Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 42 (2007)

Black Box Polynomial Identity Testing of Generalized Depth-3 Arithmetic Circuits with Bounded Top Fan-in

Zohar S. Karnin* Amir Shpilka*

Abstract

In this paper we consider the problem of determining whether an unknown arithmetic circuit, for which we have oracle access, computes the identically zero polynomial. This problem is known as the black-box polynomial identity testing (PIT) problem. Our focus is on polynomials that can be written in the form $f(\bar{x}) = \sum_{i=1}^{k} h_i(\bar{x}) \cdot g_i(\bar{x})$, where each h_i is a polynomial that depends on only ρ linear functions, and each g_i is a product of linear functions (when $h_i = 1$, for each i, then we get the class of depth-3 circuits with k multiplication gates, also known as $\Sigma\Pi\Sigma(k)$ circuits, but the general case is much richer). When $\max_i(\deg(h_i \cdot g_i)) = d$ we say that f is computable by a $\Sigma\Pi\Sigma(k, d, \rho)$ circuit. We obtain the following results.

1. A deterministic black-box identity testing algorithm for $\Sigma\Pi\Sigma(k, d, \rho)$ circuits that runs in quasi-polynomial time (for $\rho = \text{polylog}(n + d)$). In particular this gives the first black-box quasi-polynomial time PIT algorithm for depth-3 circuits with k multiplication gates.
2. A deterministic black-box identity testing algorithm for read-k $\Sigma\Pi\Sigma$ circuits (depth-3 circuits where each variable appears at most k times) that runs in time $n^{2^{O(k^2)}}$. In particular this gives a polynomial time algorithm for $k = O(1)$.

Our results give the first sub-exponential black-box PIT algorithm for circuits of depth higher than 2. Another way of stating our results is in terms of *test sets* for the underlying circuit model. A test set is a set of points such that if two circuits get the same values on every point of the set then they compute the same polynomial. Thus, our first result gives an explicit test set, of quasi-polynomial size, for $\Sigma\Pi\Sigma(k, d, \rho)$ circuits (when $\rho = \text{polylog}(n + d)$). Our second result gives an explicit polynomial size test set for read-k depth-3 circuits.

The proof technique involves a construction of a family of affine subspaces that have a *rank-preserving* property that is inspired by the construction of *linear seeded extractors for affine sources* of Gabizon and Raz [GR05], and a generalization of a theorem of [DS06] regarding the structure of identically zero depth-3 circuits with bounded top fan-in.

[*] Faculty of Computer Science, Technion, Haifa 32000, Israel. Email: {zkarnin,shpilka}@cs.technion.ac.il. Research supported by the Israel Science Foundation (grant number 439/06).

ISSN 1433-8092

Contents

1 Introduction

Finding an algorithm for polynomial identity testing (PIT) is a widely pursued open problem: We are given as input a circuit that computes a multivariate polynomial, over some field, and we have to determine whether it computes the zero polynomial. The importance of the polynomial identity testing problem stems from its many applications: Algorithms for primality testing [AB03], for deciding if a graph contains a perfect matching [Lov79, MVV87, CRS95] and more, are based on reductions to the PIT problem (for more applications see the introduction of [LV98]). In this work we consider the problem of determining whether an arithmetic circuit for which we only have oracle access computes the identically zero polynomial. That is, the input is a black-box holding a circuit C and we must find whether the polynomial computed by the circuit C is the identically zero polynomial. In particular we can only ask the circuit for its value on points of our choice. It is clear that every such algorithm must produce a test set for the circuit. Namely, a set of points such that if the circuit vanishes on all the points then the circuit computes the zero polynomial. Note that the values of a circuit on the points in the test set completely determine the circuit[1], as if two circuits agree on all the points then their difference is zero on all points of the set and therefore their difference must be zero.

1.1 Known results

The complexity of the PIT problem is not well understood. It is one of a few problems for which we have coRP algorithms but no deterministic sub-exponential time algorithms. The first randomized black-box PIT algorithm was discovered independently by Schwartz [Sch80] and Zippel [Zip79]. In [LV98, AB03, CK00] randomized algorithms that use fewer random bits were given, however these algorithms need to get the circuit as input, whereas the Schwartz-Zippel algorithm is in the black-box model. The problem of finding an efficient deterministic algorithm, or proving that no such algorithm exists, is believed to be difficult. In particular, Kabanets and Impagliazzo [KI04] and Agrawal [Agr05] showed that efficient deterministic algorithms for PIT imply lower bounds for arithmetic circuits. Conversely, [KI04] showed that from super-polynomial lower bounds on the size of arithmetic circuits one can construct a sub-exponential time deterministic algorithm for black-box PIT. However, known lower bounds are too weak and do not yield deterministic sub-exponential time PIT algorithms as suggested by [KI04].

Nevertheless, deterministic polynomial time algorithms for several restricted classes are known: For depth-2 arithmetic circuits (i.e. circuits computing sparse multivariate polynomials) there are many works giving black-box PIT algorithms over various fields [GK87, BOT88, GKS90, CDGK91, Wer94, SS96, KS96, KS01], for non-commutative arithmetic formulas there is a non black-box algorithm [RS05] and for the class of read-once arithmetic formulas, sub-exponential time black box algorithms were recently given in [SV08].

The question of giving efficient black-box polynomial identity testing algorithm for $\Sigma\Pi\Sigma(3)$ circuits (depth-3 circuits with only 3 multiplication gates) was raised by Klivans and Spielman [KS01]. In the non black-box model this question was first solved in [DS06]. Their algorithm gets as an input a depth-3 arithmetic circuit with bounded top fan in, and determines whether the circuit computes the zero polynomial or not. The crux of that work is a theorem on the structure of depth-3 arithmetic circuits that compute the zero polynomial. Specifically, for every depth-3 arithmetic circuit with bounded top fan in, if the circuit is simple (i.e. no linear function appears in all of the multiplication gates) and minimal (i.e. no subset of the multiplication gates amounts

[1] However, it is a very interesting question (and very difficult) to reconstruct the circuit from its values on the test set.

to a circuit computing the zero polynomial), then the dimension of the linear space spanned by all the linear functions in the circuit is small. The algorithm of [DS06] runs in quasi-polynomial time. This result was later improved by Kayal and Saxena [KS06] who gave a polynomial time algorithm (in the non black-box model) using a different approach. Recently a similar result with a different proof was given by Arvind and Mukhopadhyay [AM07]. For our results however, we shall need the structural theorem of [DS06].

In this work we give a sub-exponential deterministic black-box algorithm for PIT of generalized depth-3 circuits with bounded top fan-in. More precisely, the running time of our algorithm is similar to the running time of the non black-box algorithm of [DS06]. This is the first sub-exponential PIT algorithm in the black-box model for a class of circuits other than the widely studied class of depth-2 circuits (the recent result of [SV08] also gives a sub-exponential black-box PIT algorithm). In particular, our result answers the black-box version of the question of Klivans and Spielman [KS01]. Before giving a formal statement of our results we need some definitions.

1.2 Some definitions and statement of our results

In this work we study a generalization of depth-3 circuits that we denote by $\Sigma\Pi\Sigma(k,d,\rho)$ circuits. A polynomial $f(\bar{x})$ that is computed by a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit has the following form

$$f(\bar{x}) = \sum_{i=1}^{k} M_i = \sum_{i=1}^{k} \left(\prod_{j=1}^{d_i} L_{i,j}(\bar{x}) \right) \cdot h_i \left(\tilde{L}_{i,1}(\bar{x}), \ldots, \tilde{L}_{i,\rho_i}(\bar{x}) \right) \qquad (1)$$

where the $L_{i,j}$'s and the $\tilde{L}_{i,j}$'s are linear functions in the variables $\bar{x} = (x_1, \ldots, x_n)$, over $\mathbb{F}$. Every h_i is a polynomial in $\rho_i \leq \rho$ variables, and the functions $\{\tilde{L}_{i,j}\}_{j=1}^{\rho_i}$ are linearly independent. We shall assume, w.l.o.g., that each h_i depends on all its ρ_i variables. We call $M_1, \ldots, M_k$ the multiplication gates of the circuit ($M_i = \prod_{j=1}^{d_i} L_{i,j} \cdot h_i(\tilde{L}_{i,1}(\bar{x}), \ldots, \tilde{L}_{i,\rho_i}(\bar{x}))$). For a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit C we denote with $d = \deg(C)$ the maximal degree of its multiplication gates (i.e. $\max_{i=1\ldots k}\{\deg(M_i)\}$). When $\rho = 0$ (i.e. each h_i is a constant function) we get the class of depth-3 circuits with k multiplication gates and degree d, also known as $\Sigma\Pi\Sigma(k,d)$ circuits. When k and d are arbitrary we get the class of depth-3 circuits that we denote with $\Sigma\Pi\Sigma$. The following theorems summarize our results for $\Sigma\Pi\Sigma(k,d,\rho)$ arithmetic circuits:

Theorem 1 (Deterministic algorithm for $\Sigma\Pi\Sigma(k,d,\rho)$ circuits)**.** *Let k,d,ρ,n be integers and $\mathbb{F}$ a field. Then there is a deterministic black-box algorithm that on input k,d,ρ,n and black-box access to a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit C in n indeterminates over $\mathbb{F}$, determines whether C computes the zero polynomial. The running time of the algorithm is $\text{poly}(n) \cdot \exp\left((\log d)^{k-1} + k\rho \log d\right)$. If*[2] *$|\mathbb{F}| \leq O(d^2 \cdot n \cdot (k\rho + (\log d)^{k-2}))$ then the algorithm is allowed to make queries to C from an algebraic extension field of $\mathbb{F}$.*

In particular this gives a quasi-polynomial black-box PIT algorithm for $\Sigma\Pi\Sigma(k,d)$ circuits for a constant k. Our second result is for read-k depth-3 circuits. A read-k depth-3 circuit is a depth-3 circuit in which every variable appears at most k times (that is, every variable belongs to at most k linear functions). We obtain the following results for read-k $\Sigma\Pi\Sigma$ circuits.

Theorem 2 (Deterministic algorithm for read-k circuits)**.** *Let k,n be integers and $\mathbb{F}$ a field. Then there is a deterministic black-box algorithm that on input k,n and black-box access to a read-k depth-3 circuit C in n indeterminates over $\mathbb{F}$, runs in time $n^{2^{O(k^2)}}$ and determines whether C computes*

[2] When k is a constant and $\rho = o(\deg(C))$ all we need is a field of size larger than, say, $n \cdot \deg(C)^3$.

the zero polynomial. If $|\mathbb{F}| \leq O(n^3 \cdot k^4)$, then the algorithm is allowed to make queries to C from an algebraic extension field of $\mathbb{F}$.

In particular this result gives a polynomial time black box PIT algorithm for multilinear depth-3 circuits with a constant number of multiplication gates. As corollaries of the above constructions we get the following results.

Theorem 3 (Randomized algorithm for general circuits)**.** *Let C be a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit over a field $\mathbb{F}$, in n indeterminates, for some k,d,ρ,n. Then there is a coRP randomized black-box algorithm that on input ϵ,k,d,ρ,n makes a single query to (a black-box holding) C and determines whether $C \equiv 0$. Namely, if $C \not\equiv 0$ then the algorithm outputs "non-zero circuit" with probability at least $1-\epsilon$ and if $C \equiv 0$ then the algorithm always outputs "zero circuit". The number of random bits used by the algorithm is $O\left((\log(d)+\log(1/\epsilon)) \cdot \left(\log(d)^{k-2}+k\rho\right)+\log(n)\right)$. As in Theorem 1, if $|\mathbb{F}| \leq O(\deg(C)^2 \cdot n \cdot (k\rho + (\log d)^{k-2}))$ then the algorithm is allowed to make queries to C from an algebraic extension field of $\mathbb{F}$.*

Another corollary is a generalization of Theorem 1 for the case where each variable appears in at most k multiplication gates.

Theorem 4. *Let C be a $\Sigma\Pi\Sigma(m,d,\rho)$ circuit over a field $\mathbb{F}$, in n indeterminates, where each input variable appears in at most k multiplication gates, for some integers m,k,d,ρ,n. That is, there are m multiplication gates but each variable belongs to at most k of them. Then there is a deterministic black-box algorithm that on input m,k,d,ρ,n and black-box access to C determines whether C computes the zero polynomial. The running time of the algorithm is* $\mathrm{poly}(n)\cdot\exp\left((\log d)^{2k-1}+k\rho\log d\right)$. *If $|\mathbb{F}| \leq O(\deg(C)^2 \cdot n \cdot (k\rho + (\log d)^{k-2}))$ then the algorithm is allowed to make queries to C from an algebraic extension field of $\mathbb{F}$ of polynomial size.*

Note that the circuit considered in the theorem is stronger than read-k $\Sigma\Pi\Sigma(m,d,r)$ circuits, as every input variable can appear in each multiplication gate many times (we just bound the number of gates in which the variable appears).

1.3 Our techniques

The idea behind our algorithms is the following: we consider several linear subspaces of $\mathbb{F}^n$ of "low" dimension, and for each subspace V we verify that $C|_V \equiv 0$. Note, that the verification step requires $O(\deg(C)^{\dim(V)})$ time using a simple brute force interpolation. Clearly if $C \equiv 0$ then we will get that $C|_V \equiv 0$. However, it is not clear why $C \equiv 0$ if all we know is that $C|_V \equiv 0$, for every subspace V in our family. Indeed, for general depth-3 circuits we cannot show that such a naive approach works, but in the case of $\Sigma\Pi\Sigma(k,d,\rho)$ circuits we have a structural theorem[3] due to [DS06] that (roughly) says that if $C \equiv 0$ then it can be written as a sum of circuits, that are all identically zero, and such that each of the circuits essentially depends on a few linear functions (the complete statement of this theorem is given in Section 2.1, and our strengthening is given in Lemma 4.2). Thus, the structural theorem implies that for every subspace V, if $C|_V \equiv 0$ then it has the above structure. If we were guaranteed that for some V the "structure" of C remains (more or less) the same when we restrict it to V, then the fact that $C|_V \equiv 0$ will imply that $C \equiv 0$. Indeed, our family of subspaces has the guarantee that for every $\Sigma\Pi\Sigma(k,d,\rho)$ circuit C there will be at least one subspace in the family (in fact most subspaces in the family will have the property) for which the "structure" of C does not change much when restricted to V.

[3] Actually the theorem of [DS06] speaks about $\Sigma\Pi\Sigma(k,d)$ circuits, but we prove a similar result for $\Sigma\Pi\Sigma(k,d,\rho)$ circuits.

The idea behind the construction of the family of subspace on which we will evaluate the restriction of C comes from the construction of *linear seeded extractors for affine sources* of [GR05]. In their work Gabizon and Raz constructed a set of linear transformations from $\mathbb{F}^n$ to $\mathbb{F}^r$ such that for every linear subspace of dimension r, at least one of the transformations (actually most of the transformations) maps it onto the entire space. It turns out that by applying the idea of [GR05] we can construct a family of subspaces that retains the structure of $\Sigma\Pi\Sigma(k,d,\rho)$ circuits, and therefore get a deterministic black-box PIT algorithm.

1.4 Organization

The paper is organized as follows. In section 2 we give some background on depth-3 arithmetic circuits. In section 3 we provide the main idea behind our algorithms (Theorem 3.4). Section 4 contains the proofs of Theorem 1, 3 and 4. Finally, in section 5 we prove Theorem 2.

2 Preliminaries

For a positive integer k we denote $[k] = \{1,\ldots,k\}$. Let $\mathbb{F}$ be a field. We denote with $\mathbb{F}^n$ the n'th dimensional vector space over $\mathbb{F}$. For a vector $v \in \mathbb{F}^n$ we denote with $|v|$ the number of non zero entries of v. We denote with $\{e_i\}_{i\in[n]}$, the natural basis for $\mathbb{F}^n$. That is, e_i is an n-dimensional vector that has 1 in the i-th coordinate and zeros elsewhere. We shall use the notation $\bar{x} = (x_1,\ldots,x_n)$ to denote the vector of n indeterminates. For a linear functions L we denote its homogenous part with L^H (i.e., for $L = a_0 + \sum_{i=1}^n a_i x_i$ we define $L^H = \sum_{i=1}^n a_i x_i$). For two linear functions L_1, L_2 we write $L_1 \sim L_2$ whenever L_1 and L_2 are linearly dependent. The same notation will be used for vectors. Let $V = V_0 + v_0 \subseteq \mathbb{F}^n$ be an affine subspace, where $v_0 \in \mathbb{F}^n$ and $V_0 \subseteq \mathbb{F}^n$ is a linear subspace. Let $L(\bar{x})$ be a linear function. We denote with $L|_V$ the restriction of L to V. Assume that the dimension of V_0 is t, then $L|_V$ can be viewed as a linear function of t indeterminates in the following way: Let $\{v_i\}_{i\in[t]}$ be a basis for V_0. For $v \in V$ let $v = \sum_{i=1}^t y_i \cdot v_i + v_0$ be its representation according to the basis. We get that

$$L(v) = \sum_{i=1}^{t} y_i \cdot L(v_i) + L(v_0) \overset{\triangle}{=} L|_V(y_1,\ldots,y_t).$$

We shall abuse notation and use both $L|_V(v)$ and $L|_V(y_1,\ldots,y_t)$ to denote the value of L on $v \in V$. Note that the representation of $L|_V(\bar{y})$ depends on the chosen basis for V, but the value of $L|_V(v)$ does not[4]. A linear function L will sometimes be viewed as a vector of $n+1$ entries. Namely, the function $L(x_1,\ldots,x_n) = \sum_{i=1}^n \alpha_i \cdot x_i + \alpha_0$ corresponds to the vector of coefficients $(\alpha_0, \alpha_1, \ldots, \alpha_n)$. Accordingly, we define the span of a set of linear functions of n variables as the span of the corresponding vectors (i.e. as a subspace of $\mathbb{F}^{n+1}$). For an affine subspace $V = V_0 + v_0$ of dimension t, the linear function $L|_V$ can be viewed as a vector of $t+1$ entries. Thus, V, equipped with a basis $\{v_i\}_{i\in[t]}$ for V_0, defines a linear transformation from $\mathbb{F}^{n+1}$ to $\mathbb{F}^{t+1}$, that sends $L(\bar{x})$ to $L|_V(\bar{y})$. We shall sometimes refer to this transformation as the linear transformation corresponding to the affine subspace V, and denote it with T_V.

[4] In order for $L|_V(y_1,\ldots,y_t)$ to be well defined, it must correspond to some "default" basis of V_0. When not stated otherwise, we choose the gaussian elimination of some basis of V as its default basis

2.1 Generalized Depth 3 Arithmetic Circuits

We first recall the usual definition of depth-3 circuits. A depth-3 circuit with k multiplication gates of degree d (also known as $\Sigma\Pi\Sigma(k,d)$ circuit) has the following form:

$$C = \sum_{i=1}^{k} M_i = \sum_{i=1}^{k} \prod_{j=1}^{d_i} L_{i,j}(x_1,\ldots,x_n) \tag{2}$$

where each $L_{i,j}$ is a linear function in the input variables and $d = \max_{i=1\ldots k}\{\deg(M_i)\}$. When k and d are unimportant or unknown we just refer to the circuit as a $\Sigma\Pi\Sigma$ circuit. Recall that we defined a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit (see Equation 1) to be a circuit of the form

$$C = \sum_{i=1}^{k} M_i = \sum_{i=1}^{k} \left(\prod_{j=1}^{d_i} L_{i,j}(\bar{x})\right) \cdot h_i\left(\tilde{L}_{i,1}(\bar{x}),\ldots,\tilde{L}_{i,\rho_i}(\bar{x})\right). \tag{3}$$

We thus see that in a generalized depth-3 circuit multiplication gates can have an additional term that is a polynomial that depends on (at most) ρ linear functions. The following notions will be used throughout this paper.

Definition 2.1. *Let C be a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit that computes a polynomial as in Equation (3).*

1. *For every multiplication gate M_i we define $\mathrm{Lin}(M_i) = \prod_{j=1}^{d_i} L_{i,j}(\bar{x})$. That is, $\mathrm{Lin}(M_i)$ is the product of all the linear factors of M_i (we can assume w.l.o.g. that h_i, the non-linear term of M_i, has no linear factors). In particular, for a $\Sigma\Pi\Sigma$ circuit, $\mathrm{Lin}(M_i) = M_i$.*

2. *The derived $\Sigma\Pi\Sigma(k,d)$ circuit is defined as $\widehat{C} \stackrel{\triangle}{=} \sum_{i=1}^{k} \mathrm{Lin}(M_i)$. This definition is interesting only when C is a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit (as if $\rho = 0$ then $\widehat{C} = C$).*

3. *For each $A \subseteq [k]$, we define $C_A(\bar{x})$ to be a sub-circuit of C as follows: $C_A(\bar{x}) = \sum_{i\in A} M_i(\bar{x})$.*

4. *Define $\gcd(C)$ as the product of all the non-constant linear functions that divide all the multiplication gates. In other words, $\gcd(C) = \text{g.c.d.}(\mathrm{Lin}(M_1),\ldots,\mathrm{Lin}(M_k))$. A circuit will be called simple if $\gcd(C) = 1$.*

5. *The simplification of C, $\mathrm{sim}(C)$, is defined as $\mathrm{sim}(C) \stackrel{\triangle}{=} C/\gcd(C)$. Notice that $\mathrm{sim}(C)$ is a $\Sigma\Pi\Sigma(k,d',\rho)$ circuit for $d' = d - \deg(\gcd(C))$.*

6. *We define $\mathrm{Lin}(C) \stackrel{\triangle}{=} \{L_{i,j}^H\}_{i\in[k],j\in[d_i]} \cup \left(\bigcup_{i=1}^{k} \mathrm{span}\left\{(\tilde{L}_{i,j})^H\right\}_{j\in[\rho_i]}\right)$. Notice that we take every linear function in the span of each $\{\tilde{L}_{i,j}^H\}_{j\in[\rho_i]}$ to be in $\mathrm{Lin}(C)$.*

7. *We define $\mathrm{rank}(C)$ as the dimension of the span of the homogenous part of the linear functions in C. That is, $\mathrm{rank}(C) = \dim(\mathrm{Lin}(C))$.*

A word of clarification is needed regarding the definition of $\mathrm{Lin}(C)$ and $\mathrm{rank}(C)$. Notice that the definition seems to depend on the specific choice of linear functions $\tilde{L}_{i,j}$. That is, it may be the case (and it is indeed the case) that every polynomial $h_i(\tilde{L}_{i,1},\ldots,\tilde{L}_{i,\rho_i})$ can be represented as a (different) polynomial in some other set of linear functions. However the following lemma from [Shp07] shows that the specific representation that we chose does not change the rank nor the set $\mathrm{Lin}(C)$.

Lemma 2.2 (Lemma 20 in [Shp07]). *Let $h(\bar{x})$ be a polynomial in exactly k linear functions*[5]. *Let $P(\ell'_1,\ldots,\ell'_k) = h = Q(\ell_1,\ldots,\ell_k)$ be two different representations for h. Then* $\mathrm{span}(\{(\ell'_i)^H\}_{i\in[k]}) = \mathrm{span}(\{(\ell_i)^H\}_{i\in[k]})$.

We shall use the notation $C \equiv 0$ to denote the fact that a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit computes the identically zero polynomial. Notice that this is a syntactic definition, we are thinking of the circuit as computing a polynomial and not a function over the field. We say that a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit C is minimal if there is no $\emptyset \neq A \subsetneq [k]$ such that $C_A \equiv 0$. The following theorem of [DS06] gives a bound on the rank of $\Sigma\Pi\Sigma(k,d)$ identically zero circuits (the case that $\rho = 0$).

Theorem 2.3 (Lemma 5.2 of [DS06]). *Let $k \geq 3$ and $C \equiv 0$ be a simple and minimal $\Sigma\Pi\Sigma(k,d)$ circuit, of degree $d \geq 2$. Then* $\mathrm{rank}(C) < 2^{O(k^2)}\log^{k-2}(d)$.

For convenience, we define $R(k,d) = 2^{O(k^2)}\log^{k-2}(d)$ as the bound on the rank given by Theorem 2.3. It follows that $R(k,d)$ is larger than the rank of any identically zero simple and minimal $\Sigma\Pi\Sigma(k,d)$ circuit.

3 Rank Preserving Subspaces

As mentioned in Section 1.3, we would like to find a family of subspaces that for each possible circuit contains at least one subspace that preserves, to some extent, the "structure" of the circuit. In this section we state a list of properties for a subspace V and circuit C such that when held, $C|_V \equiv 0$ implies that $C \equiv 0$. Later we shall see how to construct a family of subspaces having the required properties (the construction is slightly different for the case that C is a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit and for the case that C is a read-k depth-3 circuit). We now define r-rank-preserving subspaces. Notice that this definition does not rely on the family of circuits that we work with. In particular when we speak of depth-3 circuits we shall mean $\Sigma\Pi\Sigma(k,d,\rho)$ circuits or ordinary $\Sigma\Pi\Sigma$ circuits.

Definition 3.1. *Let C be a depth-3 circuit and V an affine subspace. We say that V is r-rank-preserving for C if the following properties hold:*

1. *Any two linearly independent linear functions that appear in $\widehat{C}$, that neither of them was restricted to a constant function on V, remain linearly independent when restricted to V.*
2. $\forall A \subseteq [k],\ \mathrm{rank}(\mathrm{sim}(C_A)|_V) \geq \min\{\mathrm{rank}(\mathrm{sim}(C_A)), r\}$.
3. *No multiplication gate $M \in C$ vanishes on V. In other words $M|_V \not\equiv 0$ for every multiplication gate $M \in C$.*
4. $\mathrm{Lin}(M)|_V = \mathrm{Lin}(M|_V)$ *for every multiplication gate M in C (that is, the polynomial computed by* $\mathrm{sim}(M)|_V$ *has no linear factors).*

The following lemma lists some useful properties of rank-preserving subspaces.

Lemma 3.2. *Let C be a depth-3 circuit and V an r-rank-preserving affine subspace for C. Then we have the following:*

1. *For every $\emptyset \neq A \subseteq [k]$, V is r-rank-preserving for C_A.*
2. *V is r-rank-preserving for* $\mathrm{sim}(C)$.

[5] That is, h can be written as a polynomial in k linear functions but not in $k-1$ linear functions.

3. $\gcd(C)|_V = \gcd(C|_V)$.

4. $\mathrm{sim}(C)|_V = \mathrm{sim}(C|_V)$.

Proof. The first and second claims follow immediately from the definition of V. To prove the third claim we note that as $\mathrm{Lin}(M)|_V = \mathrm{Lin}(M|_V)$ for every multiplication gate, we have that $\gcd(C|_V) = \text{g.c.d.}\{\mathrm{Lin}(M|_V)\}_{M \in C} = \text{g.c.d.}\{\mathrm{Lin}(M)|_V\}_{M \in C}$. Since $\mathrm{Lin}(M)$ is a product of linear functions from C we get that $\text{g.c.d.}\{\mathrm{Lin}(M)|_V\}_{M \in C} = (\text{g.c.d.}\{\mathrm{Lin}(M)\}_{M \in C})\,|_V = \gcd(C)|_V$. Where the first equality holds as no new non-constant linear functions from the $\mathrm{Lin}(M_i)$'s were added to the g.c.d. after the restriction (as otherwise there will be two linearly independent linear functions in $\widehat{C}$ that become non-constant and dependent when restricted to V, in contradiction to Property 1 of Definition 3.1). The second equality is simply the definition of $\gcd(C)|_V$. The fourth claim is a direct consequence of the third claim and the definition of $\mathrm{sim}(C)$. We note that the proof of the third and forth claims did not use Property 2 of Definition 3.1. □

We are now ready for the main theorem of this section. In order to state it in the most general form we shall speak of a family of circuits having some closure properties. In this way we will not have to state different results for different families of circuits. The following definition states the required closure properties we want a family of depth-3 circuits to have.

Definition 3.3 (Closure property)**.** *Let $V \subseteq \mathbb{F}^n$ be a linear subspace. A family $\mathcal{F}$ of depth-3 circuits in n indeterminates is* closed with respect to V *if whenever C is a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit in the family we have that*

- *$C|_V \in \mathcal{F}$.*
- *$C_A \in \mathcal{F}$, for every $A \subseteq [k]$.*
- *$\mathrm{sim}(C) \in \mathcal{F}$.*

We now give the statement of the theorem. In order to better understand it one can have in mind the family of $\Sigma\Pi\Sigma(k,d)$ circuits and $R_{\mathcal{F}} = R(k,d)$ as defined after Theorem 2.3.

Theorem 3.4. *Let $\mathcal{F}$ be a family of depth-3 circuits. Assume that there exists $R_{\mathcal{F}} \in \mathbb{N}$ such that for every $C \in \mathcal{F}$ that is simple, minimal and computes the zero polynomial* $\mathrm{rank}(C) < R_{\mathcal{F}}$*. Let $V \subseteq \mathbb{F}^n$ be a subspace, such that $\mathcal{F}$ is closed with respect to V. Let C be a circuit in $\mathcal{F}$ and assume further that V is an $R_{\mathcal{F}}$-rank-preserving subspace for C. Then, if $C|_V \equiv 0$ then $C \equiv 0$.*

Proof. Let k be the number of multiplication gates in C. The proof is in three steps. We first prove the theorem for the case that $C|_V$ (which is identically zero) is simple and minimal. We then remove the simplicity assumption, and finally we remove the minimality assumption.

Assume that $C|_V$ is identically zero simple and minimal. As $C_V \in \mathcal{F}$ we get, by the assumption on $R_{\mathcal{F}}$, that $\mathrm{rank}(C|_V) < R_{\mathcal{F}}$. From the fact that V is $R_{\mathcal{F}}$-rank-preserving for C and from Property 2 of Definition 3.1 (applied to $A = [k]$) we get that $\mathrm{rank}(C|_V) \geq \mathrm{rank}(C)$, and thus $\mathrm{rank}(C|_V) = \mathrm{rank}(C)$. Denote by r the rank of the circuit C. Let $L_1, \ldots, L_r$ be linear functions forming a basis of $\mathrm{Lin}(C)$. It follows that there exists a polynomial P such that $C \equiv P(L_1, \ldots, L_r)$. Obviously, $C|_V \equiv P(L_1|_V, \ldots, L_r|_V) \equiv 0$.

We now prove that the linear functions $(L_1|_V)^H, \ldots, (L_r|_V)^H$ span $\mathrm{Lin}(C|_V)$. Let L be a linear function appearing in C. Then $L = a_0 + \sum_{i=1}^{r} a_i L_i$, and $L|_V = a_0 + \sum_{i=1}^{r} a_i L_i|_V$. Hence, $(L|_V)^H = \sum_{i=1}^{r} a_i (L_i|_V)^H$. Since $\mathrm{rank}(C|_V) = \mathrm{rank}(C) = r$, we have that $(L_1|_V)^H, \ldots, (L_r|_V)^H$ are linearly independent. Hence, P is the zero polynomial and $C \equiv P(L_1, \ldots, L_r) \equiv 0$. This completes the proof for the case that $C|_V$ is simple and minimal. We now remove the simplicity assumption. Assume that $C|_V$ is an identically zero minimal circuit. In a nutshell, the proof for this case has the following form:

$$C|_V \equiv 0 \overset{(1)}{\Rightarrow} \mathrm{sim}(C|_V) \equiv 0 \overset{(2)}{\Rightarrow} \mathrm{sim}(C)|_V \equiv 0 \overset{(3)}{\Rightarrow} \mathrm{sim}(C) \equiv 0 \overset{(4)}{\Rightarrow} C \equiv 0. \tag{4}$$

We now explain each of the implications in Equation (4).

- Implication (1) follows from property 3 of Definition 3.1 and Lemma 3.2 (as the lemma implies that $\gcd(C)|_V \neq 0$).

- The second implication follows immediately from Lemma 3.2.

- To prove implication (3) we recall that by the closure property of $\mathcal{F}$ we have that $\mathrm{sim}(C) \in \mathcal{F}$, hence $\mathrm{sim}(C)|_V \in \mathcal{F}$. Therefore, $\mathrm{sim}(C)|_V$ is a simple (as $\mathrm{sim}(C)|_V = \mathrm{sim}(C|_V)$) and minimal (by assumption) identically zero circuit in $\mathcal{F}$. As V is also $R_{\mathcal{F}}$-rank-preserving for $\mathrm{sim}(C)$ we get (by the case of simple and minimal $C|_V$) that $\mathrm{sim}(C) \equiv 0$.

- Step (4) follows immediately from the definition of $\mathrm{sim}(C)$.

We now prove the general case, that is we just assume that $C|_V \equiv 0$. Clearly there exists a partition $A_1, \ldots, A_s$ of $[k]$ (That is, the A_i's are disjoint subsets of [k] whose union is [k]) such that for every $i \in [s]$ we have that $C_{A_i}|_V$ is an identically zero minimal depth-3 circuit. Recall that Definition 3.1 implies that V is also $R_{\mathcal{F}}$-rank-preserving for each C_{A_i}. Furthermore, since $\mathcal{F}$ is closed w.r.t. V, for each $i \in [s]$, both $C_{A_i}|_V$ and C_{A_i} belong to $\mathcal{F}$. Hence, by what we just showed for minimal circuits, we get that $C_{A_i} \equiv 0$. It follows that $C = \sum_{i=1}^{s} C_{A_i} \equiv 0$. This completes the proof of the theorem. □

4 Black-box PIT for $\Sigma\Pi\Sigma(k, d, \rho)$ circuits

In this section we prove Theorem 1. The proof relies on Theorem 3.4. Therefore, in order to use the theorem we have to understand what is $R_{\mathcal{F}}$ for the family of $\Sigma\Pi\Sigma(k, d, \rho)$ circuits, and prove closure properties for this family. As a first step we notice that for every subspace V, the family of $\Sigma\Pi\Sigma(k, d, \rho)$ is closed with respect to V. the proof is immediate from the definition of the circuits.

Lemma 4.1. *The family of n variate $\Sigma\Pi\Sigma(k, d, \rho)$ circuits is closed w.r.t. any subspace $V \subseteq \mathbb{F}^n$.*

Next we give a bound on $R_{\mathcal{F}}$ where $\mathcal{F}$ is the family of $\Sigma\Pi\Sigma(k, d, \rho)$ circuits (for some k, d, ρ). That is, we give an upper bound, which we denote by $R(k, d, \rho)$, on the rank of a simple and minimal $\Sigma\Pi\Sigma(k, d, \rho)$ circuit computing the zero polynomial. Our bound is related to $R(k, d)$ (whose definition is given after Theorem 2.3).

Lemma 4.2. *Let C be a simple and minimal $\Sigma\Pi\Sigma(k, d, \rho)$ circuit in n indeterminates computing the zero polynomial. Then* $\mathrm{rank}(C) < R(k, d, \rho) \overset{\Delta}{=} R(k, d) + k \cdot \rho$.

Proof. For convenience we shall use the notations of Equation (3). That is, we denote

$$C = \sum_{i=1}^{k} M_i = \sum_{i=1}^{k} \left(\prod_{j=1}^{d_i} L_{i,j}(\bar{x}) \right) \cdot h_i \left(\tilde{L}_{i,1}(\bar{x}), \ldots, \tilde{L}_{i,\rho_i}(\bar{x}) \right).$$

Let $r = \dim\left(\mathrm{span}\{(\tilde{L}_{i,j})\}_{i\in[k],j\in[\rho_i]}\right)$. Clearly, $r \leq k\cdot\rho$. Assume for simplicity and w.l.o.g. that $x_1,\ldots,x_r$ form a basis to the linear space spanned by $\{\tilde{L}_{i,j}\}_{i\in[k],j\in[\rho_i]}$. Let $\overline{\mathbb{F}}$ be the algebraic closure of $\mathbb{F}$. For each $\bar{u} \in \overline{\mathbb{F}}^r$ define $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ to be the circuit resulting from substituting u_i to x_i for $i\in[r]$. Notice that for each such $\bar{u}$, all the functions $h_i|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ are set to constants. In particular, $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ is a (non-generalized) $\Sigma\Pi\Sigma$ circuit with (at most) k multiplication gates, of degree bounded by d, that computes the zero polynomial. We shall now prove the existence of $\bar{u}\in\overline{\mathbb{F}}^r$ such that $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ is simple and minimal. For this $\bar{u}$ we will get that

$$\mathrm{rank}(C) \leq \mathrm{rank}(C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}) + r < R(k,d) + k\cdot\rho = R(k,d,\rho). \tag{5}$$

We prove the existence of such $\bar{u}$ by giving a non-zero r-variate polynomial $q(y_1,\ldots,y_r)$ such that for each $\bar{u}\in\overline{\mathbb{F}}^r$, if $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ is not simple or minimal then $q(\bar{u})=0$. As $q\not\equiv 0$, there are many $\bar{u}\in\overline{\mathbb{F}}^r$ for which $q(\bar{u})\neq 0$ and so Equation (5) holds. The polynomial q will be the product of two polynomials. One of them will "take care" of the simplicity requirement and the other will "take care" of the minimality requirement.

Lemma 4.3. *Let C be a simple $\Sigma\Pi\Sigma(k,d,\rho)$ circuit in n indeterminates, given by Equation (3). Let $r<n$ be an integer. Assume that the linear functions $\{\tilde{L}\}_{i\in[k],j\in[\rho_i]}$ depend only on the variables $x_1,\ldots,x_r$. Then there exists a non-zero r-variate polynomial p such that for every assignment $\bar{u}$ to $x_1,\ldots,x_r$, if $p(\bar{u})\neq 0$ then $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ is also a simple circuit (that is, after substituting u_i to x_i, for $i\in[r]$, the resulting circuit is simple).*

Proof. Assume that for some vector $\bar{u}$, $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ is not simple. Assume further that no M_i was set to zero by the assignment $\bar{u}$. Then it must be the case that $\gcd(C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}})\neq 0$. In particular, for some pair of linearly independent linear functions L, L' in $\widehat{C}$ (see Definition 2.1), their restrictions $L(u_1,\ldots,u_r,x_{r+1},\ldots,x_n)$ and $L'(u_1,\ldots,u_r,x_{r+1},\ldots,x_n)$ are non-constant linearly dependent linear functions. Note that there exists at most one $\gamma_{L,L'}\in\mathbb{F}$ (that is independent of $\bar{u}$) such that $L(u_1,\ldots,u_r,x_{r+1},\ldots,x_n) - \gamma_{L,L'}\cdot L'(u_1,\ldots,u_r,x_{r+1},\ldots,x_n) = 0$. For each such pair of linearly independent linear functions we define $p_{L,L'}(x_1,\ldots,x_n) \triangleq L(x_1,\ldots,x_n) - \gamma_{L,L'}\cdot L'(x_1,\ldots,x_n)$. Since L and L' are linearly independent, it follows that $p_{L,L'}\neq 0$. Let the polynomial p' be defined as:

$$p'(x_1,\ldots,x_n) \triangleq \prod_{i=1}^{k} M_i \cdot \prod_{L\neq L'\in C} p_{L,L'}.$$

That is, p' is the product of all of polynomials corresponding to the different pairs of linearly independent linear functions times the product of all the multiplication gates. Clearly, $p'\not\equiv 0$. In particular there exists an assignment $\bar{w}\in\overline{\mathbb{F}}^{n-r}$ to $(x_{r+1},\ldots,x_n)$ such that $p(x_1,\ldots,x_r) \triangleq p'(x_1,\ldots,x_r,w_1,\ldots,w_{n-r}) \not\equiv 0$. Furthermore, for any vector $\bar{u}$ for which $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ is not simple, $p(\bar{u})=0$ (because if none of the h_i was set to zero by $\bar{u}$ then one of the linear factors of p' must vanish on $\bar{u}$). This completes the proof of Lemma 4.3. □

We now construct a polynomial that will vanish on $\bar{u}$ only if $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ is not minimal.

Lemma 4.4. *Let C be a minimal $\Sigma\Pi\Sigma(k,d,\rho)$ circuit in n indeterminates. Let $r<n$ be an integer. Then there exists a non-zero r-variate polynomial p such that for every assignment $\bar{u}$ to $x_1,\ldots,x_r$, if $p(\bar{u})\neq 0$ then $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ is also a minimal circuit (that is, after substituting u_i to x_i, for $i\in[r]$, the resulting circuit is minimal).*

Proof. For every subset $\emptyset \neq A \subsetneq [k]$ let $p_A = C_A$. That is, p_A is the polynomial computed by C_A. As C is minimal we get that $p_A \not\equiv 0$. Let p' be the product of all the different p_A's. That is, $p'(x_1, \ldots, x_n) = \prod_{\emptyset \neq A \subsetneq [k]} p_A(x_1, \ldots, x_n)$. Clearly p' is not the zero polynomial. In particular there exists a substitution $\bar{w} \in \overline{\mathbb{F}}^{n-r}$, such that $p(x_1, \ldots, x_r) \overset{\Delta}{=} p'(x_1, \ldots, x_r, w_1, \ldots, w_{n-r}) \not\equiv 0$. Now, if $\bar{u}$ is such that $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ is not minimal then, in particular, for some $\emptyset \neq A \subsetneq [k]$, we have that $(C_A)|_{(x_1,\ldots,x_r)\leftarrow\bar{u}} \equiv 0$. In other words, we have that $p_A(u_1, \ldots, u_r, w_1, \ldots, w_{n-r}) = 0$. This implies that $p(\bar{u}) = 0$. This completes the proof of Lemma 4.4. □

To complete the proof of Lemma 4.2 we define the polynomial q to be the product of the two polynomials guaranteed by Lemma 4.3 and Lemma 4.4. It follows that if for some $\bar{u} \in \overline{\mathbb{F}}^r$, $q(\bar{u}) \neq 0$, then $C|_{(x_1,\ldots,x_r)\leftarrow\bar{u}}$ is minimal and simple. By the discussion before Equation 5 this is enough to complete the proof of the lemma. □

Next we construct a family of subspaces containing at least one $R(k, d, \rho)$-rank-preserving subspace for every possible $\Sigma\Pi\Sigma(k, d, \rho)$ circuit.

4.1 Construction of rank-preserving subspaces

In this section we construct a small set of affine subspaces that contains an $R(k, d, \rho)$-rank-preserving subspace for every possible $\Sigma\Pi\Sigma(k, d, \rho)$ circuit. By Theorem 3.4 and Lemmas 4.1 and 4.2 we know that if the restriction of a $\Sigma\Pi\Sigma(k, d, \rho)$ circuit to each of the subspaces in the set computes the zero polynomial, then so does the circuit itself. We note that the properties of rank-preserving subspaces can be formalized as properties of the linear transformations corresponding to the subspaces (recall the definition from Section 2). In [GR05] Gabizon and Raz make use of linear transformations with very similar properties. As a consequence, our construction relies heavily on the construction of [GR05].

The section is organized as follows. We first present a lemma from [GR05] that was slightly modified to suit our notations and needs. We proceed by defining a subspace such that the transformation corresponding to it is the same transformation defined in [GR05]. We end the section with a theorem proving that the rank preserving properties of the transformations of [GR05] give exactly what we need.

Lemma 4.5 (Lemma 6.1 of [GR05]). *For an element $0 \neq \alpha \in \mathbb{F}$ and integers $m \geq t > 0$ define $\varphi_{\alpha,t,m} : \mathbb{F}^m \to \mathbb{F}^t$ to be the following linear transformation*

$$\varphi_{\alpha,t,m}(a_0, \ldots, a_{m-1}) = \left(\sum_{i=0}^{m-1} a_i \alpha^i, \sum_{i=0}^{m-1} a_i \alpha^{2i}, \ldots, \sum_{i=0}^{m-1} a_i \alpha^{t \cdot i} \right).$$

Fix any number of subspaces $W_1, \ldots, W_s \subseteq \mathbb{F}^m$ of dimension at most t. Then there are at most $s \cdot (m-1) \cdot \binom{t+1}{2}$ elements $\alpha \in \mathbb{F}$ for which there exists $i \in [s]$ such that $\dim(\varphi_\alpha(W_i)) < \dim(W_i)$. In other words, for all but $s \cdot (m-1) \cdot \binom{t+1}{2}$ elements of $\mathbb{F}$ we have that $\forall i \in [s]$, $\dim(\varphi_\alpha(W_i)) = \dim(W_i)$.

We now define, for each $\alpha \in \mathbb{F}$, an affine linear subspace V_α such that its corresponding linear transformation is $\varphi_{\alpha,R(k,d,\rho)+1,n+1}$. That is, by the notations of Section 2, $T_{V_\alpha} = \varphi_{\alpha,R(k,d,\rho)+1,n+1}$. For convenience, we denote $\varphi_\alpha \overset{\Delta}{=} \varphi_{\alpha,R(k,d,\rho)+1,n+1}$.

Definition 4.6. *Let $\alpha \in \mathbb{F}$ be a field element. Set $r = R(k, d, \rho)$.*

- *For $0 \leq j \leq r$ define $v_{j,\alpha} \in \mathbb{F}^n$ as $v_{j,\alpha} \overset{\Delta}{=} (\alpha^{j+1}, \ldots, \alpha^{n(j+1)})$.*

- *Let P_α be the $n \times r$ matrix whose j-th column (for $1 \leq j \leq r$) is $v_{j,\alpha}$. Namely,*

$$P_\alpha = (v_{1,\alpha}, \ldots, v_{r,\alpha}) = \begin{pmatrix} \alpha^2 & \alpha^3 & \ldots & \alpha^{r+1} \\ \alpha^4 & \alpha^6 & \ldots & \alpha^{2(r+1)} \\ \vdots & \ddots & & \vdots \\ \alpha^{2n} & & \ldots & \alpha^{n(r+1)} \end{pmatrix}.$$

- *Let $V_{0,\alpha}$ be the linear subspace spanned by $\{v_{j,\alpha}\}_{j\in[r]}$. Let $V_\alpha \subseteq \mathbb{F}^n$ be the affine subspace $V_\alpha = V_{0,\alpha} + v_{0,\alpha}$. In other words,*

$$V_\alpha = \left\{ P_\alpha \bar{y} + v_{0,\alpha} \;:\; \bar{y} \in \mathbb{F}^r \right\}.$$

Lemma 4.7. *For every $\alpha \in \mathbb{F}$, we have that $T_{V_\alpha} = \varphi_\alpha$, where every linear function $L|_{V_\alpha}(y_1, \ldots, y_r)$ is defined w.r.t. the basis $\{v_{j,\alpha}\}_{j\in[r]}$ of $V_{0,\alpha}$.*

Proof. Let L be a linear function in n variables. Denote $L(x_1, \ldots, x_n) = a_0 + \sum_{i=1}^n a_i x_i$. We need to show that the vector corresponding to $L|_{V_\alpha}$ is equal to $\varphi_\alpha(a_0, \ldots, a_n)$. Namely, we would like to show that the vector of coefficients of $L|_{V_0}$, with respect to the basis $\{v_{i,\alpha}\}_{i\in[r]}$ of $V_{0,\alpha}$, is

$$\left(\sum_{i=0}^n a_i \alpha^i, \sum_{i=0}^n a_i \alpha^{2i}, \ldots, \sum_{i=0}^n a_i \alpha^{(r+1)i} \right).$$

For convenience, we denote $L|_{V_\alpha}(y_1, \ldots, y_r) = \sum_{i=1}^r b_i y_i + b_0$. In other words, b_i $(0 \leq i \leq r)$ is the i'th entry of the vector corresponding to $L|_{V_\alpha}$. Denote $\bar{a} = (a_1, \ldots, a_n)$. We get that

$$L|_{V_\alpha}(\bar{y}) = L\left(\sum_{i=1}^r y_i \cdot v_{i,\alpha} + v_{0,\alpha} \right) = L(P_\alpha \cdot \bar{y} + v_{0,\alpha}) = \bar{a} \cdot (P_\alpha \cdot \bar{y}) + \bar{a} \cdot v_{0,\alpha} + a_0 = (\bar{a} \cdot P_\alpha) \cdot \bar{y} + \bar{a} \cdot v_{0,\alpha} + a_0.$$

The free term in this equation is

$$b_0 = \bar{a} \cdot v_{0,\alpha} + a_0 = \sum_{i=0}^n a_i \alpha^i.$$

For $1 \leq j \leq r$ we have that

$$b_j = (\bar{a} \cdot P_\alpha)_j = \sum_{i=0}^n a_i \alpha^{(j+1)i}$$

as required. □

We now prove the main theorem of this section that shows that for a fixed $\Sigma\Pi\Sigma(k, d, \rho)$ circuit C, except of a small number of $\alpha \in \mathbb{F}$, we have that V_α is $R(k, d, \rho)$-rank-preserving for C.

Theorem 4.8. *Let C be a $\Sigma\Pi\Sigma(k, d, \rho)$ circuit over a field $\mathbb{F}$. Then there are at most*

$$\left(\binom{dk}{2} + 2^k \right) \cdot n \cdot \binom{R(k, d, \rho) + 2}{2}$$

different $\alpha \in \mathbb{F}$ such that V_α is not $R(k, d, \rho)$-rank-preserving for C.

Proof. The proof is in two steps. We first construct several subspaces (that are defined using linear functions from C), each of dimension $\leq R(k,d,\rho)+1$, such that if φ_α $(= \varphi_{\alpha,R(k,d,\rho)+1,n+1})$, the linear transformation given in Lemma 4.5, preserves the rank of all of them (in the sense of Lemma 4.5) then V_α is a $R(k,d,\rho)$-rank-preserving subspace for C. We then use lemma 4.5 to prove that except a small number of α-s, φ_α indeed preserves the rank of all those subspaces.

We shall use the following notations during the proof. Assume that $C = \sum_{i=1}^{k} M_i$ as given by Equation (3). Recall the definition of $\widehat{C} = \sum_{i=1}^{k} \mathrm{Lin}(M_i)$. We now define several sets of subspaces such that if V preserves the rank of all of them then V is rank preserving for C.

1. For each pair of linear functions $L \neq L'$ that appear in $\widehat{C}$, where v and v' are the corresponding vectors of coefficients, we define $W_{L,L'} = \mathrm{span}(v, v')$. Clearly $\dim(W_{L,L'}) \leq 2$. The number of such subspaces is at most $\binom{dk}{2}$.

2. For every i, let W_i be the subspace spanned by the vectors corresponding to the linear functions $\{(\tilde{L}_{i,j})^H\}_{j\in[\rho_i]}$ and 1 (i.e., the constant function whose output is the field element 1). Clearly $\dim(W_i) \leq \rho_i + 1 \leq \rho + 1$.

3. Let $W = \cup_{i=1}^{k} W_i$. Clearly $\dim(W) \leq k \cdot \rho + 1$.

4. For every $\emptyset \neq A \subseteq [k]$, let $\hat{r}_A = \mathrm{rank}(\mathrm{sim}(C_A))$. Let $r_A = \min(\hat{r}_A, R(k,d,\rho))$. Let $\{L_i^H\}_{i=1}^{r_A}$ be a set of linearly independent linear functions from $\mathrm{Lin}(\mathrm{sim}(C_A))$. Let $\{v_i\}_{i=0}^{r_A}$ be their corresponding vectors and the vector corresponding to the constant function 1. Set $W_A = \mathrm{span}\,\{v_i\}_{i=0}^{r_A}$. Clearly $\dim(W_A) = r_A + 1 \leq R(k,d,\rho)+1$. The number of such subspaces is at most $2^k - 1$.

Note that for every i we have that $W_i \subseteq W$, but in order to ease the presentation we defined the W_i's as well. We now show that if φ_α preserves all these subspaces (i.e. for every subspace U in our family $\mathrm{rank}(\varphi_\alpha(U)) = \mathrm{rank}(U)$) then V_α is $R(k,d,\rho)$-rank-preserving for C. For this, we will prove that V_α satisfies all the properties of Definition 3.1. Consider the first property. Let L, L' be two linearly independent linear functions appearing in $\widehat{C}$. As φ_α preserves the rank of $W_{L,L'}$ we get that $\dim(\varphi_\alpha(W_{L,L'})) = \dim(W_{L,L'})$, hence $L|_{V_\alpha}, L'|_{V_\alpha}$ remain linearly independent.

To see Property 3 of Definition 3.1 we note that as φ_α preserves the rank of every $W_{L,L'}$, no linear function in $\mathrm{Lin}(M_i)$ was restricted to zero. Hence $\mathrm{Lin}(M_i)|_{V_\alpha} \not\equiv 0$. We also note that since $\dim(\varphi_\alpha(W_i)) = \dim(W_i)$, we have that $(\tilde{L}_{i,1}|_{V_\alpha})^H, \ldots, (\tilde{L}_{i,\rho_i}|_{V_\alpha})^H$ are linearly independent. As h_i (the non-linear term of M_i) is not the zero polynomial then $h_i(\tilde{L}_{i,1}|_{V_\alpha}, \ldots, \tilde{L}_{i,\rho_i}|_{V_\alpha}) \not\equiv 0$. Hence, $M_i = \mathrm{Lin}(M_i) \cdot h_i$ was not restricted to zero.

We note that by the same argument we also get Property 4, as basically each h_i remains the same polynomial after the restriction to V_α (up to applying an invertible linear transformation on its inputs) and therefore it has the same factorization before and after the restriction. Hence no new linear factors where added to $\mathrm{sim}(M_i)|_{V_\alpha}$.

To see that Property 2 of Definition 3.1 is satisfied, we consider some sub-circuit C_A, for some $\emptyset \neq A \subseteq [k]$. As we just showed that Properties 1,3 and 4 hold, we get by Lemma 3.2 that $\mathrm{sim}(C_A)|_{V_\alpha} = \mathrm{sim}(C_A|_{V_\alpha})$ (recall that the proof of this item from Lemma 3.2 did not use Property 2 of Definition 3.1). Since φ_α preserves the rank of W_A, and $\varphi_\alpha(W_A)$ is contained in $\mathrm{span}\left(\mathrm{Lin}(\mathrm{sim}(C_A|_{V_\alpha})) \cup \{1\}\right)$, we get that

$$\mathrm{rank}(\mathrm{sim}(C_A|_{V_\alpha})) \geq \dim(\varphi_\alpha(W_A)) - 1 = \dim(W_A) - 1 = r_A$$

$$= \min(\hat{r}_A, R(k,d,\rho)) = \min\left(\mathrm{rank}(\mathrm{sim}(C_A)), R(k,d,\rho)\right)$$

as required.

We now bound the number of α's for which V_α does not preserve the rank of (at least) one of the subspaces that we defined. The number of subspaces that we defined is clearly bounded by $\binom{dk}{2} + 2^k$. Therefore, by Lemma 4.5 (for $t = R(k,d,\rho)+1$ and $m = n+1$) we get that there are at most

$$\left(\binom{dk}{2} + 2^k\right) \cdot n \cdot \binom{R(k,d,\rho)+2}{2}$$

bad α's. In other words, except for $\left(\binom{dk}{2} + 2^k\right) \cdot n \cdot \binom{R(k,d,\rho)+2}{2}$ many α's, all the V_α's are $R(k,d,\rho)$-rank-preserving for C. This completes the proof of the theorem. □

The following corollary shows how to get a (relatively) small set of subspaces such that for every $\Sigma\Pi\Sigma(k,d,\rho)$ circuit C, most of the subspaces are $R(k,d,\rho)$-rank-preserving for C.

Corollary 4.9. *Let $S \subseteq \mathbb{F}$ be a set of $n\left(\binom{kd}{2} + 2^k\right)\binom{R(k,d,\rho)+2}{2}/\epsilon$ different elements of the field*[6]. *Then, for every $\Sigma\Pi\Sigma(k,d,\rho)$ circuit C over $\mathbb{F}$, there are at least $(1-\epsilon)|S|$ elements $\alpha \in S$ such that V_α is $R(k,d,\rho)$-rank-preserving subspace for C.*

4.2 The PIT algorithm for $\Sigma\Pi\Sigma(k,d,\rho)$ circuits

We now present our algorithms and prove Theorems 1 and 3. Algorithm 1 gives a quasi-polynomial time deterministic algorithm for PIT of $\Sigma\Pi\Sigma(k,d,\rho)$ circuits (when ρ is not too large) using the method described in section 3. Algorithm 2 gives an efficient randomized algorithm that makes a single query to the black-box.

Algorithm 1 Deterministic black-box PIT algorithm for $\Sigma\Pi\Sigma(k,d,\rho)$ circuits

Input: $k,n,d,\rho \in \mathbb{N}$, and oracle access to a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit C in n indeterminates.
Output: Determine whether $C \equiv 0$.

For $\alpha \in \mathbb{F}$ let P_α be the $n \times R(k,d,\rho)$ matrix for which $(P_\alpha)_{i,j} = \alpha^{i(j+1)}$. Let $v_{0,\alpha} = (\alpha, \alpha^2, \ldots, \alpha^n)$. Let $S,T \subseteq \mathbb{F}$ be subsets such that $|S| = n\left(\binom{kd}{2} + 2^k\right)\binom{R(k,d,\rho)+2}{2} + 1$ and $|T| = d+1$. Define

$$\mathcal{H} = \left\{ P_\alpha \bar{y} + v_{0,\alpha} \ : \ \alpha \in S \ \text{ and } \ \bar{y} \in T^{R(k,d,\rho)} \right\}.$$

If for every point $\bar{z} \in \mathcal{H}$, $C(\bar{z}) = 0$, then return "zero circuit".
Else, return "non-zero circuit".

Lemma 4.10. *Let C be a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit. Then Algorithm 1, when given k,d,ρ,n as input and black-box access to C, returns "zero circuit" if and only if $C \equiv 0$. The running time of the algorithm is $|S| \cdot (d+1)^{R(k,d,\rho)}$ $(= \mathrm{poly}(n) \cdot \exp((\log d)^{k-1} + k\rho \log d))$.*

Proof. The claim regarding the running time is clear as the running time is equal to $|\mathcal{H}|$ and we have

$$|\mathcal{H}| = |S| \cdot |T|^{R(k,d,\rho)} = \left(n\left(\binom{kd}{2} + 2^k\right)\binom{R(k,d,\rho)+2}{2} + 1 \right) \cdot (d+1)^{R(k,d,\rho)}.$$

We now prove the correctness of the algorithm. For $\alpha \in S$ let $V_\alpha = \left\{ P_\alpha \bar{y} + v_{0,\alpha} \ : \ \bar{y} \in \mathbb{F}^{R(k,d,\rho)} \right\}$. Denote $\mathcal{H}_\alpha = \left\{ P_\alpha \bar{y} + v_{0,\alpha} \ : \ \bar{y} \in T^{R(k,d,\rho)} \right\}$. In other words, $\mathcal{H}_\alpha$ corresponds to a box isomorphic

[6] Recall our assumption that if $|\mathbb{F}|$ is not large enough then we work over an algebraic extension field of $\mathbb{F}$.

to $T^{R(k,d,\rho)}$ inside V_α. Theorem 4.8 implies that for some $\alpha \in S$, V_α is $R(k,d,\rho)$-rank-preserving for C. As V_α is closed w.r.t. the family of $\Sigma\Pi\Sigma(k,d,\rho)$ circuits (Lemma 4.1), we get by theorem 3.4 that if $C \not\equiv 0$ then $C|_{V_\alpha} \not\equiv 0$. Note that as $C|_{V_\alpha}$ is a polynomial of degree at most d in $\{y_i\}_{i\in[R(k,d,\rho)]}$ then by the Schwartz-Zippel lemma below (see [Sch80, Zip79]) we have that $C|_{V_\alpha} \equiv 0$ if and only if $C|_{\mathcal{H}_\alpha} = 0$. In particular $C \equiv 0$ if and only if $C|_{\mathcal{H}} = 0$. □

Lemma 4.11 (Schwartz-Zippel)**.** *Let $f(x_1, ..., x_m)$ be a non-zero m-variate polynomial of degree d, over a field $\mathbb{F}$. Let $S \subseteq \mathbb{F}$ be a subset of the field. Then the probability that f vanishes on a randomly chosen input from S^m is bounded by*

$$\Pr_{\mathrm{C}}[f(x_1, ..., x_m) = 0] \le \frac{d}{|S|}.$$

In particular, if $|S| > d$ and $f \neq 0$ then $f|_{S^m} \neq 0$. Moreover, if f is of degree at most d in each variable (so the total degree can be $d \cdot m$) and $|S| > d$ then there exists some $\bar{x} \in_R S^m$ such that $f(x_1, ..., x_m) \neq 0$.

Theorem 1 now follows easily.

Proof of Theorem 1. By Lemma 4.10 we have that Algorithm 1 decides correctly whether $C \equiv 0$ and runs in time $\left(n\left(\binom{kd}{2}+2^k\right)\binom{R(k,d,\rho)+2}{2}+1\right)\cdot(d+1)^{R(k,d)}$. As $R(k,d,\rho) = O\left((\log d)^{k-2} + k\rho\right)$ the theorem follows. □

From Lemma 4.11 it is clear that if we make the set T large enough then if $C \not\equiv 0$ then a random input from $\mathcal{H}$ will be a non-zero of C with high probability. This is formalized in Algorithm 2.

Algorithm 2 Randomized black-box PIT algorithm for $\Sigma\Pi\Sigma(k,d,\rho)$ circuits

Input: $\epsilon, k, n, d, \rho \in \mathbb{N}$, and oracle access to a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit C in n input variables.
Output: Determine whether $C \equiv 0$.

For $\alpha \in \mathbb{F}$ let P_α be the $n \times R(k,d,\rho)$ matrix for which $(P_\alpha)_{i,j} = \alpha^{i(j+1)}$. Let $v_{0,\alpha} = \left(\alpha, \alpha^2, \ldots, \alpha^n\right)$. Let $S_\epsilon, T_\epsilon \subseteq \mathbb{F}$ be subsets such that $|S_\epsilon| = 2n\left(\binom{kd}{2}+2^k\right)\binom{R(k,d,\rho)+2}{2}/\epsilon$ and $|T_\epsilon| = 2d/\epsilon$. Define

$$\mathcal{H}_\epsilon = \left\{P_\alpha \bar{y} + v_{0,\alpha} \;:\; \alpha \in S_\epsilon \text{ and } \bar{y} \in T_\epsilon^{R(k,d,\rho)}\right\}.$$

Pick a random point $\bar{z} \in \mathcal{H}$. If $C(\bar{z}) = 0$ then return "zero circuit".
Else, return "non-zero circuit".

Lemma 4.12. *Let C be a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit. Let $\epsilon > 0$ be a constant. If $C \not\equiv 0$ then Algorithm 2, when given ϵ, k, d, n, ρ as input and black-box access to C, returns "non-zero circuit" with probability at least $1-\epsilon$. If $C \equiv 0$ then the algorithm always answers "zero circuit". The number of random bits used by the algorithm is*

$$\log|\mathcal{H}_\epsilon| = \log|S_\epsilon| + R(k,d,\rho)\log|T_\epsilon| = O\left(R(k,d,\rho)\log 1/\epsilon + R(k,d,\rho)\log d + \log n\right)$$

Proof. As before, for $\alpha \in S_\epsilon$ let

$$V_\alpha = \left\{P_\alpha \bar{y} + v_{0,\alpha} \;:\; \bar{y} \in \mathbb{F}^{R(k,d,\rho)}\right\}.$$

Denote

$$\mathcal{H}_{\alpha,\epsilon} = \left\{ P_\alpha \bar{y} + v_{0,\alpha} \ : \ \bar{y} \in T_\epsilon^{R(k,d,\rho)} \right\}.$$

Corollary 4.9 implies that if $C \not\equiv 0$ then for $(1-\epsilon/2)$ of the elements $\alpha \in S_\epsilon$, we have that $C|_{V_\alpha} \not\equiv 0$. For such an α we have that $C|_{V_\alpha}$ is a polynomial of degree at most d in $\{y_i\}_{i\in[R(k,d,\rho)]}$ and by the Schwartz-Zippel lemma (Lemma 4.11) we have that

$$\mathrm{Pr}_{\bar{z}\in_{\mathrm{R}}\mathcal{H}_{\alpha,\epsilon}}[C(\bar{z}) = 0] \leq \frac{d}{|T_\epsilon|} = \epsilon/2.$$

In particular, if $C \not\equiv 0$ then with probability at least $1-\epsilon$ the algorithm outputs "non-zero circuit". The claim regarding the number of random bits is clear. □

As before, Theorem 3 is an immediate corollary of Lemma 4.12. We note that the set $\mathcal{H}$ defined in Algorithm 1, and the set $\mathcal{H}_\epsilon$ defined in Algorithm 2 give rise to test sets for $\Sigma\Pi\Sigma(k,d,\rho)$ circuits. More accurately, let $\mathcal{H}$ and $\mathcal{H}_\epsilon$ be the sets corresponding to $\Sigma\Pi\Sigma(2k,d,\rho)$ circuits. Then, as an immediate consequence of Theorem 1, we get that any two $\Sigma\Pi\Sigma(k,d,\rho)$ circuit that agree on all the points of $\mathcal{H}$ compute the same polynomial. Similarly we get that any two $\Sigma\Pi\Sigma(k,d,\rho)$ circuits that compute different polynomials get different values on $1-\epsilon$ of the points in $\mathcal{H}_\epsilon$.

4.3 PIT for generalized $\Sigma\Pi\Sigma(k,d,\rho)$ circuits

In this section we prove Theorem 4. The theorem concerns n-variate $\Sigma\Pi\Sigma(m,d,\rho)$ circuits where each variable appears in at most k multiplication gates. The number of multiplication gates m will not play an important role in our results. Hence, we refer to this type of circuits as k-$\Sigma\Pi\Sigma(\cdot,d,\rho)$ circuits. Obviously, a $\Sigma\Pi\Sigma(k,d,\rho)$ circuit is also a k-$\Sigma\Pi\Sigma(\cdot,d,\rho)$ circuit. We give a PIT algorithm for k-$\Sigma\Pi\Sigma(\cdot,d,\rho)$ circuits by reducing it to the case of PIT to $\Sigma\Pi\Sigma(2k,d,\rho)$ circuits.

Let C be an n-variate k-$\Sigma\Pi\Sigma(\cdot,d,\rho)$ circuit. Algorithm 3 deterministically verifies whether $C \equiv 0$. The idea is based on the following simple observation: Since each input variable appears in at most k multiplication gates, then $C' \equiv C - C|_{x_n=0}$, is a $\Sigma\Pi\Sigma(2k,d,\rho)$ circuit[7].

Algorithm 3 Deterministic PIT for k-$\Sigma\Pi\Sigma(\cdot,d,\rho)$ circuits

Input: $k,n,d,\rho \in \mathbb{N}$, and oracle access to a k-$\Sigma\Pi\Sigma(\cdot,d,\rho)$ circuit C in n input variables.
Output: Determine whether $C \equiv 0$.

Recursively verify that the k-$\Sigma\Pi\Sigma(\cdot,d,\rho)$ circuit, $C|_{x_n=0}$ (that has only $n-1$ inputs) computes the zero polynomial. If not then return "non-zero circuit". If $C|_{x_n=0} \equiv 0$ then run Algorithm 1 on C, viewed as an n-variate $\Sigma\Pi\Sigma(2k,d,\rho)$ circuit and return its output.

Lemma 4.13. *Algorithm 3 deterministically determines whether the given circuit computes the zero polynomial. The running time of the circuit is bounded by $O(n)$ times the running time of Algorithm 1.*

Proof. We begin by showing the algorithm correctness. In the first stage, if we find that $C|_{x_n=0} \neq 0$ then obviously, $C \neq 0$ and the algorithm outputs the correct answer. If indeed $C|_{x_n=0} \equiv 0$, then

[7] Formally, for a $\Sigma\Pi\Sigma(m,d,\rho)$ circuit C, the circuit $C - C|_{x_n=0}$ has $2m$ multiplication gates. However, we remove every pair of multiplication gates that cancel each other (this removes all gates in which x_n does not appear) and the resulting circuit has at most 2k multiplication gates.

$C \equiv C' \overset{\triangle}{=} C - C|_{x_n=0}$. Moreover, C' is a $\Sigma\Pi\Sigma(2k, d, \rho)$ circuit. Hence, Algorithm 1 will determine whether C', and therefore C, computes the zero polynomial.

The claim regarding the running time follows easily from the recursion formula $T_n = T_{n-1} + A_n$, where T_n is the running time of the algorithm when there are n variables (the parameters k, d, ρ are part of T_n), and A_n is the running time of Algorithm 1 when given $2k, d, \rho, n$ as parameters. □

5 PIT for read-k $\Sigma\Pi\Sigma$ Circuits

In this section we deal with $\Sigma\Pi\Sigma$ circuits in n variables in which every input variable appears in at most k linear functions.[8] This model is known as read-k $\Sigma\Pi\Sigma$ circuit. Notice that a multilinear $\Sigma\Pi\Sigma(k)$ circuit is also a read-k $\Sigma\Pi\Sigma$ circuit (recall that a multilinear $\Sigma\Pi\Sigma$ circuit is a circuit in which every multiplication gate computes a multilinear polynomial). The main result of this section is a deterministic polynomial time black-box PIT algorithm for read-k $\Sigma\Pi\Sigma$ circuits.

Using similar methods to those in section 4.3 we can reduce the problem of PIT for read-k $\Sigma\Pi\Sigma$ circuits to PIT of read-$2k$ $\Sigma\Pi\Sigma$ circuits with at most $2k$ multiplication gates. This can be seen by noticing that as in section 4.3, the circuits C and $C|_{x_n=0}$ differ in at most k multiplication gates. We define $\mathcal{F}_k$ to be the family of $\Sigma\Pi\Sigma$ circuits that have at most $2k$ multiplication gates and each multiplication gate is read-k. In particular, for a read-k $\Sigma\Pi\Sigma$ circuit C, we have that $C - C|_{x_n=0}$ belongs to $\mathcal{F}_k$. Our proof follows the same line as Theorem 3.4. In order to apply the theorem we need to bound the rank of a simple and minimal circuit from $\mathcal{F}_k$ that computes the zero polynomial. Then we have to find a family of subspaces that is rank-preserving and that $\mathcal{F}_k$ is closed with respect to them. The following lemma gives a simple lower bound on the rank of every circuit in $\mathcal{F}_k$.

Lemma 5.1. *Let $\{L_i(\bar{x})\}_{i=1}^d$ be a set of d linear functions, such that every input variable appears in at most k of the linear functions. Then* $\mathrm{rank}\left(\{L_i(\bar{x})\}_{i=1}^d\right) \geq d/k$*. In particular, if C is a circuit in $\mathcal{F}_k$ then* $\mathrm{rank}(C) \geq \deg(C)/k$*.*

Proof. The proof is by a induction on d. When there are $1 \leq d \leq k$ linear functions the claim is obvious. Now, for $k < d$ assume w.l.o.g. that x_1 appears in the linear functions $L_1, \ldots, L_t$ for some $t \leq k$, and in no other linear function. By the induction hypothesis we have that $\mathrm{rank}\left(\{L_i(\bar{x})\}_{i=t+1}^d\right) \geq (d-t)/k \geq d/k - 1$. Clearly L_1 is not in the span of $\{L_i(\bar{x})\}_{i=t+1}^d$ (as x_1 does not appear in any of those linear functions). Therefore the total rank is at least $d/k - 1 + 1 = d/k$.

To prove the claim regarding a circuit C in $\mathcal{F}_k$, we recall that every multiplication gate in such a circuit is read-k. By the previous argument it follows that the rank of every multiplication gate of degree d is at least d/k and so the rank of the circuit is at least $\deg(C)/k$. □

Combining the result of the lemma with Theorem 2.3 we get a bound on the rank of a zero, simple and minimal circuit in $\mathcal{F}_k$.

Corollary 5.2. *There exists an integer function $R(k) = 2^{O(k^2)}$ such that for every simple and minimal zero circuit C in $\mathcal{F}_k$,* $\mathrm{rank}(C) < R(k)$*.*

Proof. Let $d = \deg(C)$. By combining Lemma 5.1 with Theorem 2.3 we get that $d/k \leq \mathrm{rank}(C) < 2^{O(k^2)} \cdot \log^{k-2}(d)$. It follows that $d = 2^{O(k^2)}$ and so $\mathrm{rank}(C) < 2^{O(k^2)}$. □

[8] Note that we do not put a restriction on the number of multiplication gates nor on the degree of the circuit. However it is clear that neither can exceed $n \cdot k$.

We now have to come up with a set of rank-preserving subspaces that $\mathcal{F}_k$ is closed with respect to each of them. The delicate point here is that if we consider an arbitrary subspace V then most likely if C is a read-k circuit, then $C|_V$ will not be read-k any more. For example, consider the subspace of co-dimension 1 defined by the equation $x_n = x_1 + x_2 + \ldots + x_{n-1}$. In the circuit $C|_V$ we have to replace every appearance of x_n with $x_1 + \ldots + x_{n-1}$ (we can replace a different variable instead of x_n but the argument will not change). In particular, every linear function that contained x_n can now, possibly, contain all the variables. If we do it for subspaces of larger co-dimension (and in our case $\dim(V)$ is small) then we may lose the read-k property. In order to avoid this kind of trouble we construct rank-preserving subspaces that have a very special form - each variable is either restricted to a constant or is shifted by a constant (that is, we do not "mix" different coordinates). The construction is given in the next subsection.

5.1 Construction of rank-preserving subspaces for the family $\mathcal{F}_k$

In this section we construct a set of subspaces that contain an r-rank-preserving subspace for every circuit in $\mathcal{F}_k$, for some given integer r. Each subspace will be composed from a projection on a small set of coordinates and a shift. It is clear that the restriction of a read-k circuit to such a subspace is again a read-k circuit. The projections alone will preserve the read-k property and will satisfy Property 2 of Definition 3.1, but not Properties 1, 3 and 4. However, as we shall see, the shifted projections will have all the required properties.

Definition 5.3. *Let $B \subseteq [n]$ be a non-empty subset of the coordinates and $\alpha \in \mathbb{F}$ be a field element.*

- *Define V_B as the following subspace: $V_B = \mathrm{span}\{e_i : i \in B\}$, where e_i is the vector that has 1 in the i'th coordinate and zeros elsewhere.*
- *Let $v_{0,\alpha}$ be, as before, the vector $v_{0,\alpha} = \left(\alpha, \alpha^2, \ldots, \alpha^n\right)$.*
- *Let $V_{B,\alpha} = V_B + v_{0,\alpha}$.*

Obviously, for a read-k circuit C, the restricted circuit $C|_{V_{B,\alpha}}$ is also read-k, for every B and α (the restriction assigns the value α^i to every x_i for $i \notin B$, and shifts x_i to $x_i + \alpha^i$ for $i \in B$). In particular we get that $\mathcal{F}_k$ is closed with respect to any subspace $V_{B,\alpha}$. The following theorem shows that if we just consider the set of all V_B-s for $|B| \leq 4^k \cdot r$ then this set contains a subspace that has Property 2 of Definition 3.1.

Theorem 5.4. *Let $C \in \mathcal{F}_k$ be a circuit. Then, for every $r > 0$, there exists a subset $B \subseteq [n]$ such that $|B| \leq 4^k \cdot r$ and B has the following properties*[9]*:*

1. $\forall \emptyset \neq A \subseteq [2k]$, $\mathrm{rank}(\mathrm{sim}(C_A)|_{V_B}) \geq \min\{\mathrm{rank}(\mathrm{sim}(C_A)), r\}$.
2. $C|_{V_B} \in \mathcal{F}_k$.

Proof. It is clear that $C|_{V_B}$ has at most $2k$ multiplication gates and that every multiplication gate is still read-k, and so we turn to prove that the first claim of the theorem holds. Let $A_1, A_2, \ldots, A_{4^k-1}$ be the non-empty subsets of $[2k]$. We first show that for each A_i, there exists a subset $B_i \subseteq [n]$ such that $|B_i| \leq r$ and $\mathrm{rank}(\mathrm{sim}(C_{A_i})|_{V_{B_i}}) = \min\{\mathrm{rank}(\mathrm{sim}(C_{A_i})), r\}$. Indeed, let $R_i = \mathrm{rank}(\mathrm{sim}(C_{A_i}))$, and let $L_1, \ldots, L_{R_i} \in \mathrm{Lin}(\mathrm{sim}(C_{A_i}))$ be such that $(L_1)^H, \ldots, (L_{R_i})^H$ are linearly independent. Denote by Z the $R_i \times n$ matrix whose rows correspond to the vectors of coefficients of $\{(L_j)^H\}_{j \in [R_i]}$. Obviously, there are R_i linearly independent column-vectors in Z. Let $B_i \subseteq [n]$ contain the indices

[9] We assume w.l.o.g. that C has exactly $2k$ multiplication gates.

of $\min\{R_i, r\}$ columns that are linearly independent. We now observe that the matrix corresponding to the vectors of coefficients of the linear functions $\{(L_j|_{V_{B_i}})^H\}_{j\in[R_i]}$ is equal to Z on the columns of B_i, and has zeros elsewhere. As the column rank of Z is equal to its row rank (that is equal to $\min\{R_i, r\}$) we get that the rank of $\{(L_j|_{V_{B_i}})^H\}_{j\in[R_i]}$ is at least $\min\{R_i, r\}$. Hence we get that,

$$\text{rank}(\text{sim}(C_{A_i})|_{V_{B_i}}) \geq \min\{\text{rank}(\text{sim}(C_{A_i})), r\}.$$

Up till now we showed that for every A_i there is a set B_i satisfying $|B_i| = \min\{R_i, r\}$ such that V_{B_i} is good for C_{A_i}. However, it may be the case that different A_i-s need different B_i-s. Therefore we shall consider the set

$$B = \cup_{i=1}^{4^k-1} B_i.$$

It is clear that $|B| < 4^k \cdot r$. Furthermore, for each $\emptyset \neq A_i \subseteq [2k]$ we have that

$$\text{rank}(\text{sim}(C_{A_i})|_{V_B}) \geq \text{rank}(\text{sim}(C_{A_i})|_{V_{B_i}}) \geq \min\{\text{rank}(\text{sim}(C_{A_i})), r\}.$$

This concludes the proof of the theorem. □

The following is an immediate corollary of Theorem 5.4.

Corollary 5.5. *For every $C \in \mathcal{F}_k$ and integer $r > 0$, there exists a subset $B \subseteq [n]$, of size $|B| = 4^k \cdot r$, such that $C|_{V_B} \in \mathcal{F}_k$ and V_B satisfies property 2 of definition 3.1.*

Proof. Let $C \in \mathcal{F}_k$ be a circuit and $B' \subseteq [n]$ be a subset guaranteed by theorem 5.4. Let $B \subseteq [n]$ be such that $B' \subseteq B$ and $|B| = 4^k \cdot r$. It is clear that B also satisfies the requirements of theorem 5.4. □

We also note that if V_B satisfies theorem 5.4 for some circuit C, then so does $V_{B,\alpha}$ for any $\alpha \in \mathbb{F}$. The reason is that restricting to an affine shift of V_B does not decrease the rank of the restricted linear functions. The following theorem shows that for every circuit $C \in \mathcal{F}_k$ there are at most poly(n) many α-s such that $V_{B,\alpha}$ is not rank preserving for the set B guaranteed by Corollary 5.5.

Theorem 5.6. *Let $C \in \mathcal{F}_k$ be a circuit over a field $\mathbb{F}$ and $0 < r \in \mathbb{N}$. Let B be the set guaranteed by Corollary 5.5. Then there are less than $3n^3k^4$ many $\alpha \in \mathbb{F}$ such that $V_{B,\alpha}$ is not r-rank-preserving for C.*

Proof. We already know that for every α, the subspace $V_{B,\alpha}$ satisfies Property 2 of Definition 3.1. We thus have to bound the number of α's for which either Property 1, Property 3 or Property 4 are not satisfied. As we discuss $\Sigma\Pi\Sigma$ circuits, Property 4 is clearly satisfied, so we only have to take care of Properties 1 and 3. We first bound the number of α-s for which $V_{B,\alpha}$ does not satisfy Property 3. Consider a linear function L that appears in C given by $L(x_1, \ldots, x_n) = a_0 + a_1x_1 + \ldots + a_nx_n$, and the subspace $V_{B,\alpha}$ for some α. Then the restriction of L to $V_{B,\alpha}$ is given by $\sum_{i\in B} a_ix_i + L(v_{0,\alpha}) = \sum_{i\in B} a_ix_i + \sum_{i=0}^{n} a_i\alpha^i$. It follows that $L|_{V_{B,\alpha}} = 0$ if and only if L is supported on $[n] \setminus B$ (that is, $a_i = 0$ for $i \in B$) and $a_0 + a_1\alpha + \ldots + a_n\alpha^n = 0$. In particular α must be a root of the polynomial

$$p_L(x) \overset{\Delta}{=} a_0 + a_1x + \ldots a_nx^n$$

(notice that this polynomial does not depend on the set B). As $p_L(x)$ is a non-zero polynomial of degree n it has at most n distinct roots. Going over all linear functions in C we see that there are at most $2n^2k^2$ (specifically, there are $2nk^2$ linear functions appearing in C and each function gives at most n distinct roots) bad α-s for C (that is, these are the only α's that are roots of one of the p_L's).

We now bound the number of α-s for which $V_{B,\alpha}$ violates Property 1. For simplicity we shall only consider those α-s for which Property 3 is satisfied. Let $L, \tilde{L}$ be two linearly independent linear functions appearing in C $(= \widehat{C})$. We have three cases. The first case is that both L and $\tilde{L}$ are supported on $[n] \setminus B$. In this case it is clear that the restriction of both functions to $V_{B,\alpha}$ is constant, for any α, and so all α-s are good. The second case is that exactly one of the functions is supported on $[n] \setminus B$, say L. In this case L is restricted to a constant non-zero function and $\tilde{L}$ is restricted to a non-constant function (no matter what α is) and so they remain linearly independent. The third, and more interesting, case is when both functions are restricted to non constants. Denote $L(\bar{x}) = a_0 + a_1x_1 + \ldots a_nx_n$ and $\tilde{L}(\bar{x}) = \tilde{a}_0 + \tilde{a}_1x_1 + \ldots \tilde{a}_nx_n$. For $L|_{V_{B,\alpha}}$ and $\tilde{L}|_{V_{B,\alpha}}$ to be linearly dependent there must exist a constant $\gamma \in \mathbb{F}$, independent of α, such that $L|_{V_B} = \gamma \cdot \tilde{L}|_{V_B}$. For this γ we have that α must satisfy that $L(v_{0,\alpha}) = \gamma \cdot \tilde{L}(v_{0,\alpha})$ or, equivalently, that $(L - \gamma \cdot \tilde{L})(v_{0,\alpha}) = 0$. As we assumed that L and $\tilde{L}$ are linearly independent we have that $L - \gamma \cdot \tilde{L} \neq 0$. Define the polynomial $p_{L-\gamma\cdot\tilde{L}}(x)$ as before. We see that it must be the case that $p_{L-\gamma\cdot\tilde{L}}(\alpha) = 0$. Thus, α is a root of a degree n polynomial that depends only on $L, \tilde{L}$ and B. In particular, for our B there are at most $n \cdot \binom{2nk^2}{2} < 2n^3k^4$ many α-s such that $V_{B,\alpha}$ violates Property 1.

Concluding, we see that for our B there are less than $2n^2k^2 + 2n^3k^4 < 3n^3k^4$ many α-s for which $V_{B,\alpha}$ is not rank-preserving for C. This concludes the proof of the theorem. □

Corollary 5.7. *Let $S \subseteq \mathbb{F}$ be of size $3n^3k^4$. Let $C \in \mathcal{F}_k$. Then there exists $B \subseteq [n]$ of size $|B| = 4^k \cdot R(k)$ and $\alpha \in S$ such that $V_{B,\alpha}$ is $R(k)$-rank-preserving for C.*

Proof. Follows immediately from Corollary 5.5 and Theorem 5.6. □

5.2 The PIT algorithm for read-k ΣΠΣ circuits

In this section we give the PIT algorithm to read-k ΣΠΣ circuits and prove Theorem 2. Algorithm 5.2 is a deterministic PIT algorithm for read-k ΣΠΣ circuits.

Algorithm 4 Deterministic PIT for read-k ΣΠΣ circuits

Input: $k, n \in \mathbb{N}$, and oracle access to a read-k ΣΠΣ circuit C in n input variables.
Output: Determine whether $C \equiv 0$.

Let $\{0\} \subseteq T \subseteq \mathbb{F}$ be a subset of size $k+1$. If $n = 1$ then if C vanishes on the different points of T then output "zero". Otherwise output "non-zero".
For $n > 1$, recursively run the algorithm on the circuit $C|_{x_n=0}$, with parameters $k, n-1$. If the answer is "non-zero" then return "non-zero". Otherwise let $S \subseteq \mathbb{F}$ be a subset of size $3n^3k^4$. For $\alpha \in \mathbb{F}$ let $v_{0,\alpha} = (\alpha, \ldots, \alpha^n) \in \mathbb{F}^n$. Define $\mathcal{H}$ as

$$\mathcal{H} = \left\{ v + v_{0,\alpha} \; : \; v \in T^n, \; |v| \leq 4^k \cdot R(k), \; \alpha \in S \right\},$$

where $|v|$ is the number of non-zero coordinates in v. If for every point $\bar{z} \in \mathcal{H}$, $C(\bar{z}) = 0$ then output "zero". Otherwise, return "non-zero".

The following lemma shows that Algorithm 5.2 is correct, and gives a trivial upper bound on its running time. Theorem 2 is an immediate corollary of the lemma.

Lemma 5.8. *Let C be a read-k ΣΠΣ circuit. Then Algorithm 5.2, when given k, n as input and oracle access to C, determines whether $C \equiv 0$. the running time of the algorithm is $n^{2^{O(k^2)}}$.*

Proof. Certainly if $C \equiv 0$ then the algorithm returns zero-circuit. So assume that $C \not\equiv 0$. If $n = 1$, then as C is a read-k circuit, its degree (as a univariate polynomial) is at most k. According to the Schwartz-Zippel lemma (Lemma 4.11), if C vanishes on $k+1$ different points then $C \equiv 0$.

For $n > 1$ notice that if $C|_{x_n=0} \not\equiv 0$ then the algorithm outputs "non-zero". So assume that $C|_{x_n=0} \equiv 0$. It follows that $C' \overset{\Delta}{=} C - C|_{x_n=0} \not\equiv 0$. Notice that $C' \in \mathcal{F}_k$.

By Corollary 5.7 we see that there exists a set $B \subseteq [n]$ of size $4^k \cdot R(k)$ and $\alpha \in S$ such that $V_{B,\alpha}$ is $R(k)$-rank-preserving for C'. Theorem 3.4 combined with Corollary 5.2 assures us that $C'|_{V_{B,\alpha}}$, which is also in $\mathcal{F}_k$, is not the zero polynomial. Let $\bar{x}_B$ be the vector of indeterminates that is supported on B, namely, replace x_i with 0 for $i \notin B$. From the definition of $V_{B,\alpha}$ we get that $C'|_{V_{B,\alpha}}$ can be represented as $C'(\bar{x}_B + v_{0,\alpha})$. Note, that $C'(\bar{x}_B + v_{0,\alpha})$ is a polynomial in $|B| = 4^k \cdot R(k)$ variables of degree at most k in each variable (each variable appears at most k times in every multiplication gate). By the Schwartz-Zippel lemma (Lemma 4.11) we get that there is some $\bar{w} \in T^{4^k \cdot R(k)}$ such that[10] $C(\bar{w} + v_{0,\alpha}) = C'(\bar{w} + v_{0,\alpha}) \neq 0$. We can think of $\bar{w}$ as an n-dimensional vector $\bar{w} \in T^n$ of weight $|\bar{w}| \leq |B| = 4^k \cdot R(k)$. Therefore $\bar{z} = \bar{w} + v_{0,\alpha} \in \mathcal{H}$ and so the algorithm will output "non-zero circuit".

To bound the running time we notice that we have the recursion formula

$$T(n) = T(n-1) + |\mathcal{H}| = T(n-1) + \left((k+1)^{4^k \cdot R(k)}\right) \cdot \binom{n}{4^k \cdot R(k)} \cdot (3n^3k^4),$$

where $T(n)$ is the running time of the algorithm on n inputs (k does not change during the execution). The solution to the recursion is

$$T(n) = n^{O\left(4^k \cdot R(k)\right)} = n^{2^{O(k^2)}}.$$

□

References

[AB03] M. Agrawal and S. Biswas. Primality and identity testing via chinese remaindering. *JACM*, 50(4):429–443, 2003.

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

[AM07] V. Arvind and P. Mukhopadhyay. The ideal membership problem and polynomial identity testing. *ECCC Report TR07-095*, 2007.

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

[CDGK91] M. Clausen, A. W. M. Dress, J. Grabmeier, and M. Karpinski. On zero-testing and interpolation of k-sparse multivariate polynomials over finite fields. *Theoretical Computer Science*, 84(2):151–164, 1991.

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

[10] We abuse notations and "redefine" $\bar{w}$ as an n-dimensional vector having zeros in the indices that are not in B and its original elements in the other indices.

[CRS95] S. Chari, P. Rohatgi, and A. Srinivasan. Randomness-optimal unique element isolation with applications to perfect matching and related problems. *SIAM J. on Computing*, 24(5):1036–1050, 1995.

[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.

[GK87] D. Grigoriev and M. Karpinski. The matching problem for bipartite graphs with polynomially bounded permanents is in NC (extended abstract). In *Proceedings of the 28th Annual FOCS*, pages 166–172, 1987.

[GKS90] D. Grigoriev, M. Karpinski, and M. F. Singer. Fast parallel algorithms for sparse multivariate polynomial interpolation over finite fields. *SIAM J. on Computing*, 19(6):1059–1063, 1990.

[GR05] A. Gabizon and R. Raz. Deterministic extractors for affine sources over large fields. In *46th Annual FOCS*, pages 407–418, 2005.

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

[KS96] M. Karpinski and I. Shparlinski. On some approximation problems concerning sparse polynomials over finite fields. *Theoretical Computer Science*, 157(2):259–266, 1996.

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

[KS06] N. Kayal and N. Saxena. Polynomial identity testing for depth 3 circuits. In *Proceedingds of the 21st Annual IEEE Conference on Computational Complexity*, pages 9–17, 2006.

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

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

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

[RS05] R. Raz and A. Shpilka. Deterministic polynomial identity testing in non commutative models. *Computational Complexity*, 14(1):1–19, 2005.

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

[Shp07] A. Shpilka. Interpolation of depth-3 arithmetic circuits with two multiplication gates. In *Proceedings of the 39th Annual STOC*, pages 284–293, 2007.

[SS96] R. E. Schapire and L. M. Sellie. Learning sparse multivariate polynomials over a field with queries and counterexamples. *J. of Computer and System Sciences*, 52(2):201–213, 1996.

[SV08] A. Shpilka and I. Volkovich. Read-once polynomial identity testing. Manuscript, 2008.

[Wer94] K. Werther. The complexity of sparse polynomial interpolation over finite fields. *Applicable Algebra in Engineering, Communication and Computing*, 5:91–103, 1994.

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

A Proof of Lemma 4.5

Notice that by the union bound it is enough to prove the theorem for the case that $s = 1$. Hence, we assume w.l.o.g. that $s = 1$ and that we have only one subspace, W. We shall also assume that $\dim(W) = t$, as any subspace W such that $\dim(W) < t$, is contained in a subspace $W \subseteq W'$ of dimension t, and the equality $\dim(\varphi_\alpha(W')) = \dim(W')$ implies that $\dim(\varphi_\alpha(W)) = \dim(W)$.

Let $\tilde{w}^{(1)}, \ldots, \tilde{w}^{(t)}$ be a basis of W. For convenience we denote $\tilde{w}^{(l)} = (\tilde{w}_0^{(l)}, \ldots, \tilde{w}_{m-1}^{(l)})$. For $j \in [t]$, let j_{max} to be the maximal $i \in \{0, \ldots, m-1\}$ such that $\tilde{w}_i^{(j)}$ is non-zero. Note that (e.g. by using Gaussian elimination) there exists a basis $w^{(1)}, \ldots, w^{(t)}$ of W such that

$$0 \leq 1_{max} < 2_{max} < \ldots < (t)_{max}.$$

Denote with B the $m \times t$ matrix who's j-th column is $w^{(j)}$. That is,

$$B = (w^{(1)}, \ldots, w^{(t)}).$$

Let $P_{\varphi_{\alpha,t,m}}$ be the matrix corresponding to the linear transformation $\varphi_{\alpha,t,m}$ (with respect to the basis $\{e_i\}_{i \in \{0,1,\ldots,m-1\}}$). As $W = B(\mathbb{F}^t)$ we have that

$$\varphi_{\alpha,t,m}(W) = (P_{\varphi_{\alpha,t,m}} \cdot B)(\mathbb{F}^t).$$

Let C_α the $t \times t$ matrix $P_{\varphi_{\alpha,t,m}} \cdot B$. That is,

$$(C_\alpha)_{j,l} = \sum_{i=0}^{m-1} \alpha^{ji} \cdot w_i^{(l)}.$$

Recall that $C_\alpha(\mathbb{F}^t) = \mathbb{F}^t$ if and only if $Det(C_\alpha) \neq 0$. Thus, our result will follow if we show that for most α-s the determinant of C_α is non zero. Let $f(\alpha) = Det(C_\alpha)$. We will show that $f(\alpha)$ is a non-zero polynomial of degree not larger than $(m-1) \cdot \binom{t+1}{2}$ in α. Hence, $Det(C_\alpha) = 0$ for at most $(m-1) \cdot \binom{t+1}{2}$ values of α and the lemma follows. Consider the following representation of f

$$f(\alpha) = Det(C_\alpha) = \sum_{\sigma \in S_t} sgn(\sigma) \cdot f_\sigma(\alpha),$$

where S_t is the group of all permutations of t elements and

$$f_\sigma(\alpha) = \prod_{j=1}^{t} (C_\alpha)_{j,\sigma(j)}.$$

Let $\mathrm{Id} \in S_t$ be the identity permutation. We will show that for every $\sigma \neq \mathrm{Id}$ in S_t, we have that $\deg(f_\sigma) < \deg(f_{\mathrm{Id}})$. Assume for a contradiction that there exists $\sigma \neq \mathrm{Id}$ such that $\deg(f_\sigma) \geq \deg(f_{\mathrm{Id}})$. Fix a permutation $\sigma \neq \mathrm{Id}$ that maximizes $\deg(f_\sigma)$. That is, $\deg(f_\sigma) \geq \deg(f_{\sigma'})$

for every $\sigma' \in S_t$. By definition, $(C_\alpha)_{j,\sigma(j)}$ is a polynomial of degree $j \cdot \sigma(j)_{max}$ in α (as $w_i^{(\sigma(j))} = 0$ for $i > \sigma(j)_{\max}$). Therefore, f_σ has degree

$$\deg(f_\sigma) = \sum_{j=1}^{t} j \cdot \sigma(j)_{max}. \tag{6}$$

By our assumption, $\sigma \neq \mathrm{Id}$, and so there exist $j_1 < j_2$ such that $\sigma(j_1) > \sigma(j_2)$. Let $\tau = (\sigma(j_1), \sigma(j_2)) \cdot \sigma$, i.e. the permutation τ consists of applying σ and then "switching" between $\sigma(j_1)$ and $\sigma(j_2)$. By Equation (6) we get that

$$\begin{aligned}\deg(f_\tau) - \deg(f_\sigma) &= j_2\tau(j_2)_{max} + j_1\tau(j_1)_{max} - j_2\sigma(j_2)_{max} - j_1\sigma(j_1)_{max} \\ &= j_2\sigma(j_1)_{max} + j_1\sigma(j_2)_{max} - j_2\sigma(j_2)_{max} - j_1\sigma(j_1)_{max} \\ &= (j_2 - j_1)(\sigma(j_1)_{max} - \sigma(j_2)_{max}) > 0\end{aligned}$$

which contradicts the maximality of $\deg(f_\sigma)$.

Hence, for any $\sigma \neq \mathrm{Id}$, $\deg(f_\sigma) < \deg(f_{\mathrm{Id}})$. Thus, the highest degree monomial in f_{Id} cannot be cancelled out by the other summands in $f(\alpha)$, and therefore $f(\alpha)$ is a non-zero polynomial of degree

$$\deg(f) = \deg(f_{\mathrm{Id}}) = \sum_{j=1}^{t} j \cdot j_{max} \leq (m-1) \cdot \sum_{j=1}^{t} j = (m-1)\binom{t+1}{2}.$$

This completes the proof of the lemma. □