdots, v_{k-1}$ are all orthogonal to T.

Consider any $v \in T$. By the SG_k-closure of S, there exists some linear combination of $\{v, v_1, \cdots, v_{k-1}\}$ in S. Call this the *image* of v. Take two different $v, v' \in T$. We argue that their images are distinct. Let the image of v be $\alpha v + \sum_{i \leq k-1} \alpha_i v_i$ and the image of v' be $\beta v' + \sum_{i \leq k-1} \beta_i v_i$. Note that $\alpha\beta \neq 0$ by the way we have chosen $v_1, v_2, \cdots, v_{k-1}$. If these images are equal, then $\alpha v - \beta v' = \sum_{i \leq k-1} (\beta_i - \alpha_i) v_i$. Since v and v' are not multiples

of each other, the right hand side is non-zero. The left hand side is a vector that is spanned by $e_{r'+1}, \cdots, e_r$ but the right hand side is spanned by $e_1, \cdots, e_{r'}$. Contradiction. Hence, all images are distinct.

Starting from any subset $T \subset S$ of rank $t \leq r/9$, there exist linearly independent $v_1, \cdots, v_{k-1} \in S$ such that $S \cap \mathrm{sp}\{v_1, \cdots, v_{k-1}, e_{r'+1}, \cdots, e_r\}$ contains at least $2|T|$ vectors. The rank of this intersection is $t + (k-1)$. Starting from T being just one vector, we keep repeating this process. This can go on for h iterations, where h is the smallest integer such that $1 + (k-1)h \geq r/9$. So, $h = \lceil (r/9 - 1)/(k-1) \rceil$ which is at least $r/9k$, since $r \geq 9k$. We have $|S| \geq 2^h \geq 2^{r/9k}$. Contradiction. $\qquad\square$

7. Conclusion

In this work we developed the strongest methods, to date, to study depth-3 identities. The ideal methods hinge on a classification of zerodivisors of the ideals generated by gates of a $\Sigma\Pi\Sigma$ circuit (eg. Lemmas 41, 44 and 47). That is useful in proving an ideal version of Chinese remaindering tailor-made for $\Sigma\Pi\Sigma$ circuits, which is in turn useful to show a connection between all the gates involved in an identity. As a byproduct, it shows the existence of a low rank *nucleus identity* C' sitting inside *any* given $\Sigma\Pi\Sigma(k, d)$ identity C (when C is not minimal, C' can still be defined but it might not be homogeneous). This nucleus identity is quite mysterious and it might be useful for PIT to understand (or classify) it further. For example, **can the rank bound for the nucleus identity be improved to $O(k)$?**

We generalize the notion of Sylvester-Gallai configurations to *any* field and define a parameter $\mathrm{SG}_k(\mathbb{F}, m)$ associated with field $\mathbb{F}$. This number seems to be a fundamental property of a field, and as we show, is very closely related to $\Sigma\Pi\Sigma$ identities. It would be interesting to obtain bounds for $\mathrm{SG}_k(\mathbb{F}, m)$ for different $\mathbb{F}$. For example, as also asked by [KS09b], **can we nontrivially bound the number $\mathrm{SG}_k(\mathbb{F}, m)$ for interesting fields: $\mathbb{C}$, finite fields with large characteristic, or even p-adic fields?** Other than the bounds for $\mathbb{R}$, all that was known before is $\mathrm{SG}_2(\mathbb{C}, m) \leq 3$ [EPS06]. We shed (a little) light on high dimension SG rank bounds by showing $\mathrm{SG}_k(\mathbb{F}, m) = O(k \log m)$. We conjecture: $\mathrm{SG}_k(\mathbb{F}, m)$ is $O(k)$ for zero characteristic fields, while $O(k + \log_p m)$ for fields of characteristic $p > 1$.

We also prove a property of a general collection $\mathfrak{P}$ of partitions of a universe $\mathcal{U}$, namely, if $\mathfrak{P}$ has at least $|\mathcal{U}| - 1$ partitions then it has an unbroken chain. It is tight and gives an idea of how a Sylvester-Gallai configuration in the non-nucleus part of a $\Sigma\Pi\Sigma(k, d)$ identity "spreads" around.

Finally, we ask: **Can the rank bound for simple minimal real $\Sigma\Pi\Sigma(k, d)$ identities be improved to $O(k)$?** The best constructions known, since [DS06], have rank $4(k - 2)$. Likewise, over other fields, our upper bound of $O(k^2 \log d)$ still leaves some gap in understanding the exact dependence on k.

REFERENCES

[AB03]   M. Agrawal and S. Biswas. Primality and identity testing via Chinese remaindering. *Journal of the ACM*, 50(4):429–443, 2003. (first appeared in FOCS 1999).

[Agr05]   M. Agrawal. Proving lower bounds via pseudo-random generators. In *Proceedings of the 25th Annual Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, pages 92–105, 2005.

[Agr06]   M. Agrawal. Determinant versus permanent. In *Proceedings of the 25th International Congress of Mathematicians (ICM)*, volume 3, pages 985–997, 2006.

[AM07]   V. Arvind and P. Mukhopadhyay. The monomial ideal membership problem and polynomial identity testing. In *Proceedings of the 18th International Symposium on Algorithms and Computation (ISAAC)*, pages 800–811, 2007.

[AS09]   M. Agrawal and R. Saptharishi. Classifying polynomials and identity testing. Technical report, IIT Kanpur, http://www.cse.iitk.ac.in/manindra/survey/Identity.pdf, 2009.

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

[BE67]   W. Bonnice and M. Edelstein. Flats associated with finite sets in P^d. *Niew. Arch. Wisk.*, 15:11–14, 1967.

[CK00]   Z. Chen and M. Kao. Reducing randomness via irrational numbers. *SIAM J. on Computing*, 29(4):1247–1256, 2000. (first appeared in STOC 1997).

[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. (first appeared in STOC 2005).

[EPS06]   N. Elkies, L. Pretorius, and K. Swanepoel. Sylvester-gallai theorems for complex numbers and quaternions. *Discrete and computational geometry*, 35(3):361–373, 2006.

[GKST02]   O. Goldreich, H. Karloff, L. Schulman, and L. Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. In *Proceedings of the 17th Annual Computational Complexity Conference (CCC)*, pages 175–183, 2002.

[Han65]   S. Hansen. A generalization of a theorem of Sylvester on the lines determined by a finite point set. *Mathematica Scandinavia*, 16:175–180, 1965.

[KI04]   V. Kabanets and R. Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1):1–46, 2004. (first appeared in STOC 2003).

[KMSV09]   Z. Karnin, P. Mukhopadhyay, A. Shpilka, and I. Volkovich. Deterministic identity testing of depth 4 multilinear circuits with bounded top fan-in. Technical Report TR09-116, ECCC, http://eccc.hpi-web.de/report/2009/116/, 2009.

[KS01]   A. Klivans and D. A. Spielman. Randomness efficient identity testing of multivariate polynomials. In *Proceedings of the 33rd Symposium on Theory of Computing (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. (first appeared in CCC 2006).

[KS08]   Z. 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 Conference on Computational Complexity (CCC)*, pages 280–291, 2008.

[KS09a]   Z. S. Karnin and A. Shpilka. Reconstruction of generalized depth-3 arithmetic circuits with bounded top fan-in. In *Proceedings of the 24th Annual Conference on Computational Complexity (CCC)*, pages 274–285, 2009.

[KS09b]   N. Kayal and S. Saraf. Blackbox polynomial identity testing for depth 3 circuits. In *Proceedings of the 50th Annual Symposium on Foundations of Computer Science (FOCS)*, 2009.

[LV98]   D. Lewin and S. Vadhan. Checking polynomial identities over any field: Towards a derandomization? In *Proceedings of the 30th Annual Symposium on the Theory of Computing (STOC)*, pages 428–437, 1998.

[O'D09]   R. O'Donnell. Lecture 9: The Chernoff bound. http://www.cs.cmu.edu/odonnell/prob/lecture9.pdf, 2009.

[Sax08]   N. Saxena. Diagonal circuit identity testing and lower bounds. In *Proceedings of the 35th Annual International Colloquium on Automata, Languages and Programming (ICALP)*, pages 60–71, 2008.

[Sax09]   N. Saxena. Progress on polynomial identity testing. *Bulletin of the European Association for Theoretical Computer Science (EATCS)- Computational Complexity Column*, (99):49–79, 2009.

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

[Shp09]   A. Shpilka. Interpolation of depth-3 arithmetic circuits with two multiplication gates. *SIAM J. Comput.*, 38(6):2130–2161, 2009. (first appeared in STOC 2007).

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

[SV09]   A. Shpilka and I. Volkovich. Improved polynomial identity testing for read-once formulas. In *Proceedings of the 13th International Workshop on Randomization and Computation (RANDOM)*, pages 700–713, 2009.

[Zip79]   R. Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposium on Symbolic and Algebraic Manipulation (EUROSAM)*, pages 216–226, 1979.

Appendix A. Technical, Algebraic Lemmas

We denote the polynomial ring $\mathbb{F}[x_1, \ldots, x_n]$ by R.

Lemma 40 (Monic forms). *Let $|\mathbb{F}| > d$ and C be a $\Sigma\Pi\Sigma(k,d)$ identity, over $\mathbb{F}$, with nucleus K. Let $y_0 \in L(R)^*$ and U be a subspace of $L(R)$ such that $L(R) = \mathbb{F}y_0 \oplus U \oplus K$. Then there exists an invertible linear transformation $\tau : L(R) \to L(R)$ that fixes K and :*

1) *$\tau(C)$ is also a $\Sigma\Pi\Sigma(k,d)$ identity with nucleus K and the same simplicity, minimality properties.*
2) *Every form in $L_K^c(\tau(C)) = \tau(L_K^c(C))$ is monic wrt y_0.*

Proof. Let $r := \mathrm{rk}(\mathbb{F}y_0 \oplus U)$. Fix a basis $\{y_0, \ldots, y_{r-1}\}$ of $\mathbb{F}y_0 \oplus U$ and let $\overline{y}$ denote the column vector $[y_0 \cdots y_{r-1}]^T$. Let $\ell \in L_K^c(T_1)$. Then there is a unique nonzero (column) vector $\overline{\alpha_\ell} \in \mathbb{F}^r$ and a $v_\ell \in K$, such that $\ell = \overline{\alpha_\ell}^T \cdot \overline{y} + v_\ell$. We intend τ to be a linear transformation that fixes each element in K and maps $\overline{y}$ to $A\overline{y}$ where $A \in \mathbb{F}^{r\times r}$. Such a τ will map ℓ to $\tau(\overline{\alpha_\ell}^T \cdot \overline{y}) + v_\ell = \overline{\alpha_\ell}^T \cdot \tau(\overline{y}) + v_\ell = \overline{\alpha_\ell}^T A\overline{y} + v_\ell$. To make $\tau(\ell)$ monic in y_0 we need to choose A such that the first coordinate in $\overline{\alpha_\ell}^T A$ is nonzero, i.e. $\overline{\alpha_\ell}^T A_{*1} \neq 0$ where A_{*1} is the first column of A. Thus, we want an A such that $\prod_{\ell \in L_K^c(T_1)} \overline{\alpha_\ell}^T A_{*1} \neq 0$.

Now the nonzero multivariate polynomial $f(\overline{Y}) := \prod_{\ell \in L_K^c(T_1)} \overline{\alpha_\ell}^T \overline{Y}$ has degree at most $d < |\mathbb{F}|$. Hence, by the Schwartz-Zippel lemma [Sch80, Zip79] there exists a point $\overline{Y} \in \mathbb{F}^r$ at which f is nonzero. We can fix A_{*1} to be that point. This fixes just one column of A to a nonzero vector and we can arbitrarily fix the rest such that A is an invertible matrix. Thus, the corresponding invertible τ makes each $\ell \in L_K^c(T_1)$ monic in y_0. Since τ fixes the nucleus K, matching property of the nucleus tells us that every form in $L_K^c(\tau(C)) = \tau(L_K^c(C))$ is monic in y_0.

Since τ is an invertible linear transformation, it is actually an automorphism of $L(R)$ and, in particular, the zeroness, simplicity and minimality properties of C are invariant under it. $\square$

An *ideal* I of R with generators $f_i, i \in [m]$, is the set $\{\sum_{i\in[m]} q_i f_i | q_i\text{'s} \in R\}$ and is denoted by the notation $\langle f_1, \ldots, f_m \rangle$. For any $f \in R$, the three notations $f \equiv 0 \pmod{I}$, $f \equiv 0 \pmod{f_1, \ldots, f_m}$ and $f \in I$, mean the same.

An $f \in R$ is called a *zerodivisor* of an ideal I (or mod I) if $f \notin I$ and there exists a $g \in R \setminus I$ such that $fg \in I$.

Let $u, v \in R$. It is easy to see that if u is nonzero mod I and is a *non*-zerodivisor mod I then: $uv \in I$ iff $v \in I$. This can be seen as some sort of a "cancellation rule" for

non-zerodivisors. We show such a cancellation rule in the case of ideals arising in $\Sigma\Pi\Sigma$ circuits.

Lemma 41 (Non-zerodivisor). *Let $f_1, \ldots, f_m$ be multiplication terms generating an ideal I, let $\ell \in L(R)$ and $g \in R$. If $\ell \notin radsp(I)$ then: $\ell g \in I$ iff $g \in I$.*

Proof. Assume $\ell \notin radsp(I)$. If $I = \{0\}$ then the lemma is of course true. So let us assume that $I \neq \{0\}$ and $rk(radsp(I)) =: r \in [n-1]$. As $\ell \notin radsp(I)$ there exists an invertible linear transformation $\tau : L(R) \to L(R)$ that maps each form of $radsp(I)$ to $sp(x_1, \ldots, x_r)$ and maps ℓ to x_n. Now suppose that $\ell g \in I$. This means that there are $q_1, \ldots, q_m \in R$ such that $\ell g = \sum_{i=1}^{m} q_i f_i$. Apply τ on this to get:

$$(5) \qquad x_n g' = \sum_{i=1}^{m} q_i' \tau(f_i).$$

We know that $\tau(f_i)$'s are free of x_n. Express g', q_i'-s as polynomials wrt x_n, say

$$(6) \qquad g' = \sum_{j \geq 0} a_j x_n^j, \text{ where } a_j \in \mathbb{F}[x_1, \ldots, x_{n-1}]$$

$$(7) \qquad q_i' = \sum_{j \geq 0} b_{i,j} x_n^j, \text{ where } b_{i,j} \in \mathbb{F}[x_1, \ldots, x_{n-1}]$$

Now for some $d \geq 1$ compare the coefficients of x_n^d on both sides of Equation (5). We get $a_{d-1} = \sum_{i=1}^{m} b_{i,d} \tau(f_i)$, thus a_{d-1} and $a_{d-1} x_n^{d-1}$ are in $\langle \tau(f_1), \ldots, \tau(f_m) \rangle$. Doing this for all $d \geq 1$, we get $g' \in \langle \tau(f_1), \ldots, \tau(f_m) \rangle$, hence $g = \tau^{-1}(g') \in \langle f_1, \ldots, f_m \rangle = I$. This finishes the proof. $\square$

All the ideals arising in this work are *homogeneous*, i.e. their generators are homogeneous polynomials. These ideals have some nice properties, as shown below. Degree $\deg(\cdot)$ refers to the total degree unless there is a subscript specifying the variable as well.

Lemma 42 (Homogeneous ideals). *Say, $f_1, \ldots, f_m, g$ are homogeneous polynomials in R. Then,*
  1) If $\deg(g) < \deg(f_m)$ then: $g \in \langle f_1, \ldots, f_m \rangle$ iff $g \in \langle f_1, \ldots, f_{m-1} \rangle$.
  2) If $\deg(g) = \deg(f_m)$ then: $g \in \langle f_1, \ldots, f_m \rangle$ iff $\exists a \in \mathbb{F}, (g + af_m) \in \langle f_1, \ldots, f_{m-1} \rangle$.

Proof. Say, $g \in \langle f_1, \ldots, f_m \rangle$. Then, by definition, there exist q's in R such that,

$$(8) \qquad g = \sum_{i=1}^{m} q_i f_i.$$

Let $d := \deg(g)$. If we compare the monomials of degree d on both sides of Equation (8) then the LHS gives g. In the RHS we see that an f_i of degree d_i contributes $[q_i]_{(d-d_i)} f_i$, where $[q]_j$ is defined to be the sum of the degree j terms of q (and, zero if $j < 0$). Thus, $g = \sum_{i=1}^{m} [q_i]_{d-d_i} f_i$. This equation proves both the properties at once. $\square$

We show below that a congruence of two multiplication terms modulo an ideal, generated by terms, leads to a matching via the radical-span.

Definition 43 $(L_U(\cdot), L_U^c(\cdot))$. *For a multiplication term f and a subspace $U \subseteq L(R)$ define $L_U(f) := L(f) \cap U$ and $L_U^c(f) := L(f) \setminus U$.*

Lemma 44 (Congruence to Matching)**.** *Let I be an ideal generated by multiplication terms $\{f_1, \ldots, f_m\}$ and define $U := radsp(I)$. Let g, h be multiplication terms such that $g \equiv h \not\equiv 0 \pmod{I}$. Then there is a U-matching between $L_U(g), L_U(h)$ and one between $L_U^c(g), L_U^c(h)$.*

Proof. Define $g_0 := M(L_U(g))$ and $h_0 := M(L_U(h))$. Suppose the list $L_U(g)$ is larger than the list $L_U(h)$. By the congruence we have $h \in \langle I, g_0 \rangle$. As $radsp(I, g_0) = U$, by Lemma 41 we can drop the non-U forms of h to get $h_0 \in \langle I, g_0 \rangle$. As $\langle I, g_0 \rangle$ is a homogeneous ideal and $\deg(h_0) < \deg(g_0)$ we get by Lemma 42 that $h_0 \in I$. But this means $h \in I$, which contradicts the hypothesis. Thus, $\deg(h_0) \geq \deg(g_0)$ and by symmetry we get them infact equal. Thus, the lists $L_U(g), L_U(h)$ are of equal size, which trivially U-matches them.

We will show that for any $\ell \in L(R) \setminus U$, the number of forms that are similar to ℓ mod U in $L_U^c(g)$ is equal to that in $L_U^c(h)$. This fact will prove the lemma as it shows that every form in $L_U^c(g)$ can be U-matched to a distinct form in $L_U^c(h)$.

Pick an $\ell \in L(R) \setminus U$. Let g_1 be the product of the forms that are similar to ℓ mod U in $L_U^c(g)$ (if none exist then set $g_1 = 1$), similarly define h_1 from h. Suppose $\deg(h_1) < \deg(g_1) =: d$. By the congruence we have $h \in \langle I, g_1 \rangle$. As $radsp(I, g_1) = U \oplus \mathbb{F}\ell$, by Lemma 41 we can drop the non $sp(U, \ell)$ forms of h to get

$$(9) \qquad\qquad h_0 h_1 \in \langle I, g_1 \rangle.$$

Define $r := rk(U)$ which has to be > 0, as otherwise $I = \langle 1 \rangle$ contradicting $h \notin I$. Pick an invertible linear transformation $\tau : L(R) \to L(R)$ such that forms in U are mapped inside $sp(x_1, \ldots, x_r)$ and $\ell \mapsto x_n$. Apply τ on Equation (9) to get

$$(10) \qquad\qquad h_0' h_1' = \sum_{i=1}^{m} q_i f_i' + q g_1',$$

where h_0' and f_i'-s are in $\mathbb{F}[x_1, \ldots, x_r]$; h_1' is 1 or is a polynomial with $\deg_{x_n} \in [d-1]$; g_1' is a polynomial with $\deg_{x_n} = d$; and q's $\in R$. With these conditions if we compare the coefficients of x_n^d on both sides of Equation (10) then we get $q \in \langle f_1', \ldots, f_m' \rangle$, hence $\tau^{-1}(q) \in \langle f_1, \ldots, f_m \rangle = I$. Thus, applying τ^{-1} on Equation (10) we get $h_0 h_1 \in I$, so $h \in I$, contradicting the hypothesis. Thus, $\deg(h_1) \geq \deg(g_1)$ and by symmetry we get them infact equal. This shows the number of forms that are similar to ℓ mod U in $L_U^c(g)$ is equal to that in $L_U^c(h)$, finishing the proof. $\qquad\square$

One pleasant consequence of K-matching all the multiplication terms in an identity is that we get a smaller identity, using linear forms solely from K, called the *nucleus identity*. To see that we use a metric associated with matchings, first introduced in [SS09].

Definition 45 (Scaling factor)**.** *Let K be a subspace of $L(R)$ and L_1, L_2 be two lists of linear forms in $L(R) \setminus K$. Let π be a K-matching between L_1, L_2. Then for every $\ell \in L_1$, there is a unique $c_\ell \in \mathbb{F}^*$ such that $\pi(\ell) \in c_\ell \ell + K$ (if there is another $d \in \mathbb{F}$ with $\pi(\ell) \in d\ell + K$, then $(c_\ell - d)\ell \in K$, implying $\ell \in K$, a contradiction).*
We define the scaling factor *of π, $sc(\pi) := \prod_{\ell \in L_1} c_\ell$.*

Lemma 46 (Nucleus identity)**.** *Suppose $C = \sum_i T_i$ is a $\Sigma\Pi\Sigma(k, d)$ identity and K is a subspace of $L(R)$ such that T_1, T_i are K-matched, for all $i \in [k]$. Then the terms $M(L_K(T_i))$, for $i \in [k]$, are all of the same degree, say d', and form a $\Sigma\Pi\Sigma(k, d')$ identity $\sum_{i \in [k]} \alpha_i M(L_K(T_i))$, for some $\alpha_i \in \mathbb{F}^*$.*

Proof. Since T_1, T_i are K-matched, we get from the definition of matching that terms $M(L_K(T_1)), M(L_K(T_i))$ have the same degree $d' \geq 0$. Furthermore, $M(L_K^c(T_1))$ and

$M(L_K^c(T_i))$ are also K-matched, call this induced matching π_i. As all the forms in $L_K^c(T_1)$ are outside K, the scaling factor $\mathrm{sc}(\pi_i)$ is well defined, for all $i \in [k]$.

Fix a subspace U such that $L(R) = K \oplus U$ and let $r := \mathrm{rk}(K)$. Fix an invertible linear transformation $\tau : L(R) \to L(R)$ that maps K to $\mathrm{sp}(x_1, \ldots, x_r)$. It follows that for any form $\ell \in L_K^c(T_1)$, $\tau(\ell)$ is a form with a nonzero coefficient wrt some x_i, $i > r$ (otherwise $\tau(\ell) \in \mathrm{sp}(x_1, \ldots, x_r)$, thus $\ell \in K$, a contradiction). Call the largest such i, j_ℓ. If we look at the product (note: it is over a list so it could have repeated factors),

$$(11) \qquad \alpha_1 := \prod_{\ell \in L_K^c(T_1)} [x_{j_\ell}]\tau(\ell)$$

($[\overline{x}^{\overline{i}}]f$ gives the coefficient of the monomial $\overline{x}^{\overline{i}}$ in f), then it is the coefficient of $\prod_{\ell \in L_K^c(T_1)} x_{j_\ell}$ in $\tau(M(L_K^c(T_1)))$, in other words, α_1 is its leading coefficient wrt lexicographic ordering of variables. Note that, for $i \in [k]$, π_i still $\tau(K)$-matches $\tau(L_K^c(T_1)), \tau(L_K^c(T_i))$ with the same scaling factor (if $\pi_i(\ell) \in c_\ell \ell + K$ then $\tau(\pi_i(\ell)) \in c_\ell \tau(\ell) + \tau(K)$). This means that the leading coefficient of $\tau(M(L_K^c(T_i)))$ is $\mathrm{sc}(\pi_i) \cdot \alpha_1 =: \alpha_i$, for all $i > 1$. Thus, we have pinpointed the coefficient of $\prod_{\ell \in L_K^c(T_1)} x_{j_\ell}$ in $\tau(M(L_K^c(T_i)))$ as α_i, for all $i \in [k]$. Now compare the coefficients of $\prod_{\ell \in L_K^c(T_1)} x_{j_\ell}$ in the identity $\tau(C) = 0$. This gives $\sum_{i \in [k]} \alpha_i \cdot \tau(M(L_K(T_i)))$ $= 0$. Applying the inverse of τ, we get the nucleus identity. $\square$

In Lemma 41 we have already come across a cancellation rule for non-zerodivisors. Here we see a situation in which it is stronger.

Lemma 47 (Cancellation). *Let K be some subspace of $L(R)$ and let $\ell_1, \ldots, \ell_m \in L(R) \setminus K$ be linearly independent modulo K. Let $f_1, \ldots, f_m$ be multiplication terms similar to powers of $\ell_1, \ldots, \ell_m$ respectively modulo K (i.e. each form in f_i is in $(\mathbb{F}^*\ell_i + K)$). Let $\ell \in L(R)^*$ such that for some $s \in [m]$, $\ell \in \mathbb{F}\ell_s + K$. Then, for any polynomial $f \in R$,*

$$\ell f \in \langle f_1, \ldots, f_m \rangle \text{ iff } f \in \left\langle f_1, \ldots, \frac{f_s}{\gcd(f_s, \ell)}, \ldots, f_m \right\rangle.$$

Proof. Suppose $\ell f \in \langle f_1, \ldots, f_m \rangle$. Then, by definition, there exist q's in R such that,

$$(12) \qquad \ell f = \sum_{i=1}^m q_i f_i.$$

Additionally assume these q_i-s to be such that the set $J := \{j \in [m] \setminus \{s\} \mid \ell \nmid q_j\}$ is the smallest possible. If $\ell | q_i$, for all $i \in [m] \setminus \{s\}$, then ℓ has to divide $q_s f_s$. This means that ℓ has to divide $q_s \gcd(\ell, f_s)$, thus we get,

$$f = \sum_{i \in [m] \setminus \{s\}} \frac{q_i}{\ell} f_i \; + \; \frac{q_s \gcd(\ell, f_s)}{\ell} \cdot \frac{f_s}{\gcd(f_s, \ell)}$$

and we are done.

So the remaining case is when the set $J := \{j \in [m] \setminus \{s\} \mid \ell \nmid q_j\}$ is nonempty. Fix an element $j^* \in J$. Consider ideal $I := \langle \{\ell, f_s\} \cup \{f_j | j^* \neq j \in J\} \rangle$. Reducing Equation (12) modulo I we get, $q_{j^*} f_{j^*} \equiv 0 \pmod{I}$. Note that $\mathrm{radsp}(I) \subseteq K + \mathrm{sp}(\{\ell_j | j^* \neq j \in [m]\})$ while each form in $L(f_{j^*})$ is in $(\mathbb{F}^*\ell_{j^*} + K)$ disjoint from $\mathrm{radsp}(I)$, thus by Lemma 41 we can drop f_{j^*} from the last congruence and get $q_{j^*} \in I$. This means $q_{j^*} f_{j^*} \in \langle \{\ell f_{j^*}, f_s\} \cup \{f_j | j^* \neq j \in J\} \rangle$. We plug this in the j^*-th summand of Equation (12) and

after simplifications get (verify that the $[m] \setminus (\{s\} \cup J)$ summands are unaffected):

$$\ell f = \sum_{i=1}^{m} q_i f_i$$

$$= q'_s f_s + (\ell q'_{j^*}) f_{j^*} + \sum_{j \in J \setminus \{j^*\}} q'_j f_j + \sum_{j \in [m] \setminus (\{s\} \cup J)} q_j f_j$$

Notice that for $j \in [m] \setminus (\{s\} \cup J)$, ℓ divides q_j, thus the above equation contradicts the assumed minimality of J. This shows that J was empty to begin with, thus finishing the proof. $\qquad\square$