Hitting-sets for low-distance multilinear depth-3

Manindra Agrawal, Rohit Gurjar, Arpita Korwar, and Nitin Saxena

Indian Institute of Technology, Kanpur, India

Abstract. The depth-3 model has recently gained much importance, as it has become a stepping-stone to understanding general arithmetic circuits. Its restriction to *multilinearity* has known exponential lower bounds but no nontrivial blackbox identity tests. In this paper, we take a step towards designing such hitting-sets. We define a notion of *distance* for multilinear depth-3 circuits (say, in n variables and k product gates) that measures how far are the partitions from a mere *refinement*. The 1-distance strictly subsumes the set-multilinear model, while n-distance captures general multilinear depth-3. We design a hitting-set in time $\text{poly}(n^{\delta \log k})$ for δ-distance. Further, we give an extension of our result to models where the distance is large (close to n) but it is small when restricted to certain variables. This implies the first subexponential whitebox PIT for the sum of constantly many set-multilinear depth-3 circuits.

It is known that invertible width-3 ABPs capture general computation. We explore read-once algebraic branching programs (ROABP) where the factor-matrices are *invertible* (called invertible-factor ROABP). We design a hitting-set in time $\text{poly}(\text{size}^{w^2})$ for width-w invertible-factor ROABP. Further, we could do *without* the invertibility restriction when $w = 2$. Previously, the best result for width-2 ROABP was quasi-polynomial time (Forbes-Saptharishi-Shpilka, STOC 2014). Our results are arithmetic analogues of certain boolean models.

The common thread in all these results is the phenomenon of low-support 'rank concentration'. We exploit the structure of these models to prove rank-concentration after a 'small shift' in the variables. Our proof techniques and results are stronger than the relevant results of Agrawal-Saha-Saxena (STOC 2013) and Forbes-Saptharishi-Shpilka (STOC 2014); giving us quasi-polynomial-time hitting-sets for models where no subexponential *whitebox* algorithms were known before.

1 Introduction

The problem of *Polynomial Identity Testing* is that of deciding if a given polynomial is nonzero. The complexity of the question depends crucially on the way the polynomial is input to the PIT test. For example, if the polynomial is given as a set of coefficients of the monomials, then we can easily check whether the polynomial is nonzero in polynomial time. The problem has been studied for different input models. Most prominent among them is the model of arithmetic circuits. Arithmetic circuits are the arithmetic analog of boolean circuits and are defined over a field $\mathbb{F}$. They are directed acyclic graphs, where every node is a '$+$' or '$\times$' gate and each input gate is a constant from the field $\mathbb{F}$ or a variable from $\overline{x} = \{x_1, x_2, \ldots, x_n\}$. Every edge has a weight from the underlying field $\mathbb{F}$. The computation is done in the natural way. Clearly, the output gate computes a polynomial in $\mathbb{F}[\overline{x}]$. We can restate the PIT problem as: Given an arithmetic circuit $\mathcal{C}$, decide if the polynomial computed by $\mathcal{C}$ is nonzero in time polynomial in the circuit size. Note that, given a circuit, computing the polynomial explicitly is not possible, as it can have exponentially many monomials. However, given the circuit, it is easy to compute an evaluation of the polynomial by substituting the variables with constants.

Though there is no known *deterministic* algorithm for PIT, there are easy randomized algorithms, e.g. [Sch80]. These randomized algorithms are based on the theorem: A nonzero polynomial, evaluated at a random point, gives a nonzero value with a good probability. Observe that such an algorithm does not need to see the structure of the circuit, it just uses the evaluations; it is a *blackbox* algorithm. The other kind of algorithms, where the structure of the input is used, are called *whitebox* algorithms. Whitebox algorithms for PIT have many known applications. E.g. graph matching reduces to PIT. On the other hand, blackbox algorithms (or *hitting-sets*) have connections to circuit lower bound proofs. Arguably, this is currently the only concrete approach

towards lower bounds, see [Mul12a, Mul12b]. See the surveys by Saxena [Sax09, Sax14] and Shpilka & Yehudayoff [SY10] for more applications.

The PIT problem has been studied for various restricted classes of circuits. One such class is depth-3 circuits. A depth-3 circuit is usually defined as a $\Sigma\Pi\Sigma$ circuit: the circuit gates are in three layers, the top layer has an output gate which is $+$, second layer has all $\times$ gates and the last layer has all $+$ gates. In other words, the polynomial computed by a $\Sigma\Pi\Sigma$ circuit is of the form $C(\overline{x}) = \sum_{i=1}^{k} a_i \prod_{j=1}^{n_i} \ell_{ij}$, where n_i is the number of input lines to the i-th product gate and ℓ_{ij} is a linear polynomial of the form $b_0 + \sum_{r=1}^{n} b_r x_r$. An efficient solution for depth-3 PIT is still not known. Recently, it was shown by Gupta et al. [GKKS13], that depth-3 circuits are almost as powerful as general circuits. A polynomial time hitting-set for a depth-3 circuit implies a quasi-poly-time hitting-set for general circuits. Till now, for depth-3 circuits, efficient PIT is known when the top fan-in is assumed to be constant [DS07, KS07, KS09, KS11, SS11, SS12, SS13] and for certain other restrictions [Sax08, SSS13, ASSS12].

On the other hand, there are exponential lower bounds for depth-3 *multilinear* circuits [RY09]. Since there is a connection between lower bounds and PIT [Agr05], we can hope that solving PIT for depth-3 multilinear circuits should also be feasible. This should also lead to new tools for general depth-3.

A polynomial is said to be multilinear if the degree of every variable in every term is at most 1. The circuit $C(\overline{x})$ is a multilinear circuit if the polynomial computed at every gate is multilinear. A polynomial time algorithm is known only for a sub-class of multilinear depth-3 circuits, called *depth-3 set-multilinear circuits*. This algorithm is due to Raz and Shpilka [RS05] and is whitebox. In a depth-3 multilinear circuit, since every product gate computes a multilinear polynomial, a variable occurs in at most one of the n_i linear polynomials input to it. Thus, each product gate naturally induces a *partition* of the variables, where each *color* (i.e. part) of the partition contains the variables present in a linear polynomial ℓ_{ij}. Further, if the partitions induced by all the k product gates are the same then the circuit is called a depth-3 set-multilinear circuit.

Agrawal et al. [ASS13] gave a quasi-polynomial time blackbox algorithm for the class of depth-3 set-multilinear circuits. Their approach is to view the vector of k products, $D(\overline{x}) := (\prod_{j=1}^{n_i} \ell_{ij})_{i=1}^{k}$ as a polynomial over the Hadamard algebra, $\mathbb{H}_k(\mathbb{F})$, and to achieve a *low-support concentration* in it. Low-support concentration means that all the coefficient vectors in $D(\overline{x})$ are linearly dependent on low-support coefficient vectors. We define a new class of circuits called *multilinear depth-3 circuits with δ-distance*. We show low-support concentration for this model. To achieve that we use some deeper combinatorial properties of our model. We also use an improved version of a combinatorial property of the transfer matrix from [ASS13, Theorem 13] and we present it with a simplified proof. Recently, Forbes et al. [FSS14] have also improved [ASS13]. But their methods apply only to set-multilinear models, and not even to 1-distance circuits.

A multilinear depth-3 circuit has δ-distance if there is an ordering on the partitions induced by the product gates, say $(\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_k)$, such that for any color in the partition $\mathbb{P}_i$, there exists a set of $\leq (\delta - 1)$ other colors in $\mathbb{P}_i$ such that the set of variables in the union of these $\leq \delta$ colors are *exactly* partitioned in the upper partitions, i.e. $\{\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_{i-1}\}$. As we will see, such sets of δ colors form equivalence classes of the colors at partition $\mathbb{P}_i$. We call them friendly neighborhoods and they help us in identifying subcircuits. Intuitively, the distance measures how far away are the partitions from a mere *refinement* sequence of partitions, $\mathbb{P}_1 \leq \mathbb{P}_2 \leq \cdots \leq \mathbb{P}_k$. Our first main result gives a blackbox test for this class of circuits (Section 3).

Theorem 1. *Let $C(\overline{x})$ be a δ-distance depth-3, n-variate multilinear circuit with top fan-in k. Then there is a $n^{O(\delta \log k)}$-time hitting-set for $C(\overline{x})$.*

Note that the running time becomes quasi-polynomial when δ is poly-logarithmic in nk. Till now, no subexponential time test was known for this class, even in the whitebox setting. Also, observe that the set-multilinear class is strictly subsumed in the class of 1-distance circuits. E.g. a circuit, whose product gates induce two different partitions $\mathbb{P}_1 = \{\{1\}, \{2\}, \ldots, \{n\}\}$ and $\mathbb{P}_2 = \{\{1, 2\}, \{3, 4\}, \ldots, \{n-1, n\}\}$, has 1-distance but is not set-multilinear. So, poly-logarithmic δ is a significant improvement from set-multilinear. On the other hand, general multilinear depth-3 circuits can have at most n-distance. So, our result is also a first step towards multilinear depth-3 circuits.

Our second result further generalizes this class to *multilinear depth-3 circuits having m base sets with δ-distance*. A circuit is in this class if we can partition the set of variables into m base sets, such that when restricted to any of these base sets, the circuit has δ-distance. E.g. consider a circuit C, whose product gates induce two partitions $\mathbb{P}_1 = \{\{1,2\},\{3,4\},\ldots,\{n-1,n\}\}$ and $\mathbb{P}_2 = \{\{2,3\},\{4,5\},\ldots,\{n,1\}\}$. Clearly, C is $(n/2)$-distance. But, when restricted to any of the two base sets $B_1 = \{1,3,\ldots,n-1\}$ and $B_2 = \{2,4,\ldots,n\}$, $C|_{B_i}$ has 1-distance (in fact, it is set-multilinear). We give a quasi-polynomial time blackbox test for this class, when m and δ are poly-logarithmic and the base sets are *known* (Section 4).

Theorem 2. *If $C(\overline{x})$ is a depth-3 multilinear circuit, with top fan-in k, having m base sets (known) with δ-distance, then there is a $n^{O(m\delta \log k)}$-time hitting-set for C.*

These results generalize to suitable higher-depth circuit models. But, we focus in this paper, only on the depth-3 case to avoid unnecessary complicated notations. We apply Theorem 2 to get the first subexponential time whitebox PIT, for the sum of c (constant) set-multilinear depth-3 circuits (top fan-in k), with time complexity $n^{O(n^{1-\epsilon}\log k)}$, where $\epsilon := 1/2^{c-1}$ (see Appendix F). Thus, we can see that the ideas of δ-distance and base sets are powerful enough to capture the sum of constantly many set-multilinear depth-3 circuits. Since the results of [ASS13] are for set-multilinear circuits, they cannot be used to give a PIT for the above model. Further, no reduction of the above model to ROABPs is known. So, the results of [FSS14] also cannot be used. But the tool of low-distance circuits can be used to handle general partitions, even when the distance is *not* low.

Our third result expands the realm of low-support concentration to multilinear variants of Arithmetic Branching Programs (ABP). An ABP is another interesting model of computing polynomials. It consists of a directed acyclic graph with a source and a sink. The edges of the graph have polynomials as their weights. The weight of a path is the product of the weights of the edges present in the path. The polynomial computed by the ABP is the sum of the weights of all the paths from the source to the sink. It is well known that for an ABP, the underlying graph can seen as a layered graph such that all paths from the source to the sink have exactly one edge in each layer. And the polynomial computed by the ABP can be written as a *matrix product*, where each matrix corresponds to a layer. The entries in the matrices are weights of the corresponding edges. The maximum number of vertices in a layer, or equivalently, the dimension of the corresponding matrices is called the *width* of the ABP. Ben-Or & Cleve [BOC92] have shown that a polynomial computed by a formula of logarithmic depth and constant fan-in, can also be computed by a width-3 ABP. Moreover, Saha et al. [SSS09] showed that PIT for depth-3 circuits reduces to PIT for width-2 ABP. Hence, constant width ABP is already a strong model. Our results are for constant width ABP with some natural restrictions.

An ABP is a read once ABP (ROABP) if the entries in the different matrices come from disjoint sets of variables. Forbes et al. [FSS14] recently gave a quasi-polynomial time blackbox test for ROABP, when the entries in each matrix are essentially *constant*-degree univariate polynomials. Their approach, too, involves low-support concentration. They interpret the dual of the transfer matrix as a rank extractor. In another work Jansen et al. [JQS10] gave quasi-polynomial time blackbox test for a sum of constantly many "ROABP". Their definition of ROABP is more stringent. They assume that every variable appears in at most one entry of a factor matrix.

Our result is for ROABP with a further restriction. We assume that all the matrices in the matrix product, except the left-most and the right-most matrices, are invertible. We give a blackbox test for this class of ROABP. In contrast to [FSS14], our test works in *polynomial time* if the dimension of the matrices is constant; moreover, we can handle univariate factor matrices with any degree.

Note that the class of ABP, where the factor matrices are invertible, is quite powerful, as Ben-Or and Cleve [BOC92] actually reduce formulas to width-3 ABP with *invertible* factors. Saha, Saptharishi and Saxena [SSS09] reduce depth-3 circuits to width-2 ABP with invertible factors. But the constraints of invertibility and read-once together seem to restrict the computing power of ABP. Interestingly, an analogous class of read-once boolean branching programs called *permutation branching programs* has been studied recently [KNP11, De11, Ste12]. These works give pseudorandom generators for this class (for constant width) with seed-length $O(\log n)$. In other words, they

give polynomial size sample set which can fool these programs. Our polynomial size hitting sets for the arithmetic setting is analogous to this result.

Theorem 3 (Informal version). *Let $C(\overline{x}) = D_0^T(\prod_{i=1}^{d} D_i)D_{d+1}$ be a polynomial such that $D_0 \in \mathbb{F}^w[x_{j_0}]$ and $D_{d+1} \in \mathbb{F}^w[x_{j_{d+1}}]$ and for all $i \in [d]$, $D_i \in \mathbb{F}^{w\times w}[x_{j_i}]$ is an invertible matrix (order of the variables is unknown). Let the degree bound on D_i be δ for $0 \leq i \leq d+1$. Then there is a $\mathrm{poly}((\delta n)^{w^2})$-time hitting-set for $C(\overline{x})$.*

The proof technique here is very different from the first two theorems (now, we show rank concentration over a *non-commutative* algebra). Our algorithm works even when the factor matrices have their entries as general sparse polynomials (still over disjoint sets of variables) instead of univariate polynomials (see detailed version in Section 5). Running time in this case is quasi-polynomial.

If the matrices are 2×2, we do not need the assumption of invertibility (see Theorem 4, Appendix E). So, for width-2 ROABP our results are strictly stronger than [FSS14]. Here again, there is a comparable result in the boolean setting. Pseudorandom generators with $O(\log n)$ seed-length (polynomial size sample set) are known for width-2 boolean branching programs [BDVY13].

1.1 Main idea of Theorem 1

As mentioned earlier, the basic idea is to show low-support concentration in $D(\overline{x})$, but by an efficient shift (Lemma 21). While showing low-support concentration in set-multilinear case, the key idea of [ASS13] was to identify low degree subcircuits of $D(\overline{x})$ which have 'true coefficients', i.e. each of these subcircuits is such that, when multiplied by an appropriate constant vector, its coefficients become the coefficients of $D(\overline{x})$. Then they show that an efficient shift can ensure low-support concentration in these subcircuits. The final step is to argue that low-support concentration in the subcircuits translates to low-support concentration in the actual circuit.

Such *true* subcircuits, in a general multilinear circuit, may have a high degree. In the case of small distance circuits, there exist true subcircuits with low degree in the last partition. But they may have many high degree monomials in the upper (other) partitions. That is not good, because the degree of the subcircuit affects the hitting-set size.

This inspired us to prove concentration in phases, i.e. by induction on k. We divide the coefficients into k phases (one corresponding to each partition). Phase-j coefficients are those which have nonzero values only in the coordinates $\{1, 2, \ldots, j\}$. We show concentration in successive phases (Lemma 16). In each phase, the concentration in the previous phases is assumed.

How do we isolate phase-j coefficients? We take an appropriate partial derivative of the circuits, which ensures that the values in the coordinates $\{j+1, j+2, \ldots, k\}$ are zero. Since the partial derivatives add to the complexity, they should be of small order (Observation 11).

Now, we identify subcircuits of this *derivative polynomial*, which have low degree in the j-th coordinate. The high degree in other coordinates does not matter as we assume that there is already concentration in the previous phases. We show low-support concentration in these subcircuits (Lemma 19; it is the most technical part of the proof). This in turn implies low-support concentration in the derivative polynomial (Lemma 20) by another induction on the *neighborhoods* (defined in Section 3.1). To show low-support concentration in a subcircuit we need to show some combinatorial properties of the *transfer matrix* (Lemma 26). Finally, the concentration in the derivative polynomial implies concentration among the phase-j coefficients of $D(\overline{x})$ (Lemma 18).

Carrying the argument from phase-1 to phase-k finishes the proof. The cost of this is only quasipoly in k (because the dimension of the underlying vector space is k and there is an implicit 'doubling effect' in Lemma 26), but *exponential* in δ (because the true subcircuits have degree δ, which makes the Kronecker map expensive). Any improvement in the latter would lead to the first nontrivial PIT for multilinear depth-3 circuits.

2 Preliminaries

$[n]$ denotes the indices 1 to n. Let $2^{[n]}$ denote the set of all subsets of $[n]$. $\mathrm{Part}(S)$ denotes the set of all possible partitions of the set S. Elements in a partition are called *colors*. The *support* of a monomial is the set of variables that have degree ≥ 1 in that monomial. The *support size* of the monomial is the cardinality of its support. $\mathbb{F}^{m\times n}$ represents the set of all $m \times n$ matrices over the

field $\mathbb{F}$. $\mathbb{F}^{S\times T}$, where S and T are sets, represents the set of all $|S| \times |T|$ matrices over the field $\mathbb{F}$, indexed by the elements of S and T. The matrices in this paper are often indexed by subsets of $[n]$.

A multilinear polynomial can be represented as $\sum_{S\subseteq[n]} a_S x_S$, where x_S is the monomial $\prod_{i\in S} x_i$. We will sometimes use the notation S for the multilinear monomial x_S.

Henceforth, we will only discuss polynomials computed by depth-3 multilinear circuits, unless explicitly stated otherwise. The polynomial computed by a depth-3 circuit, $C(\overline{x}) = \sum_{i=1}^{k} a_i \prod_{j=1}^{n_i} \ell_{ij}$ can also be written as the inner product of the vector $\overline{a} = (a_1, a_2, \ldots, a_k)^T$ and $D(\overline{x})$, where $D(\overline{x})$ is a polynomial over the k-dimensional Hadamard algebra $\mathbb{H}_k(\mathbb{F})$. The Hadamard algebra $\mathbb{H}_k(\mathbb{F})$ is defined as $(\mathbb{F}^k, +, \star)$, where $+$ and $\star$ are coordinate-wise addition and multiplication. The i-th coordinate of $D(\overline{x})$ is $\prod_{j=1}^{n_i} \ell_{ij}$. Hence, $C(\overline{x}) = \overline{a}^T D(\overline{x})$. Let $\text{coef}_D(x_S)$ denote the coefficient of the monomial x_S in the polynomial $D(\overline{x})$. The coefficients $\{\text{coef}_D(x_S) \mid S \subseteq [n]\}$ of $D(\overline{x})$ form a $(\leq k)$-dimensional vector space over the base field $\mathbb{F}$. Our rough plan is to show that this vector space is spanned by the coefficients of the 'low-degree' monomials. We thus define *ℓ-concentration* for a polynomial $D(\overline{x})$ whose coefficients are vectors from a k-dimensional vector space.

Definition 1 (ℓ-concentration). *Polynomial $D(\overline{x}) \in \mathbb{H}_k(\mathbb{F})[x_1, x_2, \ldots, x_n]$ is ℓ-concentrated if* $\text{rk}_{\mathbb{F}}\{\text{coef}_D(x_S) \mid S \subseteq [n], |S| < \ell\} = \text{rk}_{\mathbb{F}}\{\text{coef}_D(x_S) \mid S \subseteq [n]\}$.

The following Lemma from [ASS13] says that a polynomial $C(\overline{x})$, with an ℓ-concentrated polynomial $D(\overline{x})$, has a hitting set (for a proof, see Section A).

Lemma 2. *If $D(\overline{x})$ is ℓ-concentrated, then there is a $n^{O(\ell)}$-time hitting-set for $C(\overline{x})$.*

However, observe that low-support concentration does not exist in all polynomials $D(\overline{x})$. E.g. in the polynomial $D(\overline{x}) = \overline{c} \cdot x_1 x_2 \ldots x_n$, there are no low-support monomials. To counter this problem, the polynomial is *shifted*. Each input x_i to the polynomial is replaced with $x_i + t_i$, where t_is are symbolic constants adjoined to the base field $\mathbb{F}$. Now, the input field is considered to be the field of fractions $\mathbb{F}(\overline{t})$, where $\overline{t} = \{t_1, t_2, \ldots, t_n\}$. Since after shifting, the coefficients of high-support monomials contribute an additive term to the coefficients of the low-support monomials, we can hope to prove ℓ-concentration over this field of fractions. We will use the notation $D'(\overline{x})$ as well as $D(\overline{x} + \overline{t})$ to mean $D(x_1 + t_1, x_2 + t_2, \ldots, x_n + t_n)$. For the example above, $D'(\overline{x}) = \sum_{S\subseteq[n]} \overline{c} \cdot t_{\bar{S}} x_S$. The dependence of $\text{coef}_D(x_S)$ over the field $\mathbb{F}(\overline{t})$ is given by $\text{coef}_{D'}(x_S) = \overline{c} \cdot t_{\bar{S}} = t_{\bar{S}} \,\text{coef}_{D'}(x_\varnothing)$. Thus, in the above example, $D'(\overline{x})$ is ℓ-concentrated for $\ell = 1$.

We conjecture that $O(\log n + \log k)$-concentration can be proven for all multilinear circuits after an appropriate shift. Agrawal, Saha & Saxena ([ASS13]) have proven $O(\log k)$-concentration of set-multilinear circuits after an efficient shift. Here, we study low-support concentration for more general models, by developing stronger techniques.

2.1 General Approach

How do we prove that the high-support coefficients of D' are dependent on the low-support coefficients? Do we even know that a given high-support coefficient is dependent on other coefficients? The reason we believe such a dependency exists is because of shifting. A monomial x_S, when the circuit is shifted, contributes its coefficient, denoted by u_S, to the coefficients of all of its subsets: $u'_T := \text{coef}_{D'}(x_T) = \sum_{S\supseteq T} u_S t_{S\setminus T}$. The polynomial $D'(\overline{x} - \overline{t}) = D(\overline{x})$, i.e. by shifting every variable x_i by $-t_i$ in $D'(\overline{x})$, we get back $D(\overline{x})$. Thus, $u_T = \sum_{S\supseteq T} u'_S t_{S\setminus T} (-1)^{|S\setminus T|}$. This can be represented as $U = U' \cdot \mathcal{M}$, where $U \in \mathbb{F}^{[k]\times 2^{[n]}}$ and $U' \in (\mathbb{F}(\overline{t}))^{[k]\times 2^{[n]}}$ represent the coefficients in the polynomials $D(\overline{x})$ and $D'(\overline{x})$ respectively. The matrix $\mathcal{M}$ has (S,T)-th entry

$$\mathcal{M}(S,T) = \begin{cases} t_{S\setminus T}(-1)^{|S\setminus T|} & \text{if } T \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently, we can write $\mathcal{M} = A^{-1}MA$, where $M(S,T) = 1$ if $T \subseteq S$ and 0 otherwise, and A is a diagonal $2^{[n]} \times 2^{[n]}$ matrix where the (T,T)-th entry is $(-1)^{|T|} t_T^{-1}$.

To analyze the dependencies among vectors in U' we take a dependency for the vectors in U and *lift* it. Suppose a dependency for the U vectors is: $\sum_T \alpha_T u_T = 0, \alpha_T \in \mathbb{F}$. Now, replace the coefficients u_T with an equivalent expression in terms of the coefficients u'_S. We get $\sum_T \alpha_T \sum_{S \supseteq T} u'_S t_{S\setminus T}(-1)^{|S\setminus T|} = 0$. In the dependency, coef(u'_S) is $\sum_{T \subseteq S} \alpha_T t_{S\setminus T}(-1)^{|S\setminus T|}$. The coefficient for each u'_S is nonzero, and thus, it participates non-trivially in a dependency.

There are two problems here. First, we do not directly get dependencies which show that high-support coefficients are in the span of low-support coefficients. There must be one such dependency for each high-support u'_S, which shows that it is in the span of low-support coefficients. Existence of such dependencies will be shown in the later sections, by considering all the dependencies of the U vectors (null vectors of U) and their lifts.

The other problem is that, even if low-support concentration does exist, we are substituting $(x_i + t_i)$s instead of x_is and then computing low-support coefficients (Lemma 2). The coefficients themselves can be exponentially large polynomials in the variables $\bar{t}$, and thus cannot be computed efficiently. It turns out that we can substitute the shift variables with an efficient univariate map and show low-support concentration.

2.2 Kronecker substitution

The shift variables t_is are replaced with powers of a single variable t. Let us say, the degree of the variable t is upper bounded by some function $g(n)$. Then the t-degree of the polynomial computed by the depth-3 multilinear circuit is bounded by $ng(n)$. The computation is thus efficient when the shift is univariate and 'small'. The univariate map we use will need to separate all $(\leq \ell)$ support monomials for some small ℓ, i.e. all $(\leq \ell)$ support monomials should be mapped to distinct powers of t (the map can be seen as acting on monomials in the natural way e.g. $\phi(t_1 t_2) = \phi(t_1)\phi(t_2)$). This map will be denoted by ϕ_ℓ. For the time complexity of generating such maps, see Lemma 25 (Appendix A). We now describe the effect of the shifting map on the final time complexity. The notation $D(\overline{x} + \phi(\bar{t}))$ will mean $D(x_1 + \phi(t_1), x_2 + \phi(t_2), \ldots, x_n + \phi(t_n))$.

Lemma 3 (ℓ-Concentration to hitting-sets). *If for a polynomial $D(\overline{x}) \in \mathbb{H}_k(\mathbb{F})[\overline{x}]$, we can construct a set of $f(n)$-many maps from $\bar{t}$ to $\{t^i\}_{i=1}^{g(n)}$ such that for at least one of the maps ϕ, the shifted polynomial $D(\overline{x} + \phi(\bar{t}))$ has ℓ-concentration, then $C(\overline{x}) = \overline{a}^T D(\overline{x})$, for any $\overline{a} \in \mathbb{F}^k$, has an $n^{O(\ell)} f(n)g(n)$-time hitting-set.* (Proof in Appendix A.)

Now that we are clear about the general technique, we can proceed to proving ℓ-concentration in some interesting models.

3 Low-distance multilinear depth-3 circuits: Theorem 1

The main model for which we study low-support concentration is depth-3 multilinear circuits with 'small distance'.

3.1 δ-distance circuits

Each product gate in a depth-3 multilinear circuit induces a partition on the variables. Let these partitions be $\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_k$.

Definition 4 (Distance for a partition sequence, $\mathrm{d}(\mathbb{P}_1, \ldots, \mathbb{P}_k)$). *Let $\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_k \in \mathrm{Part}([n])$ be the k partitions of the variables $\{x_1, x_2, \ldots, x_n\}$. Then $\mathrm{d}(\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_k) =: \delta$ if $\forall i \in \{2, 3, \ldots, k\}$, $\forall$colors $Y_1 \in \mathbb{P}_i$, $\exists Y_2, Y_3, \ldots, Y_{\delta'} \in \mathbb{P}_i$ $(\delta' \leq \delta)$ such that $Y_1 \cup Y_2 \cup \cdots \cup Y_{\delta'}$ equals a union of some colors in $\mathbb{P}_j, \forall j \in [i-1]$.*

In other words, in every partition $\mathbb{P}_i$, each color Y_1 has a set of colors called 'friendly neighborhood', $\{Y_1, Y_2, \ldots, Y_{\delta'}\}$, consisting of at most δ colors, which is exactly partitioned in the 'upper partitions'. We call $\mathbb{P}_i$, an *upper* partition relative to $\mathbb{P}_j$ (and $\mathbb{P}_j$, a *lower* partition relative to $\mathbb{P}_i$), if $i < j$. For a color X_a of a partition $\mathbb{P}_j$, let $\mathrm{nbd}_j(X_a)$ denote its friendly neighborhood. The friendly neighborhood $\mathrm{nbd}_j(x_i)$ of a variable x_i in a partition $\mathbb{P}_j$ is defined as $\mathrm{nbd}_j(\mathrm{color}_j(x_i))$, where $\mathrm{color}_j(x_i)$ is the color in the partition $\mathbb{P}_j$ that contains the variable x_i. The friendly neighborhood $\mathrm{nbd}_j(\{x_i\}_{i \in \mathcal{I}})$ of a set of variables $\{x_i\}_{i \in \mathcal{I}}$ in a partition $\mathbb{P}_j$ is given by $\bigcup_{i \in \mathcal{I}} \mathrm{nbd}_j(x_i)$.

Definition 5 (δ-distance circuits). *A multilinear depth-3 circuit C has δ-distance if its product gates can be ordered to correspond to a partition sequence $(\mathbb{P}_1, \ldots, \mathbb{P}_k)$ with $\mathrm{d}(\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_k) \leq \delta$.*

The corresponding $\Pi\Sigma$ circuit $D(\overline{x})$ over $\mathbb{H}_k(\mathbb{F})$ is also said to have δ-distance.

Every depth-3 multilinear circuit is thus an n-distance circuit. A circuit with a partition sequence, where the partition $\mathbb{P}_i$ is a refinement of the partition $\mathbb{P}_{i+1}, \forall i \in [k-1]$, exactly characterizes a 1-distance circuit. All depth-3 multilinear circuits have distance between 1 and n. Also observe that the circuits with 1-distance subsume set-multilinear circuits.

Friendly neighborhoods - To get a better picture, we ask: Given a color X_a of a partition $\mathbb{P}_j$ in a circuit $D(\overline{x})$, how do we find its friendly neighborhood $\mathrm{nbd}_j(X_a)$? Consider a graph G_j which has the colors of the partitions $\{\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_j\}$, as its vertices. For all $i \in [j-1]$, there is an edge between the colors $X \in \mathbb{P}_i$ and $Y \in \mathbb{P}_j$ if they share at least one variable. Observe that if any two colors X_a and X_b of partition $\mathbb{P}_j$ are reachable from each other in G_j, then, they should be in the same neighborhood. As reachability is an equivalence relation, *the neighborhoods are equivalence classes of colors.*

Moreover, observe that for any two variables x_a and x_b, if their respective colors in partition $\mathbb{P}_j$, $\mathrm{color}_j(x_a)$ and $\mathrm{color}_j(x_b)$ are reachable from each other in G_j then their respective colors in partition $\mathbb{P}_{j+1}$, $\mathrm{color}_{j+1}(x_a)$ and $\mathrm{color}_{j+1}(x_b)$ are also reachable from each other in G_{j+1}. Hence,

Observation 6 *If at some partition, the variables x_a and x_b are in the same neighborhood, then, they will be in the same neighborhood in all of the lower partitions. I.e.* $\mathrm{nbd}_j(x_a) = \mathrm{nbd}_j(x_b) \implies \mathrm{nbd}_i(x_a) = \mathrm{nbd}_i(x_b), \forall i \geq j$.

In other words, at the level of the variables, the neighborhoods in the upper partitions are *refinements* of the neighborhoods in the lower partitions.

We now claim that any subcircuit of a δ-distance circuit $D(\overline{x})$, also has δ-distance.

Observation 7 (Subcircuit of $D(\overline{x})$) *Let $E \in \mathbb{H}_j(\mathbb{F})[\overline{x}]$ be a subcircuit of a δ-distance circuit $D(\overline{x})$, obtained by replacing an arbitrary set of linear factors in each coordinate of $D(\overline{x})$ with 1, and restricting the circuit to the coordinates 1 to j. Then E is also a δ-distance circuit.*

Proof. A linear factor $b_{i_0} + \sum_r b_{i_r} x_{i_r}$ which is replaced with 1, can be viewed as $1 + \sum_r 0 \cdot x_{i_r}$. Such a subcircuit induces the same partition sequence as circuit D. When we restrict the circuit to the coordinates 1 to j, we get a subsequence of this partition sequence. Clearly, the subsequence also has δ-distance.

Since there exists a set of colors (linear factors in the circuit C) in $\mathbb{P}_i$ ($i \in [j]$) that exactly contain the variables of one neighborhood, $\mathrm{nbd}_j(X_a)$, we can define the following subcircuit of $D(\overline{x})$.

Definition 8 ($\mathrm{tower}_j(\mathcal{X})$). *For a neighborhood $\mathcal{X}$ in partition $\mathbb{P}_j$, we define a $\mathrm{tower}_j(\mathcal{X})$ as a polynomial over $\mathbb{H}_j(\mathbb{F})$, such that its i-th coordinate ($i \leq j$) is the product of exactly those linear factors (in the i-th product gate of the circuit C) that contain the variables of the neighborhood $\mathcal{X}$.*

We can define a tower over a union of neighborhoods $(\cup_{i=1}^r \mathcal{X}_i)$ in partition $\mathbb{P}_j$ as $\mathrm{tower}_j(\cup_{i=1}^r \mathcal{X}_i) := \mathrm{tower}_j(\mathcal{X}_1) \star \mathrm{tower}_j(\mathcal{X}_2) \star \cdots \star \mathrm{tower}_j(\mathcal{X}_r)$. For any such tower $E = \mathrm{tower}_j(\cup_{i=1}^r \mathcal{X}_i)$, $\mathrm{nbd}_j(E)$ will denote the union of neighborhoods $(\cup_{i=1}^r \mathcal{X}_i)$. For a set of variables S, $\mathrm{tower}_j(S)$ is the tower over $\mathrm{nbd}_j(S)$. For a neighborhood $\mathcal{X}$ in partition $\mathbb{P}_j$, the variables in $\mathrm{tower}_j(\mathcal{X})$ are the variables in $\mathcal{X}$, denoted by both $\mathrm{var}(\mathrm{tower}_j(\mathcal{X}))$ and $\mathrm{var}(\mathcal{X})$. Observe that the towers over any two neighborhoods in partition $\mathbb{P}_j$ are polynomials over a disjoint set of variables.

The following observation says that the coefficient of a monomial in a product of towers is equal to the product of its 'support coefficients' in the individual towers.

Observation 9 *Let R and T be two sets of variables coming from two* disjoint sets *of neighborhoods of partition $\mathbb{P}_j$. Then, the coefficients of monomial $S \subseteq R \cup T$ in $\mathrm{tower}_j(R) \star \mathrm{tower}_j(T)$ is given by $\mathrm{coef}_{\mathrm{tower}_j(R) \star \mathrm{tower}_j(T)}(S) = \mathrm{coef}_{\mathrm{tower}_j(R)}(S \cap R) \star \mathrm{coef}_{\mathrm{tower}_j(T)}(S \cap T)$.*

3.2 True coefficients

We will be proving low-support concentration in some special subcircuits $E'(\overline{x})$ (see Observation 7) of $D'(\overline{x})$ (the shifted polynomial). This would eventually prove low-support concentration of $D'(\overline{x})$. The coefficient $\text{coef}_{E'}(x_S)$ of the monomial x_S in the subcircuit $E'(\overline{x})$ is given by $\text{coef}_{D'}(x_S) = \text{const}_{E'} \star \text{coef}_{E'}(x_S)$, where $\text{const}_{E'}$ is the product of the constant parts of all linear factors not in $E'(\overline{x})$. Thus, it is enough to prove ℓ-concentration within $E'(\overline{x})$ since any dependency in $E'(\overline{x})$ translates to a dependency in $D'(\overline{x})$ by multiplying throughout with a constant.

All the monomials which participate in a dependency, and hence, all the monomials occurring in $E'(\overline{x})$ should have *true coefficients*. I.e. each monomial should have a coefficient in $E'(\overline{x})$ similar to that in $D'(\overline{x})$. I.e. $\text{const}_{E'} \star \text{coef}_{E'}(x_T) = \mathbf{0}$ or $\text{coef}_{D'}(x_T)$, for all monomials T.

In the next subsection, we will study a few properties of the subcircuit $E'(\overline{x})$, for which we prove low-support concentration.

3.3 Phases

Since true coefficients have to be used, $E'(\overline{x})$ has to be (roughly) a tower over a (small) set of neighborhoods. Thus, though the lowest partition in $E'(\overline{x})$ will have a few linear factors, the higher partitions may have $O(n)$ linear factors. This blows up the support size of the monomials in $E'(\overline{x})$. We intend to show (by induction) that these coefficients of the high-support monomials are ℓ-concentrated. Note that these high-support monomials come only from the upper partitions.

We use the refinement property of neighborhoods (Observation 6) to categorize the monomials of the polynomial $D(\overline{x})$ into k phases. Roughly speaking, the *phase* of a monomial S is the lowest partition from which S can possibly be generated. If $(\delta + 1)$ variables of S belong to the same neighborhood of partition $\mathbb{P}_{j+1}$, then S cannot be generated from this partition, or the partitions below it, as we will soon see. This motivates us to define the phase of a monomial as:

Definition 10 (Phase-j). *A monomial S is in phase-j if the partition $\mathbb{P}_j$ is the lowest partition with $\leq \delta$ of its support variables in each of its neighborhoods.*

A phase-j monomial can be characterized by the presence of a *pivot tuple*.

Observation 11 (Pivot tuple) *Let $j \in [k-1]$. A monomial S is in the j-th* phase *iff*

1. *There is no neighborhood of the upper partitions $\{\mathbb{P}_i\}_{i=1}^{j}$ that contributes more than δ variables to the monomial S, and,*
2. *there exist $(\delta+1)$ variables in its support (called a* pivot tuple*) that are in the same neighborhood of the partition $\mathbb{P}_{j+1}$. This pivot tuple is in the same neighborhood of all the lower partitions $\{\mathbb{P}_i\}_{i=j+1}^{k}$.*

A monomial is in the k-th phase if it is not in any of the previous phases. (Proof is by applying Observation 6 on Definition 10.)

Observation 12 *A monomial S belongs to exactly one of the k phases, denoted by* $\text{phase}(S)$.

Observation 13 *A monomial S may have more than one pivot tuples. Observe that* $\text{phase}(R) = \text{phase}(S)$*, for all pivots R of a monomial S.*

Since every superset of a monomial S contains its pivot tuple, $\text{pivot}(S)$,

Observation 14 *For all supersets $T \supseteq S$,* $\text{phase}(T) \leq \text{phase}(S)$.

In a δ-distance circuit, a neighborhood has at most δ colors. Hence, for any phase-j monomial, at least two of its $(\delta + 1)$ pivot variables, have the same color in the lower partitions $\{\mathbb{P}_i\}_{i=j+1}^{k}$. They come from the same linear factor of the circuit C. Thus,

Observation 15 *If a monomial S is in the j-th phase, then, its coefficient vector is zero in the coordinates $\{j+1, j+2, \ldots, k\}$.*

Since the coordinates $\{j+1, j+2, \ldots, k\}$ in the coefficients of phase-($\leq j$) monomials are 0, they can be ignored when the linear dependencies among the coefficients of phase-($\leq j$) monomials are studied. For phase-j, we discuss the linear dependencies among coefficients of phase-($\leq j$) monomials, i.e. vectors limited to the ($\leq j$)-th coordinates. Coefficients of phase-j monomials will be called phase-j coefficients.

3.4 Partial derivatives

All the monomials in phase-j are identified by the existence of at least one pivot tuple. Consider a phase-j monomial S with a pivot R of phase-j (Observation 13). We will prove ℓ-concentration (linear dependence on ($\leq \ell$)-support coefficients) among the coefficients of all the monomials that include R. Since all the monomials now include R, the ($> j$)-th coordinates in the polynomial $D'(\overline{x})$ are 0 (Observations 13, 14, 15). The polynomial could be restricted to coordinates 1 to j. We prove ℓ-concentration after taking *partial derivative* with respect to x_R. This is encapsulated in Observation 17.

In other words: In the linear factors of C corresponding to the partitions $\{\mathbb{P}_i\}_{i=1}^{j}$ that include the variables of R, we pick the monomial x_R and its coefficient. Then, we prove low-support concentration in the remaining polynomial.

3.5 ℓ-concentration

We now come to the main subsection of the paper. Let $\ell_0 = \log_2(k+1), \ell = \delta + 1 + \ell_0$ and $D'(\overline{x}) = D(\overline{x} + \phi_{\delta\ell_0}(\overline{t}))$, where $D(\overline{x}) \in \mathbb{H}_k(\mathbb{F})[\overline{x}]$ is a δ-distance circuit and $\phi_{\delta\ell_0}$ (defined in Subsection 2.2) is a univariate map which separates all ($\leq \delta\ell_0$) support monomials. We prove ℓ-concentration in the polynomial $D'(\overline{x})$ (Lemma 21), because of which, a $n^{O(\delta \log k)}$ time hitting-set exists for the δ-distance depth-3 multilinear circuit (Theorem 1).

ℓ-concentration in each phase. As already stated in Subsection 3.3, proof of ℓ-concentration in $D'(\overline{x})$ is by induction on the phases. The monomials of each phase are progressively shown to be ℓ-concentrated: In this subsection, when we say that a phase-i monomial is ℓ-concentrated, we mean that its coefficient is linearly dependent on coefficients of ($< \ell$)-support monomials of phase-($\leq i$). The following lemma shows one step of the induction.

Lemma 16 (Phase-($< j$) to Phase-j). *If the monomials in $D'(\overline{x})$ of phase-($< j$) are ℓ-concentrated, then the monomials of phase-j are ℓ-concentrated. ℓ-Concentration exists in phase-1 without any preconditions.*

Proof. Since phase-0 does not exist, the monomials of phase-0 are vacuously ℓ-concentrated. Now, when the monomials of phase-($< j$) are ℓ-concentrated, Lemma 18 proves ℓ-concentration in all phase-j monomials that include the phase-j pivot R. Every monomial of phase-j has such a pivot from Observation 11.

The following observation about $D'(\overline{x})$ will be used in the next lemma.

Observation 17 (D'_R, E'_R and Decomposition of coefficients) *For any $R \subseteq [n]$, we define $D'_R(\overline{x}) \in \mathbb{H}_k(\mathbb{F}(t))[\overline{x}]$ to be the polynomial, obtained from $D'(\overline{x})$, where each coordinate is the product of those linear factors, in the corresponding coordinate of $D'(\overline{x})$, that include the variables of R. Let $E'_R(\overline{x}) \in \mathbb{H}_k(\mathbb{F}(t))[\overline{x}]$ be the polynomial, obtained from $D'(\overline{x})$, where each coordinate is the product of those linear factors, in the corresponding coordinate of $D'(\overline{x})$, that do not include the variables of R. I.e. $D'(\overline{x}) = D'_R(\overline{x}) \star E'_R(\overline{x})$. Then, for all supersets S of R, $\text{coef}_{D'}(x_S) = \text{coef}_{D'_R}(x_R) \star \text{coef}_{E'_R}(x_{S\setminus R})$.*

Proof. The observation is easy for the coordinates where the linear factors that support R and $S \setminus R$ are disjoint. In the coordinates where at least one variable, say x_a, of $S \setminus R$ is included in the linear factors in $D'_R(\overline{x})$, $\text{coef}_{E'_R}(x_{S\setminus R})$ is zero in these coordinates because the variable $x_a \in S \setminus R$ is present in only one linear factor, which is not in E'_R. Whereas $\text{coef}_{D'}(x_S)$ is zero in these coordinates because at least two variables of S are present in the same linear factor in D'_R.

Lemma 18 (Concentration in monomials that include R). *Let R be a phase-j monomial such that $|R| = \delta + 1$. If ℓ-concentration at phase-$(< j)$ exists, then the monomials that include R are ℓ-concentrated.*

Proof. We show that the coefficients, whose support include R, are linearly dependent on: The low-support coefficients, whose support include R, and phase-$(< j)$ coefficients.

Consider the polynomial $E'_R(\overline{x})$ of $D'(\overline{x})$, restricted to coordinates 1 to j. Let S be an arbitrary monomial that includes R. Let $S' = S \setminus R$. Then, S' is a monomial in $E'_R(\overline{x})$. Lemma 20, together with Observation 7, proves ℓ_0-concentration for the monomials in $E'_R(\overline{x})$ modulo the coefficients in phase-$(< j)$. I.e.

$$\operatorname{coef}_{E'_R}(x_{S'}) = \sum_{\substack{T':|T'|<\ell_0,\\ T'\in\text{phase-}j \text{ of } E'_R}} \alpha_{T'} \operatorname{coef}_{E'_R}(x_{T'}) + \sum_{T'\in\text{phase-}(<j) \text{ of } E'_R} \alpha_{T'} \operatorname{coef}_{E'_R}(x_{T'}),$$

where $\alpha_T \in \mathbb{F}(t)$ and the monomials are from $E'_R(\overline{x})$. Since the supports of R and S' are disjoint, from Observation 17, $\operatorname{coef}_{D'}(x_{S'\cup R}) = \operatorname{coef}_{D'_R}(x_R) \star \operatorname{coef}_{E'_R}(x_{S'})$, $\forall$ monomial S' in $E'_R(\overline{x})$.

Therefore, multiplying throughout with $\operatorname{coef}_{D'_R}(x_R)$,

$$\operatorname{coef}_{D'}(x_{S'\cup R}) = \sum_{\substack{T':|T'|<\ell_0,\\ T'\in\text{phase-}j \text{ of } E'_R}} \alpha_{T'} \operatorname{coef}_{D'}(x_{T'\cup R}) + \sum_{T'\in\text{phase-}(<j) \text{ of } E'_R} \alpha_{T'} \operatorname{coef}_{D'}(x_{T'\cup R}).$$

$S = S' \cup R$ and let $T = T' \cup R$. From Observation 14, even after including R, a phase-$(< j)$ monomial remains a phase-$(< j)$ monomial and a phase-j monomial remains a phase-$(\leq j)$ monomial. $|R| = \delta + 1$. Rewriting gives,

$$\operatorname{coef}_{D'}(x_S) = \sum_{\substack{T:|T|<\ell_0+\delta+1,\\ T\in\text{phase-}(\leq j) \text{ of } D'}} \alpha_{T'} \operatorname{coef}_{D'}(x_T) + \sum_{T\in\text{phase-}(<j) \text{ of } D'} \alpha_{T'} \operatorname{coef}_{D'}(x_T).$$

$\operatorname{coef}_{D'}(x_S)$ is thus linearly dependent on: $(< \ell)$ support coefficients, and the coefficients in the upper phases. The upper phase monomials are already ℓ-concentrated. Hence, $\operatorname{coef}_{D'}(x_S)$ is also ℓ-concentrated.

ℓ_0-concentration in the "last" phase. Lemma 18 needed ℓ_0-concentration of E'_R modulo its phase-$(< j)$ coefficients. Similar to the way E'_R was defined from D' in Observation 17, we can define the polynomial E_R from D. It is easy to see that $E'_R = E_R(\overline{x} + \phi_{\delta\ell_0}(\bar{t}))$. By Observation 7, E_R is also a δ-distance circuit. We will actually give a general result that for any δ-distance circuit $E(\overline{x}) \in \mathbb{H}_j[\overline{x}]$, its shifted version $E' := E(\overline{x} + \phi_{\delta\ell_0}(\bar{t}))$ has ℓ_0-concentration modulo its phase-$(< j)$ coefficients. Here, phases for coefficients, neighborhoods in the partitions and towers over the neighborhoods, are defined for E', similar to D'. Note that j is the last phase of E', as it is over $\mathbb{H}_j(\mathbb{F}(t))$. For any $E'(\overline{x}) \in \mathbb{H}_j(\mathbb{F}(t))[\overline{x}]$, let $V_{E'}$ denote the space spanned by the phase-$(< j)$ coefficients of E'.

The proof is in two parts. Let $\|\text{nbd}_j(E')\|$ be the number of neighborhoods in E' in its partition $\mathbb{P}_j$. The first part is to show the result for smaller circuits E', when $\|\text{nbd}_j(E')\| \leq \ell_0$. This is done in Lemma 19. The second part is to prove it for larger circuits using induction. This is done in Lemma 20.

Lemma 19 (Small-neighborhood ℓ_0-concentration). *Let $E(\overline{x}) \in \mathbb{H}_j(\mathbb{F})[\overline{x}]$ be a δ-distance circuit such that $\|\text{nbd}_j(E')\| \leq \ell_0$. Then $E'(\overline{x}) = E(\overline{x} + \phi_{\delta\ell_0}(\bar{t}))$ has ℓ_0-concentration modulo $V_{E'}$.* (Proof in Appendix B.)

Lemma 20 (Subcircuit-ℓ_0-concentration). *Let $F(\overline{x}) \in \mathbb{H}_j(\mathbb{F})[\overline{x}]$ be a δ-distance circuit. Then, $F'(\overline{x}) = F(\overline{x} + \phi_{\delta\ell_0}(\bar{t}))$ has ℓ_0-concentration modulo $V_{F'}$.* (Proof in Appendix A.)

Proving Theorem 1 From Lemma 16, we directly get Lemma 21:

Lemma 21. *If $D \in \mathbb{H}_k(\mathbb{F})[\overline{x}]$ is a δ-distance circuit, then D' has ℓ-concentration.*

We now prove the main theorem. If ℓ-concentration exists, then a hitting-set exists.

Proof (of Theorem 1). We can write $C(\overline{x}) = \overline{a}^T D(\overline{x})$ for some $\overline{a} \in \mathbb{F}^k$ and $D(\overline{x}) \in \mathbb{H}_k(\mathbb{F})[\overline{x}]$. From Lemma 21, $D(\overline{x} + \phi_{\delta\ell_0}(\overline{t}))$ has ℓ-concentration. Now, the map $\phi_{\delta\ell_0}$ separates all the monomials of support $\leq \delta\ell_0$. The number of such monomials is $n^{O(\delta\ell_0)}$. Hence from Lemma 25 (Appendix A), $\phi_{\delta\ell_0}$ can be generated by trying N-many monomial maps which have degree $\leq N \log N$, where $N := n^{O(\delta\ell_0)}$. Now, from Lemma 3 we directly get a $n^{O(\delta \log k)}$-time hitting-set.

4 Base sets with δ distance: Theorem 2

In this section we further generalize the class of polynomials, for which we can give an efficient test, beyond low-distance. Basically, it is enough to have low-distance "projections".

Definition 22. *A multilinear depth-3 circuit is said to have m-base-sets-δ-distance if there is a partition of the variable set $\overline{x}$ into base sets $\{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_m\}$ such that for any $i \in [m]$, restriction of C on the i-th base set, i.e. $C|_{(\overline{x}_j = \mathbf{0}\ \forall j \neq i)}$ has δ-distance.*

As discussed in the previous section, for a depth-3 circuit C we can write $C(\overline{x}) = \overline{a}^T \cdot D(\overline{x})$, where $\overline{a} \in \mathbb{H}_k(\mathbb{F})$ and $D(\overline{x}) \in \mathbb{H}_k(\mathbb{F})[\overline{x}]$. We say a polynomial $D(\overline{x}) \in \mathbb{H}_k(\mathbb{F})[\overline{x}]$ has m-base-sets-δ-distance if the corresponding circuit C has m-base-sets-δ-distance. We will show that such a polynomial $D(\overline{x})$ will also have some appropriately low-support concentration after an efficient shift (Lemma 31, Appendix C).

From Lemma 21, we know that if a polynomial $D(\overline{x}) \in \mathbb{H}_k(\mathbb{F})[\overline{x}]$ has δ-distance then $D(\overline{x} + \phi_{\delta\ell_0}(\overline{t}))$ has ℓ-concentration. The basic idea for a polynomial $D(\overline{x})$ having m-base-sets-δ-distance, is to use a different shift variable for each base set. Hence, it is necessary that the base sets are known. Except this knowledge, the test is blackbox. The basic argument is to view the polynomial $D(\overline{x})$ as a polynomial over the variable set $\overline{x}_m$ and whose coefficients lie in $\mathbb{H}_k(\mathbb{F})[\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}]$. From this perspective, it has δ-distance and hence we can achieve low-support concentration. Further, we argue low-support concentration in its coefficients, which themselves have $(m-1)$-base-sets-δ-distance (Lemma 30). The proof is by induction on the number of base sets.

Let $\overline{t}_i$ be the set of shift variables for $\overline{x}_i$ for any $i \in [m]$ ($\overline{t} = \overline{t}_1 \sqcup \overline{t}_2 \sqcup \cdots \sqcup \overline{t}_m$). Let I_i denote the set of indices corresponding to the variables in $\overline{x}_i$, for all $i \in [m]$. We define our shifting map $\phi^m_{\delta\ell_0} : \overline{t} \to \mathbb{F}[y_1, y_2, \ldots, y_m]$ as follows: $\forall i \in [m]$, each variable in the set $\overline{t}_i$ is mapped to a power of y_i, such that for any two sets $S, T \subseteq I_i$ with $|S|, |T| \leq \delta\ell_0$, $\phi^m_{\delta\ell_0}(t_S) \neq \phi^m_{\delta\ell_0}(t_T)$, for all $i \in [m]$.

Proof (of Theorem 2). Let $C(\overline{x}) = \overline{a}^T \cdot D(\overline{x})$ for some $\overline{a} \in \mathbb{H}_k(\mathbb{F})$ and $D(\overline{x}) \in \mathbb{H}_k(\mathbb{F})[\overline{x}]$. Lemma 31 (Appendix C) shows that $D(\overline{x} + \phi^m_{\delta\ell_0}(\overline{t}))$ has $(m(\ell-1)+1)$-concentration. Hence, from Lemma 2, $C(\overline{x} + \phi^m_{\delta\ell_0}(\overline{t}))$ has $n^{O(m\ell)}$-time hitting-set. Moreover, each evaluation of $C(\overline{x} + \phi^m_{\delta\ell_0}(\overline{t}))$ is a polynomial in $\{y_1, y_2, \ldots, y_m\}$. Let us say the individual degree bound on this polynomial is d. Then the time taken to compute this polynomial would be proportional to the number of monomials in it, i.e. $(d+1)^m$.

For the degree bound we must look at the map $\phi^m_{\delta\ell_0}$. The map $\phi^m_{\delta\ell_0}$ separates all $\delta\ell_0$ support monomials. There are $n^{O(\delta\ell_0)}$ such monomials. From Lemma 25 (Appendix A), we need to try N maps each with highest degree $N \log N$ to get the desired map, where $N = n^{O(\delta\ell_0)}$. Hence, the total complexity is $n^{O(m\ell)} N^{O(m)} = n^{O(m\delta \log k)}$.

5 Sparse-Invertible Width-w ROABP: Theorem 3

An ABP is a directed graph with $d+1$ layers of vertices $\{V_0, V_1, \ldots, V_d\}$ such that the edges are only going from V_{i-1} to V_i for any $i \in [d]$. As a convention, V_0 and V_d have only one node each, let the nodes be v_0 and v_d respectively. A width-w ABP has $|V_i| \leq w$ for all $i \in [d]$. Let the set of nodes in V_i be $\{v_{i,j} \mid j \in [w]\}$. All the edges in the graph have weights from $\mathbb{F}[\overline{x}]$, for some field

$\mathbb{F}$. For an edge e, let us denote its weight by $w(e)$. For a path p from v_0 to v_d, its weight $w(p)$ is defined to be the product of weights of all the edges in it, i.e. $\prod_{e\in p} w(e)$. Consider the polynomial $C(\overline{x}) = \sum_{p\in \text{paths}(v_0,v_d)} w(p)$ which is the sum of the weights of all the paths from v_0 to v_d. This polynomial $C(\overline{x})$ is said to be computed by the ABP.

It is easy to see that this polynomial is the same as $D_0^T(\prod_{i=1}^{d-2} D_i)D_{d-1}$, where $D_0, D_{d-1} \in (\mathbb{F}[\overline{x}])^w$ and D_i for $1 \le i \le d-2$ is a $w \times w$ matrix such that

$$\begin{aligned} D_0(\ell) &= w(v_0, v_{1,\ell}) \text{ for } 1 \le \ell \le w \\ D_i(k,\ell) &= w(v_{i,k}, v_{i+1,\ell}) \text{ for } 1 \le \ell, k \le w \text{ and } 1 \le i \le d-2 \\ D_{d-1}(k) &= w(v_{d-1,k}, v_d) \text{ for } 1 \le k \le w \end{aligned}$$

An ABP is called a read once ABP (ROABP) if the edge weights in the different layers are polynomials in disjoint sets of variables (it is actually called oblivious ROABP, but we drop the word oblivious from now on). More formally, there exists an unknown partition of the variable set $\overline{x}$ into d sets $\{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_d\}$ such that in the corresponding matrix product $D_0(\prod_{i=1}^{d-2} D_i)D_{d-1}$, the entries in D_{i-1} are polynomials in variables $\overline{x}_i$, for all $i \in [d]$. It is read once in the sense that in the corresponding ABP, any particular variable contributes to at most one edge on any path.

We work with the matrix representation of ABP. We will show a hitting-set for an ROABP $D_0(\prod_{i=1}^{d} D_i)D_{d+1}$ with D_i being *invertible* matrices for all $i \in [d]$ and all the matrices being *sparse* polynomials. Hence, we name this model *sparse-invertible-factor ROABP*. Like in the previous sections, we find a hitting-set by showing a low-support concentration.

For a polynomial D, let its sparsity $\mathsf{s}(D)$ be the number of monomials in D with nonzero coefficients and let $\mu(D)$ be the maximum support of any monomial in D.

Theorem 3 (restated). Let $\overline{x} = \overline{x}_0 \sqcup \cdots \sqcup \overline{x}_{d+1}$, with $|\overline{x}| = n$. Let $C(\overline{x}) = D_0^T D D_{d+1} \in \mathbb{F}[\overline{x}]$ be a polynomial with $D(\overline{x}) = \prod_{i=1}^{d} D_i(\overline{x}_i)$, where $D_0 \in \mathbb{F}^w[\overline{x}_0]$ and $D_{d+1} \in \mathbb{F}^w[\overline{x}_{d+1}]$ and for all $i \in [d]$, $D_i \in \mathbb{F}^{w\times w}[\overline{x}_i]$ is an invertible matrix. For all $i \in \{0, 1, \ldots, d+1\}$, D_i has degree bounded by δ, $\mathsf{s}(D_i) \le s$ and $\mu(D_i) \le \mu$. Let $\ell := 1 + 2\min\{\lceil \log(w^2 \cdot s)\rceil, \mu\}$. Then there is a hitting-set of size $\text{poly}((n\delta s)^{\ell w^2})$ for $C(\overline{x})$. *(For the proof, see Appendix D.)*

Remark 23. If $\mu = 1$, i.e. each D_i is univariate or linear, then we get poly-time for constant w.

Proof Idea- Here again, we find a hitting-set by proving low-support concentration, but with a different approach. As all the matrices in the matrix product $D(\overline{x}) = \prod_{i=1}^{d} D_i(\overline{x}_i)$ are over disjoint sets of variables, any coefficient in the polynomial $D(\overline{x})$ can be uniquely written as a product of d factors, each coming from one D_i. We start with the assumption that the constant term of each polynomial D_i, denoted by $D_{i\mathbf{0}}$, is an invertible matrix. Using this we define a notion of *parent* and *child* between all the coefficients: If a coefficient can be obtained from another coefficient by replacing one of its constant factors $D_{i\mathbf{0}}$ with another term (with non-trivial support) from D_i, then former is called a parent of the latter. Observe that if we want to do this replacement by a multiplication of some matrix, then $D_{i\mathbf{0}}$ should be invertible. Moreover, all the factors on its right side (or its left side) also need to be constant terms in their respective matrices (this is because of non-commutativity). For a coefficient, the set of matrices D_i which contribute a non-trivial factor to it, is said to form the *block-support* of the coefficient.

Our next step is to show that if a coefficient linearly depends on its descendants then the dependence can be lifted to its parent (by dividing and multiplying appropriate factors) i.e. its parent also linearly depends on its descendants. As the dimension of the matrix algebra is constant, if we take an appropriately large (constant) child-parent chain, there will be a linear dependence among the coefficients in the chain. As the dependencies lift to the parent, they can be lifted all the way up. By an inductive argument it follows that every coefficient depends on the coefficients with low-block-support. Now, this can be translated to low-support concentration in D, if a low-support concentration is assumed in each D_i.

To achieve low-support concentration in each D_i, we use an appropriate shift. The sparsity of D_i is used crucially in this step. To make $D_{i\mathbf{0}}$ invertible, again an appropriate shift is used. Note that $D_{i\mathbf{0}}$ can be made invertible by a shift only when D_i itself is invertible, hence the invertible-factor assumption.

6 Discussion

We conjecture that for depth-3 multilinear circuits, low-support concentration can be achieved by an efficient shift. A first question here is to remove the knowledge of the base sets in Theorem 2.

In the case of constant width ROABP, we could show constant-support concentration, but only after assuming that the factor matrices are invertible. It seems that the invertibility assumption restricts the computing power of ROABP significantly. It is desirable to have low-support concentration without the assumption of invertibility.

As in the case of invertible ROABP and width-2 ROABP, analogous results hold in the boolean setting, it will be interesting to see if there is some connection, at the level of techniques, between pseudorandom generators for boolean and arithmetic models.

7 Acknowledgements

We thank Chandan Saha for suggestions to improve this paper. Several useful ideas about δ-distance circuits and base sets came up during discussions with him. RG thanks TCS research fellowship for support. NS thanks DST for support.

References

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

[ASS13] Manindra Agrawal, Chandan Saha, and Nitin Saxena. Quasi-polynomial hitting-set for set-depth- formulas. In *STOC*, pages 321–330, 2013.

[ASSS12] Manindra Agrawal, Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. Jacobian hits circuits: hitting-sets, lower bounds for depth-d occur-k formulas & depth-3 transcendence degree-k circuits. In *STOC*, pages 599–614, 2012.

[BDVY13] Andrej Bogdanov, Zeev Dvir, Elad Verbin, and Amir Yehudayoff. Pseudorandomness for width-2 branching programs. *Theory of Computing*, 9:283–293, 2013.

[BOC92] Michael Ben-Or and Richard Cleve. Computing algebraic formulas using a constant number of registers. *SIAM J. Comput.*, 21(1):54–58, 1992.

[De11] Anindya De. Pseudorandomness for permutation and regular branching programs. In *IEEE Conference on Computational Complexity*, pages 221–231, 2011.

[DS07] Zeev Dvir and Amir Shpilka. Locally decodable codes with two queries and polynomial identity testing for depth 3 circuits. *SIAM J. Comput.*, 36(5):1404–1434, 2007.

[FSS14] Michael A. Forbes, Ramprasad Saptharishi, and Amir Shpilka. Pseudorandomness for multilinear read-once algebraic branching programs, in any order. In *STOC*, 2014.

[GKKS13] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth three. *FOCS*, 2013.

[JQS10] Maurice J. Jansen, Youming Qiao, and Jayalal M. N. Sarma. Deterministic identity testing of read-once algebraic branching programs. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:84, 2010.

[KNP11] Michal Koucký, Prajakta Nimbhorkar, and Pavel Pudlák. Pseudorandom generators for group products: extended abstract. In *STOC*, pages 263–272, 2011.

[Kro82] Leopold Kronecker. *Grundzuge einer arithmetischen Theorie der algebraischen Grossen.* Berlin, G. Reimer, 1882.

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

[KS09] Neeraj Kayal and Shubhangi Saraf. Blackbox polynomial identity testing for depth 3 circuits. In *FOCS*, pages 198–207, 2009.

[KS11] Zohar Shay Karnin and Amir Shpilka. Black box polynomial identity testing of generalized depth-3 arithmetic circuits with bounded top fan-in. *Combinatorica*, 31(3):333–364, 2011.

[Mul12a] Ketan Mulmuley. Geometric complexity theory V: Equivalence between blackbox derandomization of polynomial identity testing and derandomization of Noether's normalization lemma. In *FOCS*, pages 629–638, 2012.

[Mul12b] Ketan D. Mulmuley. The gct program toward the p vs. np problem. *Commun. ACM*, 55(6):98–107, June 2012.

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

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

[Sax08] Nitin Saxena. Diagonal circuit identity testing and lower bounds. In *ICALP*, volume 5125 of *Lecture Notes in Computer Science*, pages 60–71. Springer, 2008.

[Sax09] Nitin Saxena. Progress on polynomial identity testing. *Bulletin of the EATCS*, 99:49–79, 2009.

[Sax14] Nitin Saxena. Progress on polynomial identity testing - 2. *CoRR*, abs/1401.0976, 2014.

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

[SS11] Nitin Saxena and C. Seshadhri. An almost optimal rank bound for depth-3 identities. *SIAM J. Comput.*, 40(1):200–224, 2011.

[SS12] Nitin Saxena and C. Seshadhri. Blackbox identity testing for bounded top-fanin depth-3 circuits: The field doesn't matter. *SIAM J. Comput.*, 41(5):1285–1298, 2012.

[SS13] Nitin Saxena and C. Seshadhri. From sylvester-gallai configurations to rank bounds: Improved blackbox identity test for depth-3 circuits. *J. ACM*, 60(5):33, 2013.

[SSS09] Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. The power of depth 2 circuits over algebras. In *FSTTCS*, pages 371–382, 2009.

[SSS13] Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. A case of depth-3 identity testing, sparse factorization and duality. *Computational Complexity*, 22(1):39–69, 2013.

[Ste12] Thomas Steinke. Pseudorandomness for permutation branching programs without the group theory. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:83, 2012.

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

A Complete proofs of Section 3

Lemma 2 (restated). If $D(\overline{x})$ is ℓ-concentrated, then there is a $n^{O(\ell)}$-time hitting-set for $C(\overline{x})$.

Proof. Assume $D(\overline{x})$ is ℓ-concentrated. We now make the following observation.

Observation 24 *The polynomial $C(\overline{x}) = 0$ iff all of its $(< \ell)$-support monomials have coefficient $= 0$.*

This observation gives an $n^{O(\ell)}$-time hitting-set for $C(\overline{x})$: We can isolate each of the $(< \ell)$-support monomials and check whether their coefficient is 0. How do we isolate the $(< \ell)$-support monomials? For each subset S of cardinality $< \ell$, set the variables in the set S to 1 and the other variables to 0. This isolates $\sum_{T \subseteq S} \text{coef}_C(x_T) = \text{coef}_C(x_S) + \sum_{T \subsetneq S} \text{coef}_C(x_T)$. If it was already known that the coefficient of the smaller subsets is 0, then $\sum_{T \subseteq S} \text{coef}_C(x_T) = 0 \implies \text{coef}_C(x_S) = 0$. Thus, starting with coefficients of cardinality 0, we test the coefficient of each set for nonzeroness. The hitting-set thus consists of substituting $\overline{x} = (b_1, b_2, \ldots, b_n), b_i \in \{0, 1\}$ with $< \ell$ many 1s. There are $\sum_{i=0}^{\ell-1} \binom{n}{i} = n^{O(\ell)}$ such tuples.

Proof (of Observation 24). The forward implication holds trivially. To prove the reverse implication, we express the coefficients of the polynomial $C(\overline{x})$ as a linear combination of its low-support coefficients.

$$\begin{aligned} C(\overline{x}) &= \overline{a}^T \cdot D(\overline{x}) \\ &= \overline{a}^T \cdot \left(\sum_{S \subseteq [n]} u_S x_S \right), \text{ where } u_S := \text{coef}_D(x_S) \\ &= \sum_{S \subseteq [n]} \overline{a}^T u_S x_S \end{aligned}$$

Since $D(\overline{x})$ is ℓ-concentrated, the coefficients of each of its monomials can be expressed as a linear combination of coefficients of low-support monomials, i.e. $\forall S \subseteq [n], u_S = \sum_{R \subseteq [n], |R| < \ell} \alpha_{R,S} u_R$.

Hence,

$$C(\overline{x}) = \sum_{S\subseteq[n]} \left(\sum_{R\subseteq[n],|R|<\ell} \alpha_{R,S} \left(\overline{a}^T u_R\right) \right) x_S.$$

Let $\beta_R := \text{coef}_C(x_R) = \overline{a}^T u_R$. Hence,

$$C(\overline{x}) = \sum_{S\subseteq[n]} \left(\sum_{R\subseteq[n],|R|<\ell} \alpha_{R,S}\beta_R \right) x_S.$$

This proves the claim.

Lemma 25 (Efficient Kronecker map [Kro82, Agr05]). *Consider a set A of n-variate monomials with maximum individual degree d and $|A| = a$. There exists a set of N-many monomial maps $\phi\colon \bar{t} \to \{t^i\}_{i=1}^{N\log N}$, such that at least one of them separates the monomials in A, where $N := nda^2\log(d+1)$.*

Proof. Since we want to separate the n-variate monomials with maximum individual degree d, we use the naive Kronecker map $\phi'\colon t_i \mapsto t^{(d+1)^{i-1}}$ for all $i \in [n]$. It can easily be seen that there is a 1-1 correspondence between the original monomials and the substituted monomials. But the degree of the new monomials can be very high.

Hence, we take the substitutions $\text{mod}(t^p - 1)$ for many small primes p. In other words, the degrees are taken modulo p. Each prime p leads to a different substitution of the variables t_is. That is our set of candidate maps. We need to bound the number N of primes that ensure that at least one substitution separates the monomials in A. We choose the smallest N primes, say $\mathcal{P}$ is the set. By the effective version of the Prime Number Theorem, the highest value in the set $\mathcal{P}$ is $N\log N$.

To bound the number N of primes: We want a p in the set $\mathcal{P}$ such that $\forall i \neq j,\ d_i - d_j \not\equiv 0 \pmod{p}$, where d_is are the degrees of t in the a monomials after the substitution ϕ'. I.e. we want a p such that $p \nmid \prod_{i\neq j}(d_i - d_j)$.

There are N such primes. Hence we want that $\prod_{p\in\mathcal{P}} p \nmid \prod_{i\neq j}(d_i - d_j)$. This can be ensured by setting $\prod_{p\in\mathcal{P}} p > \prod_{i\neq j}(d_i - d_j)$. There are $(< a^2)$ such monomial pairs and each $d_i < (d+1)^{nd}$. Also, $\prod_{p\in\mathcal{P}} p > 2^N$. Hence, $N = nda^2\log(d+1)$ suffices.

Lemma 3 (restated). If for a polynomial $D(\overline{x}) \in \mathbb{H}_k(\mathbb{F})[\overline{x}]$, there exists a set of $f(n)$-many maps from $\bar{t}$ to $\{t^i\}_{i=1}^{g(n)}$ such that for at least one of the maps ϕ, the shifted polynomial $D(\overline{x} + \phi(\bar{t}))$ has ℓ-concentration, then $C(\overline{x}) = \overline{a}^T D(\overline{x})$, for any $\overline{a} \in \mathbb{F}^k$, has an $n^{O(\ell)}f(n)g(n)$-time hitting-set.

Proof. We need to try $f(n)$-many maps, such that one of them will ensure ℓ-concentration in $D(\overline{x}+\phi(\bar{t}))$. Lemma 2 shows that ℓ-concentration implies a hitting-set of size $n^{O(\ell)}$ for $C(\overline{x}+\phi(\bar{t}))$. But each evaluation of $C(\overline{x}+\phi(\bar{t}))$ will be polynomial in t with degree at most $ng(n)$. This multiplies a factor of $ng(n)$ to the running time. So, total time complexity is $n^{O(\ell)}f(n)g(n)$.

Lemma 20 (restated). Let $F(\overline{x}) \in \mathbb{H}_j(\mathbb{F})[\overline{x}]$ be a δ-distance circuit. Let $F'(\overline{x}) = F(\overline{x} + \phi_{\delta\ell_0}(\bar{t}))$ be its shifted version. Let $V_{F'}$ be the subspace spanned by the phase-$(< j)$ coefficients of $F'(\overline{x})$. Then, $F'(\overline{x})$ has ℓ_0-concentration modulo $V_{F'}$. I.e. $\forall$ monomial T in F',

$$\text{coef}_{F'}(x_T) \in \text{span}\{\text{coef}_{F'}(x_S) \mid S \subseteq \text{monom}(F'),\ |S| < \ell_0\} + V_{F'},$$

where, $\text{monom}(F')$ denotes the set of all possible multilinear monomials that appear in F'.

Proof. The proof is by induction on the number of neighborhoods in F', $\|\text{nbd}_j(F')\|$. We prove the theorem for intermediate subcircuits E' of F'. These subcircuits E' have progressively more neighborhoods in the partition $\mathbb{P}_j$.

Base Case: Any shifted δ-distance circuit, E' over j coordinates, with $\|\text{nbd}_j(E')\| \leq \ell_0$, has ℓ_0-concentration modulo $V_{E'}$, from Lemma 19.

Induction Hypothesis: Any shifted δ-distance circuit, E' over j coordinates, such that $\|\text{nbd}_j(E')\| \leq r$ $(r \geq \ell_0)$, has ℓ_0-concentration modulo $V_{E'}$.

Induction Step: Let $\|\text{nbd}_j(E')\| = r+1$. Let E'_0 be the subcircuit corresponding to the tower over *one* neighborhood in $\text{nbd}_j(E')$. Let E'_1 be the subcircuit corresponding to the tower over the remaining neighborhoods in $\text{nbd}_j(E')$, i.e. $\|\text{nbd}_j(E'_1)\| = r$. Hence, $E' = E'_1 \star E'_0$.

We will now prove ℓ_0-concentration of an arbitrary monomial T in E'. Since E'_0 and E'_1 are disjoint towers, the variables in their supports are disjoint. Let $T_0 = \text{var}(E'_0) \cap T$ and $T_1 = \text{var}(E'_1) \cap T$. Hence, $T = T_0 \cup T_1$. By the induction hypothesis, E'_1 is ℓ_0-concentrated modulo $V_{E'_1}$. Hence,

$$\text{coef}_{E'_1}(x_{T_1}) \in \text{span}\{\text{coef}_{E'_1}(x_S) \mid S \in \text{monom}(E'_1),\ |S| < \ell_0\} + V_{E'_1},$$

where, $\text{monom}(E'_1)$ denotes the set of monomials in E'_1. We know that $\text{coef}_{E'}(x_T) = \text{coef}_{E'_1}(x_{T_1}) \star \text{coef}_{E'_0}(x_{T_0})$ from Observation 9. Multiplying the above equation with $\text{coef}_{E'_0}(x_{T_0})$, we get,

$$\text{coef}_{E'}(x_T) \in \text{span}\{\text{coef}_{E'_1}(x_S) \star \text{coef}_{E'_0}(x_{T_0}) \mid S \in \text{monom}(E'_1),\ |S| < \ell_0\} + V_{E'}.$$

This holds because, by Observation 14, $\text{coef}_{E'_0}(x_{T_0}) \star V_{E'_1} \subseteq V_{E'}$. By Observation 9, the above equation is equivalent to

$$\text{coef}_{E'}(x_T) \in \text{span}\{\text{coef}_{E'}(x_{S \cup T_0}) \mid S \in \text{monom}(E'_1),\ |S| < \ell_0\} + V_{E'}. \tag{1}$$

We will now prove that $\text{coef}_{E'}(x_{S\cup T_0})$ is ℓ_0-concentrated (modulo $V_{E'}$) for all monomials S in E'_1. This will prove that $\text{coef}_{E'}(x_T)$ is ℓ_0-concentrated (modulo $V_{E'}$).

Let $\|\text{nbd}_j(S)\|$ be the number of neighborhoods in $\text{nbd}_j(S)$. Since $|S| < \ell_0$, we know $\|\text{nbd}_j(S)\| < \ell_0$. Hence, $\|\text{nbd}_j(S \cup T_0)\| \leq \ell_0$. Let the subcircuit corresponding to $\text{tower}_j(S \cup T_0)$ be E'_2. Let the subcircuit E'_3 be such that $E' = E'_2 \star E'_3$. From Observation 9,

$$\text{coef}_{E'}(x_{S\cup T_0}) = \text{coef}_{E'_2}(x_{S\cup T_0}) \star \text{coef}_{E'_3}(x_\varnothing).$$

As $\|\text{nbd}_j(E'_2)\| \leq \ell_0$, from Lemma 19, we have ℓ_0-concentration in E'_2 modulo $V_{E'_2}$.

$$\text{coef}_{E'_2}(x_{S\cup T_0}) \in \text{span}\{\text{coef}_{E'_2}(x_R) \mid R \in \text{monom}(E'_2),\ |R| < \ell_0\} + V_{E'_2}.$$

Multiplying throughout with $\text{coef}_{E'_3}(x_\varnothing)$,

$$\text{coef}_{E'}(x_{S\cup T_0}) \in \text{span}\{\text{coef}_{E'_2}(x_R) \star \text{coef}_{E'_3}(x_\varnothing) \mid R \in \text{monom}(E'_2),\ |R| < \ell_0\} + V_{E'}.$$

This is because $\text{coef}_{E'_3}(x_\varnothing) \star V_{E'_2} \subseteq V_{E'}$. The equation is equivalent to

$$\text{coef}_{E'}(x_{S\cup T_0}) \in \text{span}\{\text{coef}_{E'}(x_R) \mid R \in \text{monom}(E'),\ |R| < \ell_0\} + V_{E'}. \tag{2}$$

Using Equations (1) and (2) we get that $\text{coef}_{E'}(T)$ is in the span of coefficients in E' with support $< \ell_0$ modulo $V_{E'}$. Hence, we get ℓ_0-concentration of E' modulo $V_{E'}$.

B Proof of Lemma 19 (Small-Neighborhood $\ell_0 -$ *concentration*)

We will now show ℓ_0-concentration in a δ-distance circuit E' modulo its phase-$(< j)$ coefficients, where the number of neighborhoods in the partition $\mathbb{P}_j$ is $\leq \ell_0$. The basic idea of the proof is to take dependencies among the coefficients of E and lift them to get new dependencies of E'.

Lemma 19 (restated). Let $E(\overline{x}) \in \mathbb{H}_j(\mathbb{F})[\overline{x}]$ be a δ-distance circuit with $\|\text{nbd}_j(E')\| \leq \ell_0$. Let $E'(\overline{x}) = E(\overline{x} + \phi_{\delta\ell_0}(\overline{t}))$ be its shifted version. Let $V_{E'}$ be the subspace spanned by the phase-$(< j)$ coefficients of $E'(\overline{x})$. Then, $E'(\overline{x})$ has ℓ_0-concentration modulo $V_{E'}$. I.e. $\forall$ monomial T in E',

$$\text{coef}_{E'}(x_T) \in \text{span}\{\text{coef}_{E'}(x_S) \mid S \subseteq \text{monom}(E'),\ |S| < \ell_0\} + V_{E'}.$$

Proof. Let the matrices U and $U' \in \mathbb{F}^{[j]\times 2^{[n]}}$ be such that their columns represent the coefficient vectors in E and E' respectively. Let us recall the relation between U and U' from Section 2.1,

$$U = U'\mathcal{M}, \tag{3}$$

where $\mathcal{M}$ is the transfer matrix defined in Section 2.1. Let $\text{Null}(U)$ denote the nullspace of U. Consider a vector $\alpha \in \text{Null}(U)$, i.e. $U\alpha = 0$. Using Equation (3), $U'\mathcal{M}\alpha = 0$. Thus, $\mathcal{M}\alpha \in \text{Null}(U')$.

Let us now study the dependencies we want for U' to show the appropriate concentration. Let $\mathcal{S}_1$ be the set of monomials which have support $\geq \ell_0$ and are in phase-j. Let $\mathcal{S}_0$ be the set of all other monomials. Let $U'_0 \in \mathbb{F}^{[j]\times\mathcal{S}_0}$ and $U'_1 \in \mathbb{F}^{[j]\times\mathcal{S}_1}$ be the submatrices of U' consisting of coefficient vectors of monomials in $\mathcal{S}_0$ and $\mathcal{S}_1$ respectively. To prove the lemma, we need to show that the columns of U'_1 are in the span of the columns of U'_0. Let $\beta \in \mathbb{F}(t)^{2^{[n]}}$ be a dependency among the columns of U', i.e. $U'\beta = \mathbf{0}$. We can break this equation into two parts as: $U'_1\beta_1 = -U'_0\beta_0$, where β_0 and β_1 are the subvectors of β, indexed by the monomials in $\mathcal{S}_0$ and $\mathcal{S}_1$ respectively. In other words,

$$U'_1\beta_1 \equiv \mathbf{0} \pmod{U'_0}, \tag{4}$$

where, by "mod U'_0" we mean modulo column-span(U'_0). Now, we want to show that $U'_1 \equiv \mathbf{0} \pmod{U'_0}$.

Let $|\mathcal{S}_1|$, the number of columns in U'_1 be N^*. Suppose there exists a set of N^* many null-vectors of U'. Let $B \in \mathbb{F}^{2^{[n]}\times[N^*]}$, be the matrix with these null-vectors as its columns. Let B_1 be the submatrix of B obtained by picking the rows indexed by the monomials in $\mathcal{S}_1$. Using Equation (4),

$$U'_1B_1 \equiv \mathbf{0} \pmod{U'_0}.$$

If B_1 is an invertible matrix then,

$$\begin{aligned} & U'_1 \equiv \mathbf{0}B_1^{-1} \pmod{U'_0} \\ \implies & U'_1 \equiv \mathbf{0} \pmod{U'_0}, \end{aligned}$$

which is the desired result.

Now, to show the existence of such a set of null-vectors B, we will take the null-vectors of U and lift them. Here, we will use the notation of a matrix, to also denote the set of its column-vectors. As seen earlier, if $\mathcal{N}^* \subseteq \text{Null}(U)$, then $B := \mathcal{M}\mathcal{N}^* \subseteq \text{Null}(U')$ (by '$\subseteq$' we mean column-wise). Let $\mathcal{M}_1$ be the submatrix of $\mathcal{M}$, obtained by picking the rows of $\mathcal{M}$ indexed by the monomials in $\mathcal{S}_1$. We intend to show that there exists a set of N^* many null-vectors $\mathcal{N}^* \subseteq \text{Null}(U)$ such that $B_1 := \mathcal{M}_1\mathcal{N}^*$ is an invertible matrix.

For this, let us analyze the monomials in $\mathcal{S}_1$. Recall that any phase-j monomial has at most δ support from any neighborhood in partition $\mathbb{P}_j$. As $\|\text{nbd}_j(E)\| \leq \ell_0$, any phase-$j$ monomial has degree at most $\delta\ell_0$. Hence, monomials in $\mathcal{S}_1$ have degree between ℓ_0 and $\delta\ell_0$. It suffices to show $\mathcal{M}_1\mathcal{N}^*$ is an invertible matrix, assuming $\mathcal{M}_1$ has rows indexed by all the monomials with degree between ℓ_0 and $\delta\ell_0$ (If this $\mathcal{M}_1\mathcal{N}^*$ is invertible, then, for every subset $\widehat{\mathcal{M}}_1$ of rows of $\mathcal{M}_1$, there exists a subset $\widehat{\mathcal{N}}^*$ of null-vectors $\mathcal{N}^*$ such that $\widehat{\mathcal{M}}_1\widehat{\mathcal{N}}^*$ is invertible). Now, we redefine N^* to be the number of such monomials.

Recall that $\mathcal{M}_1(S,T) \neq 0$ only when $T \subseteq S$. Since the rows of $\mathcal{M}_1$ are indexed by the monomials of degree at most $\delta\ell_0$, its columns indexed by the monomials of degree $> \delta\ell_0$ will be all zero. Hence, the rows of $\mathcal{N}^*$ indexed by the monomials of degree $> \delta\ell_0$ can be ignored. We thus redefine $\mathcal{M}_1$ and $\mathcal{N}^*$ as their truncated versions.

So, we can as well assume that $\mathcal{N}^* \subseteq \text{Null}(U_{\delta\ell_0})$, where $U_{\delta\ell_0}$ is the submatrix of U, obtained by picking the columns indexed by the monomials of degree $\leq \delta\ell_0$.

Let N be the number of all n-variate monomials of degree $\leq \delta\ell_0$. Let $\mathcal{N} := \text{Null}(U_{\delta\ell_0})$.

Now, in Lemma 28 we show that for any vector-space $\mathcal{N} \subseteq \mathbb{F}^N$ of dimension $\geq (N-k)$, there exists $\mathcal{N}^* \subseteq \mathcal{N}$ such that $\mathcal{M}_1\mathcal{N}^*$ is invertible. As $U_{\delta\ell_0}$ has rank at most j, $\mathcal{N}$ has dimension at least $(N-j) \geq (N-k)$ as required in Lemma 28.

This completes the proof.

To show that there exists such a set of vectors $\mathcal{N}^*$, now we will look at some properties of the transfer matrix. This is an improved result of Theorem 13 in [ASS13]. Their proof involved a complicated greedy algorithm with binary search. Our proof uses a straightforward induction, and is simpler. Their result is only for set-multilinear circuits. Our result is for arbitrary multilinear circuits. Our result has the optimal parameters. I.e. there exist nonzero linear combinations of the rows of the matrix described below that have exactly 2^ℓ nonzero entries.

Lemma 26 (Transfer matrix theorem). *Consider a matrix $M_{n,\ell}$ ($\ell \leq n$), with rows indexed by all possible n-variate multilinear monomials with support size $\geq \ell$ and columns indexed by all possible n-variate multilinear monomials. Let $M_{n,\ell}(S,T) = 1$ if $T \subseteq S$ and 0 otherwise. Then any vector formed by a nonzero linear combination of the rows over any field $\mathbb{F}$ has at least 2^ℓ nonzero entries.*

Proof. Proof is by induction on the number of variables, n.

Base case: There is only one variable, x_1. Then ℓ is 0 or 1.

When $\ell = 1$, the matrix $M_{1,1}$ has only one row, indexed by the monomial x_1. The matrix $M_{1,1}$ has two columns. They are indexed by the empty set $\varnothing$ and the monomial x_1. So, $M_{1,1} = \left(\begin{smallmatrix} 1 & 1 \end{smallmatrix}\right)$; which clearly satisfies the lemma statement.

When $\ell = 0$, both the rows and the columns are indexed by the empty set $\varnothing$ and the monomial x_1. The matrix $M_{1,0} = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$. Since the two rows are linearly independent, any nonzero linear combination will have at least $1 = 2^0$ nonzero entry.

Induction Hypothesis: Assume that for number of variables $= n-1$ and for all $\ell \leq n-1$, any nonzero linear combination of the rows has at least 2^ℓ nonzero entries.

Induction Step: We have to prove the property for $M_{n,\ell}$ for all $\ell \leq n$.

The rows of the matrix $M_{n,\ell}$ are partitioned into two sets: $\mathcal{S}_1$, the set of rows whose indices do not contain x_n and $\mathcal{S}_2$, the set of rows whose indices contain x_n. The columns of the matrix $M_{n,\ell}$ are similarly partitioned into two sets: $\mathcal{T}_1$, the set of columns whose indices do not contain x_n and $\mathcal{T}_2$, the set of columns whose indices contain x_n. Then $M_{n,\ell}$ is divided into four blocks: $\{M_{n,\ell}(\mathcal{S}_i, \mathcal{T}_j)\}, (i,j \in \{1,2\})$. Clearly,

$$M_{n,\ell}(\mathcal{S}_1, \mathcal{T}_1) = M_{n-1,\ell} \tag{5}$$
$$M_{n,\ell}(\mathcal{S}_1, \mathcal{T}_2) = \mathbf{0} \text{ and} \tag{6}$$
$$M_{n,\ell}(\mathcal{S}_2, \mathcal{T}_1) = M_{n,\ell}(\mathcal{S}_2, \mathcal{T}_2). \tag{7}$$

Note that, if $\ell = n$, Equation (5) still holds, since then, $\mathcal{S}_1 = \varnothing$. Equation 7 holds because $T \cup \{n\} \subseteq S \cup \{n\}$ iff $T \subseteq S \cup \{n\}$, where $S, T \subseteq [n-1]$.

We break the linear combination: $\sum_{S \in (\mathcal{S}_1 \cup \mathcal{S}_2)} c_S M_S = \sum_{S \in \mathcal{S}_1} c_S M_S + \sum_{S \in \mathcal{S}_2} d_S M_S$ where $c_S, d_S \in \mathbb{F}$ and M_S is the row of $M_{n,\ell}$ indexed by the set S.

$$\begin{aligned}
\sum_{S \in (\mathcal{S}_1 \cup \mathcal{S}_2)} c_S M_S &= \sum_{S \in \mathcal{S}_1} c_S \left(M(S,\mathcal{T}_1) \,\middle|\, M(S,\mathcal{T}_2) \right) + \sum_{S \in \mathcal{S}_2} d_S \left(M(S,\mathcal{T}_1) \,\middle|\, M(S,\mathcal{T}_2) \right) \\
&= \sum_{S \in \mathcal{S}_1} c_S \left(M(S,\mathcal{T}_1) \,\middle|\, \mathbf{0} \right) + \sum_{S \in \mathcal{S}_2} d_S \left(M(S,\mathcal{T}_1) \,\middle|\, M(S,\mathcal{T}_1) \right) \\
&= \left(\sum_{S \in \mathcal{S}_1} c_S M(S,\mathcal{T}_1) + \sum_{S \in \mathcal{S}_2} d_S M(S,\mathcal{T}_1) \,\middle|\, \sum_{S \in \mathcal{S}_2} d_S M(S,\mathcal{T}_1) \right) \\
&= \left(\overline{C} + \overline{D} \,\middle|\, \overline{D} \right),
\end{aligned}$$

where, $\overline{C} := \sum_{S \in \mathcal{S}_1} c_S \; M(S,\mathcal{T}_1)$ and $\overline{D} := \sum_{S \in \mathcal{S}_2} d_S \; M(S,\mathcal{T}_1)$ are row vectors in $\mathbb{F}^{2^{[n-1]}}$. The second equality holds from Equations (6) and (7).

Let us first consider the case when $\overline{C} \neq \mathbf{0}$. Since $\overline{C}$ is a nonzero linear combination of rows in $M_{n,\ell}(\mathcal{S}_1, \mathcal{T}_1)$, from Equation (5) and by the induction hypothesis, it has $\geq 2^\ell$ nonzero entries. For

any index $T \in \mathcal{T}_1$, if $\overline{C}(T) \neq 0$, but $(\overline{C} + \overline{D})(T) = 0$, then $\overline{D}(T) \neq 0$. Hence, $\left(\overline{C} + \overline{D} \,\middle\vdots\, \overline{D}\right)$ has at least as many nonzero entries as $\overline{C}$, i.e. $\geq 2^\ell$.

But, we cannot use the induction hypothesis if $\overline{C} = \mathbf{0}$, i.e. if $\mathcal{S}_1 = \varnothing$ or $c_S = 0, \forall S \in \mathcal{S}_1$. In this case, we have to show that there are $\geq 2^\ell$ nonzero entries in $\left(\overline{D} \,\middle\vdots\, \overline{D}\right)$, i.e. that there are $\geq 2^{\ell-1}$ nonzero entries in $\overline{D}$ (assuming $\ell \geq 1$). For this, observe that $M_{n,\ell}(\mathcal{S}_2, \mathcal{T}_1) = M_{n-1,\ell-1}$. Hence, we can use the induction hypothesis.

If $\ell = 0$, we need to show that that there are ≥ 1 nonzero entries in $\overline{D}$. For this, observe that $M_{n,0}(\mathcal{S}_2, \mathcal{T}_1) = M_{n-1,0}$. Hence, we can use the induction hypothesis.

Now, from this property of the transfer matrix, we will conclude linear independence of rows of a truncated transfer matrix.

Lemma 27. *Consider the matrix $M_{n,\ell}$ described in Lemma 26. Let us mark any set of at most $2^\ell - 1$ columns in $M_{n,\ell}$. Let $M'_{n,\ell}$ denote the submatrix of $M_{n,\ell}$ consisting of all the unmarked columns. The rows of $M'_{n,\ell}$ are linearly independent.*

Proof. Lemma 26 shows that any nonzero linear combination of the rows of $M_{n,\ell}$ has at least 2^ℓ nonzero entries. $M'_{n,\ell}$ has at most $2^\ell - 1$ columns missing from $M_{n,\ell}$. So, any nonzero linear combination of the rows of $M'_{n,\ell}$ has at least one nonzero entry. In other words, the rows of matrix $M'_{n,\ell}$ are linearly independent.

Now, we are ready to prove the invertibility requirement in Lemma 19. Recall that N denotes the number of all n-variate multilinear monomials with support size $\leq \delta\ell_0$ and N^* denotes the number of all n-variate multilinear monomials with support size between ℓ_0 and $\delta\ell_0$. The truncated transfer matrix (introduced in the proof of Lemma 19) $\mathcal{M}_1$ has dimension $N^* \times N$. Now, we show that the truncated transfer matrix $\mathcal{M}_1$ multiplied by an appropriate set of N^*-many null-vectors gives an invertible matrix.

Lemma 28 (Transfer matrix action). *Consider the truncated transfer matrix $\mathcal{M}_1$. Given any space $\mathcal{N} \subseteq \mathbb{F}^N$ of dimension at least $N - k$, there exists a set of N^* vectors in it, denoted by $\mathcal{N}^*_{N \times N^*}$, such that $\mathcal{M}_1\mathcal{N}^*$ is an invertible matrix.*

Proof. From Section 2.1, recall that $\mathcal{M}_1$ can be written as the product $A' M_1 A$. Here $A'_{N^* \times N^*}$ and $A_{N \times N}$ are diagonal matrices with $A(T,T) = (-1)^{|T|} \frac{1}{\phi_{\delta\ell_0}(t_T)}$. Recall that $\phi_{\delta\ell_0}$ is a univariate map, which sends all monomials t_T to distinct powers of t, when $|T| \leq \delta\ell_0$. Hence, $\phi_{\delta\ell_0}$ gives a total ordering on the monomials t_T. From now on, we assume that the columns of matrix M_1 are arranged in an increasing order according to the ordering given by $\phi_{\delta\ell_0}$.

M_1 is a matrix with columns indexed by monomials with support size $\leq \delta\ell_0$ and rows indexed by monomials with support size between ℓ_0 and $\delta\ell_0$. Also,

$$M_1(S,T) = \begin{cases} 1 & \text{if } T \subseteq S, \\ 0 & \text{otherwise} \end{cases}$$

Take a basis of the space $\mathcal{N}$ of size $N - k$ and with an abuse of notation, denote it by $\mathcal{N}$. In the matrix form $\mathcal{N}$ has dimension $N \times (N - k)$. The rows of matrix $\mathcal{N}$ are also arranged in an increasing order according to the ordering given by $\phi_{\delta\ell_0}$.

Any linear combination of the basis vectors remains in the same space. Hence, we can do column operations in the matrix $\mathcal{N}$. We can assume that after the column operations $\mathcal{N}$ has a lower triangular form. To be more precise, we can assume that in any column of $\mathcal{N}$ the first nonzero entry is 1. And if i_j denotes the index of the first nonzero entry in j-th column, then $i_1 < i_2 < \cdots < i_{N-k}$. Clearly, the rows of $\mathcal{N}$, given by the indices $I := \{i_1, i_2, \ldots, i_{N-k}\}$, are independent. The other rows corresponding to the indices $I' := [N] - I$ are dependent on the I rows. Mark the columns in M_1 corresponding to the indices in I'. We know $|I'| = k \leq 2^{\ell_0} - 1$. Now, we apply Lemma 27 to the matrix M_1. Note that in Lemma 27, the matrix $M_{n,\ell}$ has columns corresponding to all the multilinear monomials while our matrix M_1 has columns only corresponding to monomials of

support size $\leq \delta\ell_0$. So, we cannot directly apply Lemma 27 to M_1. However, note that M_1 has rows corresponding to monomials of support size between ℓ_0 and $\delta\ell_0$. Hence, any column with monomial support size $> \delta\ell_0$ will be a zero column. So, we can ignore the zero columns, and Lemma 27 implies: the rows of M_1' are linearly independent, where M_1' denotes the matrix formed by the unmarked columns of M_1. In other words, there exists a set of N^* unmarked columns in M_1, which are linearly independent. Let the set of indices corresponding to these columns be I^*.

Recall that $I^* \subseteq I$. So, we can choose a set of N^* columns from $\mathcal{N}$ such that the set of their first nonzero indices is I^*. Let the matrix corresponding to these columns be $\mathcal{N}^*_{N\times N^*}$. Now, we consider the square matrix given by $R_{N^*\times N^*} := M_1 A\mathcal{N}^*$. We claim that $\det(R) \neq 0$. To prove the claim we look at the lowest degree term in $\det(R)$. Let R_j be the j-th column of R for $1 \leq j \leq N^*$. Viewing R_j as a polynomial in $\mathbb{F}^{N^*}[t]$, let R_{j0} be the coefficient of lowest degree term in R_j. Let us define a matrix R_0 whose j-th column is R_{j0} for $1 \leq j \leq N^*$. Clearly, $\det(R_0)$ is the coefficient of the lowest degree term in $\det(R)$, in the case when $\det(R_0) \neq 0$. We claim that $\det(R_0) \neq 0$.

We know that $R_j = M_1 A\mathcal{N}^*_j$, where $\mathcal{N}^*_j$ is the j-th column of $\mathcal{N}^*$. Let i_j be the index of the first nonzero entry in $\mathcal{N}^*_j$. As the entries in $A\mathcal{N}^*_j$ have powers of t in an strictly increasing order, the coefficient of the least term in $M_1 A\mathcal{N}^*_j$ is clearly $\pm M_{i_j}$, where M_{i_j} is the i_j-th column of M_1. So, $R_{j0} = \pm M_{i_j}$ and hence $R_0 = M_{I^*}$ (upto a multiplicative factor of (-1) to the columns), where M_{I^*} is the submatrix of M_1 corresponding to I^* columns. As these columns are linearly independent, we get that $\det(R_0) \neq 0$. Which in turn implies that $\det(R) \neq 0$. Hence, the product matrix $A'M_1 A\mathcal{N}^* = \mathcal{M}_1\mathcal{N}^*$ is invertible.

C Complete proofs of Section 4

Lemma 29 (Circuits to coefficients). *Let $D_0(\overline{x}), D_1(\overline{x}), \ldots, D_h(\overline{x})$ be multilinear polynomials in $\mathbb{H}_k(\mathbb{F})[\overline{x}]$ for some field $\mathbb{F}$, where $\overline{x} = \{x_1, x_2, \ldots, x_n\}$. Let $D_i(\overline{x}) = \sum_{S\subseteq[n]} u_{iS}x_S$, $\forall i \in \{0, 1, \ldots, h\}$, where $u_{iS} \in \mathbb{H}_k(\mathbb{F})$. If $D_0(\overline{x}) \in \text{span}_{\mathbb{F}(\overline{x})}\{D_i(\overline{x}) \mid i \in [h]\}$ then*

$$\{u_{0S} \mid S \subseteq [n]\} \in \text{span}_{\mathbb{F}}\{u_{iS} \mid i \in [h],\ S \subseteq [n]\}.$$

Proof. Let us define a field $\mathbb{F}' := \mathbb{F}(\overline{x})$. $D_0 \in \text{span}_{\mathbb{F}'}\{D_i \mid i \in [h]\}$ implies that any null-vector for the vectors $D_1, D_2, \ldots, D_h$ is also a null-vector for D_0, i.e.

$$\{\alpha \in \mathbb{H}_k(\mathbb{F}') \mid \alpha^T \cdot D_i = 0,\ \forall i \in [h]\} \subseteq \{\alpha \in \mathbb{H}_k(\mathbb{F}') \mid \alpha^T \cdot D_0 = 0\}.$$

So, the statement is also true when the vector α coming from $\mathbb{H}_k(\mathbb{F})$, i.e.

$$\{\alpha \in \mathbb{H}_k(\mathbb{F}) \mid \alpha^T \cdot D_i = 0,\ \forall i \in [h]\} \subseteq \{\alpha \in \mathbb{H}_k(\mathbb{F}) \mid \alpha^T \cdot D_0 = 0\}. \tag{8}$$

It is easy to see that the set of null-vectors for a vector $D_i \in \mathbb{H}_k(\mathbb{F}')$, which are coming from $\mathbb{H}_k(\mathbb{F})$, is the same as the intersection of the sets of null-vectors of the coefficient vectors in D_i, i.e.

$$\{\alpha \in \mathbb{H}_k(\mathbb{F}) \mid \alpha^T \cdot D_i = 0\} = \{\alpha \in \mathbb{H}_k(\mathbb{F}) \mid \alpha^T \cdot u_{iS} = 0,\ \forall S \in [n]\}. \tag{9}$$

Using Equations (8) and (9), we can write,

$$\{\alpha \in \mathbb{H}_k(\mathbb{F}) \mid \alpha^T \cdot u_{iS} = 0,\ \forall i \in [h],\ \forall S \subseteq [n]\} \subseteq \{\alpha \in \mathbb{H}_k(\mathbb{F}) \mid \alpha^T \cdot u_{0S} = 0,\ \forall S \subseteq [n]\}.$$

Hence, we can write, for any $T \subseteq [n]$,

$$\{\alpha \in \mathbb{H}_k(\mathbb{F}) \mid \alpha^T \cdot u_{iS} = 0,\ \forall i \in [h],\ \forall S \subseteq [n]\} \subseteq \{\alpha \in \mathbb{H}_k(\mathbb{F}) \mid \alpha^T \cdot u_{0T} = 0\}.$$

This clearly implies, by linear algebra, that for any $T \subseteq [n]$,

$$u_{0T} \in \text{span}_{\mathbb{F}}\{u_{iS} \mid i \in [h],\ S \subseteq [n]\}.$$

Lemma 30. *Let $D(\overline{x})$ has m-base-sets-δ-distance with base sets $\{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_m\}$. Let us write $D(\overline{x}) = \sum_{S\subseteq I_m} u_S x_S$, where I_m denotes the set of indices corresponding to the variable set $\overline{x}_m$ and for all $S \subseteq I_m$, $u_S \in \mathbb{H}_k(\mathbb{F})[\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}]$. Then for any $S \subseteq I_m$, u_S has $(m-1)$-base-sets-δ-distance with base sets $\{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}\}$.*

Proof. It is easy to see that u_S is an evaluation of a derivative of D. To be more precise, $u_S = \frac{\partial D}{\partial x_S}|_{(x_j=0,\ \forall j\in I_m - S)}$, where $\frac{\partial}{\partial x_S} := \circ_{j\in S}\frac{\partial}{\partial x_j}$.

Now, D is a multilinear polynomial computed by a $\Pi\Sigma$ circuit over $\mathbb{H}_k(\mathbb{F})$, hence for each product gate, a variable occurs in at most one of its input wires. If we take the derivative with respect to that variable, the corresponding wire vanishes from the circuit. So, the circuit for u_S is essentially the same as that of D except some input wires to the product gates are missing and also some variables are replaced with 0. Hence, u_S when restricted to any variable set $\overline{x}_i$, still has δ distance, $\forall i \in [m-1]$. Moreover, the variables from $\overline{x}_m$ are not present in u_S. Hence, u_S has $(m-1)$-base-sets-δ-distance with base sets $\{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}\}$.

Lemma 31. *Let a polynomial $D(\overline{x}) \in \mathbb{H}_k(\mathbb{F})[\overline{x}]$ have m-base-sets-δ-distance with base sets being $\{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_m\}$. Then, $D(\overline{x}+\phi^m_{\delta\ell_0}(\bar{t}))$ has $(m(\ell-1)+1)$-concentration, where $\ell = \log(k+1)+\delta+1$.*

Proof. We will prove the theorem by induction on m, i.e. we first show the effect of shift in $\overline{x}_m$ variables and next invoke induction for the shift in $\cup_{i=1}^{m-1}\overline{x}_i$ variables.

Base case: When $m = 1$, $D(\overline{x}) = D(\overline{x}_1)$ is simply a polynomial with δ-distance. So, $D(\overline{x}_1 + \phi^m_{\delta\ell_0}(\bar{t}_1))$ has ℓ-concentration from Lemma 21.

Induction hypothesis: Assume that the statement is true for $m-1$ base sets.

Induction step: We will view the polynomial $D(\overline{x})$ as a polynomial on $\overline{x}_m$ variables with coefficients coming from $\mathbb{H}_k(\mathbb{F})[\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}]$. Let I_i denote the set of indices corresponding to the variables in $\overline{x}_i$. Then we can write $D = \sum_{S\subseteq I_m} u_S x_S$, where $u_S \in \mathbb{H}_k(\mathbb{F})[\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}]$ for each $S \subseteq I_m$. Let us define

$$D' := D(\overline{x}_1 + \phi^m_{\delta\ell_0}(\bar{t}_1), \overline{x}_2 + \phi^m_{\delta\ell_0}(\bar{t}_2), \ldots, \overline{x}_m + \phi^m_{\delta\ell_0}(\bar{t}_m)).$$

Also define

$$\tilde{u}_S := u_S(\overline{x}_1 + \phi^m_{\delta\ell_0}(\bar{t}_1), \overline{x}_2 + \phi^m_{\delta\ell_0}(\bar{t}_2), \ldots, \overline{x}_{m-1} + \phi^m_{\delta\ell_0}(\bar{t}_{m-1}))$$

and

$$\tilde{D} := \sum_{S\subseteq I_m} \tilde{u}_S x_S.$$

Let $\tilde{\mathbb{F}} := \mathbb{F}(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}, y_1, y_2, \ldots, y_{m-1})$. It is easy to see that $\tilde{D}$, as a polynomial in $\mathbb{H}_k(\tilde{\mathbb{F}})[\overline{x}_m]$, has δ-distance. Hence, from Lemma 21, $D' = \tilde{D}(\overline{x}_m + \phi^m_{\delta\ell_0}(\bar{t}_m))$ has ℓ-concentration over $\tilde{\mathbb{F}}(y_m)$. Note that our working field $\tilde{\mathbb{F}}$ having variables $\{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}\}$ and $\{y_1, y_2, \ldots, y_{m-1}\}$ is not a problem as y_m is a new variable. ℓ-Concentration of $D' = \sum_{S\subseteq I_m} u'_S x_S$ means that for any $T \subseteq I_m$,

$$u'_T \in \text{span}_{\tilde{\mathbb{F}}(y_m)}\{u'_S \mid S \subseteq I_m,\ |S| < \ell\}. \tag{10}$$

Next, we will show this kind of dependence among their coefficient vectors. Let us define a field $\mathbb{F}_y := \mathbb{F}(y_1, y_2, \ldots, y_m)$. Also define $\mathbb{F}' := \tilde{\mathbb{F}}(y_m) = \mathbb{F}_y(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1})$. Let $I_{[m-1]}$ denote the indices corresponding to the variables in the set $\overline{x} \setminus \overline{x}_m$. Now for any $T \subseteq I_m$, the vector $u'_T \in \mathbb{H}_k(\mathbb{F}')$ can be seen as a polynomial over $\overline{x} \setminus \overline{x}_m$ variables, i.e. $u'_T = \sum_{U\subseteq I_{[m-1]}} u'_{T,U} x_U$, where $u'_{T,U} \in \mathbb{H}_k(\mathbb{F}_y)$ for all $U \in I_{[m-1]}$. From Lemma 29, we know that dependence among polynomial vectors implies a dependence among their coefficients over the base field. So, Equation (10) together with Lemma 29 implies that for any $T \subseteq I_m$, $V \subseteq I_{[m-1]}$,

$$u'_{T,V} \in \text{span}_{\mathbb{F}_y}\{u'_{S,U} \mid S \subseteq I_m,\ |S| < \ell,\ U \subseteq I_{[m-1]}\}. \tag{11}$$

Let us view D' again as a polynomial in $\mathbb{H}_k(\mathbb{F}_y)[\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_m]$. Equation (11) shows that in D', the rank is concentrated on the coefficients corresponding to the monomials which have low-support from $\overline{x}_m$. Next, we will argue that the rank is actually concentrated on the coefficients corresponding to the monomials which have low-support from all the $\overline{x}_i$s.

To show that, we will show low-support concentration in the circuit u'_S for any $S \in I_m$. Let us define a polynomial

$$\widehat{D} := D(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}, \overline{x}_m + \phi^m_{\delta\ell_0}(\bar{t}_m)).$$

Viewing $\widehat{D}$ as a polynomial in $\overline{x}_m$ variables, we can write $\widehat{D} = \sum_{S \subseteq I_m} \hat{u}_S x_S$. We know that D has m-base-sets-δ-distance. It is easy to see that shifting a polynomial in some variables preserves this property. Hence, $\widehat{D}$ has m-base-sets-δ-distance with base sets being $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_m$. Now, $\hat{u}_S \in \mathbb{H}_k(\mathbb{F}(y_m))[\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}]$ is the coefficient of x_S in $\widehat{D}$. So, from Lemma 30, $\hat{u}_S$ will have $(m-1)$-base-sets-δ-distance, with base sets being $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{m-1}$. Now, by our induction hypothesis, $\hat{u}_S$ will have low-support concentration after appropriate shifting. To be precise, $u'_S = \hat{u}_S(\overline{x}_1 + \phi^m_{\delta\ell_0}(\bar{t}_1), \overline{x}_2 + \phi^m_{\delta\ell_0}(\bar{t}_2), \ldots, \overline{x}_{m-1} + \phi^m_{\delta\ell_0}(\bar{t}_{m-1}))$ has $((m-1)(\ell-1)+1)$-concentration. This means that for any $S \subseteq I_m$, $V \subseteq I_{[m-1]}$,

$$u'_{S,V} \in \text{span}_{\mathbb{F}_y}\{u'_{S,U} \mid U \subseteq I_{[m-1]},\ |U| < (m-1)(\ell-1)+1\}. \tag{12}$$

Combining Equation (11) and (12) we get that for any $T \subseteq I_m$, $V \subseteq I_{[m-1]}$,

$$u'_{T,V} \in \text{span}_{\mathbb{F}_y}\{u'_{S,U} \mid S \subseteq I_m,\ |S| < \ell,\ U \subseteq I_{[m-1]},\ |U| < (m-1)(\ell-1)+1\}. \tag{13}$$

Now, let us define $I_{[m]} := I_m \cup I_{[m-1]}$. Viewing at D' as a polynomial in variable set $\overline{x}$ let us write $D' = \sum_{S \subseteq I_{[m]}} u'_S x_S$. From Equation (13), we can say that for any $T \subseteq I_{[m]}$,

$$u'_T \in \text{span}_{\mathbb{F}_y}\{u'_S \mid S \subseteq I_{[m]},\ |S| < m(\ell-1)+1\}.$$

Hence, D' has $(m(\ell-1)+1)$-concentration.

D Building the Proof of Theorem 3

Our first focus will be on the matrix product $D(\overline{x}) := \prod_{i=1}^d D_i$ which belongs to $\mathbb{F}^{w\times w}[\overline{x}]$. We will show low-support concentration in $D(\overline{x})$ over the matrix algebra $\mathbb{F}^{w\times w}$ (which is non-commutative!).

D.1 Low Block-Support

Let the matrix product $D(\overline{x}) := \prod_{i=1}^d D_i$ correspond to an ROABP such that $D_i \in \mathbb{F}^{w\times w}[\overline{x_i}]$ for all $i \in [d]$. Let n_i be the cardinality of $\overline{x}_i$ and let $n = \sum_{i=1}^d n_i$. For an exponent $e = (e_1, e_2, \ldots, e_m) \in \mathbb{N}^m$, and for a set of variables $\overline{y} = \{y_1, y_2, \ldots, y_m\}$, $\overline{y}^e$ will denote ${y_1}^{e_1}{y_2}^{e_2}\ldots{y_m}^{e_m}$.

Viewing D_i as belonging to $\mathbb{F}^{w\times w}[\overline{x}_i]$, one can write $D_i := \sum_{e\in\mathbb{N}^{n_i}} D_{ie}\overline{x}_i^e$. In particular $D_{i\mathbf{0}}$ refers to the constant part of the polynomial D_i.

For any $e \in \mathbb{N}^n$, support of the monomial $\overline{x}^e$ is defined as $\mathrm{S}(e) := \{i \in [n] \mid e_i \neq 0\}$ and support size is defined as $\mathrm{s}(e) := |\mathrm{S}(e)|$. In this section, we will also define block-support of a monomial. Any monomial $\overline{x}^e$ for $e \in \mathbb{N}^n$, can be seen as a product $\prod_{i=1}^d \overline{x}_i{}^{e_i}$, where $e_i \in \mathbb{N}^{n_i}$ for all $i \in [d]$, such that $e = (e_1, e_2, \ldots, e_d)$. We define *block-support* of e, $\mathrm{bS}(e)$ as $\{i \in [d] \mid e_i \neq \mathbf{0}\}$ and *block-support size* of e, $\mathrm{bs}(e) = |\mathrm{bS}(e)|$.

Next, we will show low block-support concentration of $D(\overline{x})$ when *each* $D_{i\mathbf{0}}$ *is invertible*.

As each D_i is a polynomial over a different set of variables, we can easily see that the coefficient of any monomial $\overline{x}^e = \prod_{i=1}^d \overline{x}_i{}^{e_i}$ in $D(\overline{x})$ is

$$D_e := \prod_{i=1}^d D_{ie_i}. \tag{14}$$

Now, we will define a relation of *parent* and *children* between these coefficients.

Definition 32. *For $e^*, e \in \mathbb{N}^n$, D_{e^*} is called a parent of D_e if $\exists j \in [d]$, $j > \max \mathrm{bS}(e)$ or $j < \min \mathrm{bS}(e)$, such that $\mathrm{bS}(e^*) = \mathrm{bS}(e) \cup \{j\}$ and $e^*_i = e_i$, $\forall i \in [d]$ with $i \neq j$.*

If D_{e^*} is a parent of D_e then D_e is a *child* of D_{e^*}. Note that any coefficient has at most two children, on the other hand it can have many parents. In the case when $j > \max\{\mathrm{bS}(e)\}$ we call e, the *left* child of e^* and in the other case we call it the *right* child.

To motivate this definition, observe that if $j > \max \mathrm{bS}(e)$ then by Equation (14) we can write $D_{e^*} = D_e A^{-1} B$, where $A := \prod_{i=j}^{d} D_{i\mathbf{0}}$ and $B := D_{je_j^*} \prod_{i=j+1}^{d} D_{i\mathbf{0}}$. We will denote the product $A^{-1}B$ as $D_{e^{-1}e^*}$. Similarly, if $j < \min\{\mathrm{bS}(e)\}$ then one can write $D_{e^*} = BA^{-1}D_e$, where $A := \prod_{i=1}^{j} D_{i\mathbf{0}}$ and $B := \left(\prod_{i=1}^{j-1} D_{i\mathbf{0}}\right) D_{je_j^*}$. In this case we will denote the product BA^{-1} as $D_{e^*e^{-1}}$. Note that the invertibility of $D_{i\mathbf{0}}$s is crucial here.

We also define *descendants* of a coefficient D_e as $\mathrm{descend}(D_e) := \{D_f \mid f \in \mathbb{N}^n,\ \mathrm{bS}(f) \subset \mathrm{bS}(e)\}$. Now, we will view the coefficients as $\mathbb{F}$-vectors and look at the linear dependence between them. The following lemma shows how these dependencies lift to the parent.

Lemma 33 (Child to parent). *Let D_{e^*} be a parent of D_e. If D_e is linearly dependent on its descendants, then D_{e^*} is linearly dependent on its descendants.*

Proof. Let D_e be the left child of D_{e^*} (the other case is similar). So, we can write

$$D_{e^*} = D_e D_{e^{-1}e^*}. \tag{15}$$

Let the dependence of D_e on its descendants be the following:

$$D_e = \sum_{\substack{f \\ \mathrm{bS}(f) \subset \mathrm{bS}(e)}} \alpha_f D_f.$$

Using Equation (15) we can write,

$$D_{e^*} = \sum_{\substack{f \\ \mathrm{bS}(f) \subset \mathrm{bS}(e)}} \alpha_f D_f D_{e^{-1}e^*}.$$

Now, we just need to show that for any D_f with $\mathrm{bS}(f) \subset \mathrm{bS}(e)$, $D_f D_{e^{-1}e^*}$ is a valid coefficient of some monomial in $D(\overline{x})$ and also that it is a descendant of D_{e^*}. Recall that $D_{e^{-1}e^*} = A^{-1}B$, where $A := \prod_{i=j}^{d} D_{i\mathbf{0}}$ and $B := D_{je_j^*} \prod_{i=j+1}^{d} D_{i\mathbf{0}}$ and $\mathrm{bS}(e^*) = \mathrm{bS}(e) \cup \{j\}$. We know that $j > \max\{\mathrm{bS}(e)\}$. Hence, $j > \max\{\mathrm{bS}(f)\}$ as $\mathrm{bS}(f) \subset \mathrm{bS}(e)$. So, it is clear that $D_f D_{e^{-1}e^*}$ is the coefficient of $\overline{x}^{f^*} := \overline{x}^f \overline{x}_j{}^{e_j^*}$. It is easy to see that $\mathrm{bS}(f^*) = \mathrm{bS}(f) \cup \{j\} \subset \mathrm{bS}(e^*)$. Hence, $D_{f^*} = D_f D_{e^{-1}e^*}$ is a descendant of D_{e^*}.

Note that, the set of descendants of a coefficient strictly contains its children, grand-children, etc. Clearly, if the descendants are more than $\dim_{\mathbb{F}} \mathbb{F}^{w \times w}$, then there will be a linear dependence among them. So,

Lemma 34. *Any coefficient D_e, with $\mathrm{bs}(e) = w^2$, $\mathbb{F}$-linearly depends on its descendants.*

Proof. First of all, we show that if a coefficient D_{f^*} is nonzero then so are its children. Let us consider its left child D_f (the other case is similar). Recall that we can write $D_{f^*} = D_f D_{f^{-1}f^*}$. Hence if D_f is zero, so is D_{f^*}.

Let $k := w^2$. Now, consider a chain of coefficients $D_{e^0}, D_{e^1}, \ldots, D_{e^k} = D_e$, such that for any $i \in [k]$, $D_{e^{i-1}}$ is a child of D_{e^i}. Clearly, $\mathrm{bs}(e^i) = i$ for $0 \le i \le k$. All the vectors in this chain are nonzero because of our above argument, as D_e is nonzero (The case of $D_e = 0$ is trivial). These $k+1$ vectors lie in $\mathbb{F}^k$, hence, there exists an $i \in [k]$ such that D_{e^i} is linear dependent on $\{D_{e^0}, \ldots, D_{e^{i-1}}\}$. As descendants include children, grand-children, etc., we can say that D_{e^i} is linearly dependent on its descendants. Now, by applying Lemma 33 repeatedly, we conclude $D_{e^k} = D_e$ is dependent on its descendants.

Note that, for a coefficient D_e with $\mathrm{bs}(e) = i$, its descendants have block-support strictly smaller than i. So, Lemma 34 means that coefficients with block-support w^2 depend on coefficients with block-support $\le w^2 - 1$. Now, we show w^2-block-support-concentration in $D(\overline{x})$, i.e. any coefficient is dependent on the coefficients with block-support $\le w^2 - 1$.

Lemma 35 (w^2-Block-concentration). *Let $D(\overline{x}) = \prod_{i=1}^{d} D_i(\overline{x}_i) \in \mathbb{F}^{w \times w}[\overline{x}]$ be a polynomial with $D_{i\mathbf{0}}$ being invertible for each $i \in [d]$. Then $D(\overline{x})$ has w^2-block-support concentration.*

Proof. Let $k := w^2$. We will actually show that for any coefficient D_e with $\mathrm{bs}(e) \geq k$ (the case when $\mathrm{bs}(e) < k$ is trivial),

$$D_e \in \mathrm{span}\{D_f \mid f \in \mathbb{N}^n,\ \mathrm{bS}(f) \subset \mathrm{bS}(e) \text{ and } \mathrm{bs}(f) \leq k-1\}.$$

We will prove the statement by induction on the block-support of D_e, $\mathrm{bs}(e)$.

Base case: When $\mathrm{bs}(e) = k$, it has been already shown in Lemma 34.

Induction Hypothesis: For any coefficient D_e with $\mathrm{bs}(e) = i-1$ for $i-1 \geq k$,

$$D_e \in \mathrm{span}\{D_f \mid f \in \mathbb{N}^n,\ \mathrm{bS}(f) \subset \mathrm{bS}(e) \text{ and } \mathrm{bs}(f) \leq k-1\}.$$

Induction step: Let us take a coefficient D_e with $\mathrm{bs}(e) = i$. Consider any child of D_e, denoted by $D_{e'}$. As $\mathrm{bs}(e') = i-1$, by our induction hypothesis, $D_{e'}$ is linearly dependent on its descendants. So, from Lemma 33, D_e is linearly dependent on its descendants. In other words,

$$D_e \in \mathrm{span}\{D_f \mid \mathrm{bS}(f) \subset \mathrm{bS}(e) \text{ and } \mathrm{bs}(f) \leq i-1\}. \tag{16}$$

Again, by our induction hypothesis, for any coefficient D_f, with $\mathrm{bs}(f) \leq i-1$,

$$D_f \in \mathrm{span}\{D_g \mid \mathrm{bS}(g) \subset \mathrm{bS}(f) \text{ and } \mathrm{bs}(g) \leq k-1\}. \tag{17}$$

Combining Equations (16) and (17), we get

$$D_e \in \mathrm{span}\{D_g \mid \mathrm{bS}(g) \subset \mathrm{bS}(e) \text{ and } \mathrm{bs}(g) \leq k-1\}.$$

Now, we show low block-support concentration in the actual polynomial computed by an ROABP, i.e. in $C(\overline{x}) = D_0^T(\prod_{i=1}^d D_i)D_{d+1}$, where $D_0, D_{d+1} \in F^w[\overline{x}]$.

Corollary 36. *Let $\overline{x} = \overline{x}_0 \sqcup \overline{x}_1 \sqcup \cdots \sqcup \overline{x}_{d+1}$. Let $D(\overline{x}) \in \mathbb{F}^{w\times w}[\overline{x}_1, \ldots, \overline{x}_d]$ be a polynomial described in Lemma 35. Let $C(\overline{x}) = D_0^T D D_{d+1} \in \mathbb{F}[\overline{x}]$ be a polynomial with $D_0 \in \mathbb{F}^w[\overline{x}_0]$, $D_{d+1} \in \mathbb{F}^w[\overline{x}_{d+1}]$. Then $C(\overline{x})$ has (w^2+2)-block-support concentration.*

Proof. Let $k := w^2$. Lemma 35 shows that $D(\overline{x})$ has k-block-support concentration. The coefficient of $\overline{x}^e$ in C is $C_e := D_{0e_0}\prod_{i=1}^d D_{ie_i}D_{(d+1)e_{d+1}}$, where $e = (e_0, e_1, \ldots, e_d, e_{d+1})$. Let $D_e := \prod_{i=1}^d D_{ie_i}$. By k-block-support concentration of $D(\overline{x})$,

$$D_e \in \mathrm{span}\{D_f \mid \mathrm{bs}(f) \leq k-1\}.$$

Which implies,

$$C_e \in \mathrm{span}\{D_{0e_0} D_f D_{(d+1)e_{d+1}} \mid \mathrm{bs}(f) \leq k-1\}.$$

Clearly, $D_{0e_0}D_fD_{(d+1)e_{d+1}}$ is the coefficient of the monomial $x_0^{e_0}x_1^{f_1}\cdots x_d^{f_d}x_{d+1}^{e_{d+1}}$. Hence, $C_e \in \mathrm{span}\{C_f \mid \mathrm{bs}(f) \leq k+1\}$.

D.2 Low-support concentration

Now, we argue that if $D(\overline{x}) = \prod_{i=1}^d D_i(\overline{x}_i)$ has low block-support concentration and moreover if each D_i has low-support concentration then $D(\overline{x})$ has an appropriate low-support concentration.

Lemma 37 (Composition). *If a polynomial $D(\overline{x}) = \prod_{i=1}^d D_i(\overline{x}_i) \in \mathbb{F}^{w\times w}[\overline{x}]$ has ℓ-block-support concentration and $D_i(\overline{x}_i)$ has ℓ'-support concentration for all $i \in [d]$ then $D(\overline{x})$ has $\ell\ell'$-support concentration.*

Proof. Recall that as D_i's are polynomials over disjoint sets of variables, any coefficient D_f in $D(\overline{x})$ can be written as $\prod_{i=1}^d D_{if_i}$, where $f = (f_1, f_2, \ldots, f_d)$ and D_{if_i} is the coefficient corresponding to the monomial $\overline{x}_i^{f_i}$ in D_i. From the definition of $\mathrm{bS}(f)$, we know that $f_i = 0$, for any $i \notin \mathrm{bS}(f)$. From ℓ'-support concentration of $D_i(\overline{x}_i)$, we know that for any coefficient D_{if_i},

$$D_{if_i} \in \mathrm{span}\{D_{ig_i} \mid g_i \in \mathbb{N}^{n_i},\ \mathrm{s}(g_i) \leq \ell'-1\}.$$

Using this, we can write

$$D_f \in \text{span}\left\{\prod_{i=1}^{d} D_{ig_i} \mid g_i \in \mathbb{N}^{n_i},\ \text{s}(g_i) \le \ell' - 1,\ \forall i \in [d] \text{ and } g_i = \mathbf{0},\ \forall i \notin \text{bS}(f)\right\}.$$

Note that the product $\prod_{i\in[d]} D_{ig_i}$ will be the coefficient of a monomial $\overline{x}^g$ such that $\text{bS}(g) \subseteq \text{bS}(f)$ because $g_i = \mathbf{0},\ \forall i \notin \text{bS}(f)$. Clearly, if $\text{s}(g_i) \le \ell' - 1,\ \forall i \in \text{bS}(f)$ then $\text{s}(g) \le (\ell' - 1)\,\text{bs}(f)$. So, one can write

$$D_f \in \text{span}\{D_g \mid g \in \mathbb{N}^n,\ \text{s}(g) \le (\ell' - 1)\,\text{bs}(f)\}. \tag{18}$$

From ℓ-block-support concentration of $D(\overline{x})$, we know that for any coefficient D_e of $D(\overline{x})$,

$$D_e \in \text{span}\{D_f \mid f \in \mathbb{N}^n,\ \text{bs}(f) \le \ell - 1\}. \tag{19}$$

Using Equations (18) and (19), we can write for any coefficient D_e of $D(\overline{x})$,

$$D_e \in \text{span}\{D_g \mid g \in \mathbb{N}^n,\ \text{s}(g) \le (\ell' - 1)(\ell - 1)\}.$$

Hence, $D(\overline{x})$ has $((\ell - 1)(\ell' - 1) + 1)$-support concentration and hence $\ell\ell'$-support concentration.

Now, we just need to show low-support concentration of each D_i. To achieve that we will use some efficient shift. Shifting will serve a dual purpose. Recall that for Lemma 35, we need invertibility of the constant term in D_i, i.e. $D_{i\mathbf{0}}$, for all $i \in [d]$. In case $D_{i\mathbf{0}}$ is not invertible for some $i \in [d]$, after a shift it might become invertible, since D_i is assumed invertible in the sparse-invertible model. For the shifted polynomial $D'_i(\overline{x}_i) := D_i(\overline{x}_i + \phi(\bar{t}_i))$, its constant term $D'_{i\mathbf{0}}$ is just an evaluation of $D_i(\overline{x})$, i.e. $D_i|_{\overline{x}_i=\phi(\bar{t}_i)}$. Now, we want a shift for D_i which would ensure that $\det(D'_{i\mathbf{0}}) \neq 0$ and that D'_i has low-support concentration. For both the goals we use the sparsity of the polynomial.

For a polynomial D, let its sparsity set $\mathsf{S}(D)$ be the set of monomials in D with nonzero coefficients and $\mathsf{s}(D)$ be its sparsity, i.e. $\mathsf{s}(D) = |\mathsf{S}(D)|$. Let, for a polynomial $D(\overline{x}) \in \mathbb{F}^{w\times w}[\overline{x}]$, $S = \mathsf{S}(D)$ and $s = |S|$. Then it is easy to see that for its determinant polynomial $\mathsf{S}(\det(D)) \subseteq S^w$, where $S^w := \{m_1 m_2 \cdots m_w \mid m_i \in S,\ \forall i \in [w]\}$. Hence $\mathsf{s}(\det(D)) \le s^w$. Now, suppose $\det(D) \neq 0$. We will describe an efficient shift which will make the constant term, of the shifted polynomial, invertible. Let $\phi\colon \bar{t} \to \{t^i\}_{i=0}^{\infty}$ be a monomial map which separates all the monomials in $\det(D(\bar{t}))$, i.e. for any two $\bar{t}^{e_1}, \bar{t}^{e_2} \in \mathsf{S}(\det(D(\bar{t})))$, $\phi(\bar{t}^{e_1}) \neq \phi(\bar{t}^{e_2})$. It is easy to see that if we shift each x_i by $\phi(t_i)$ to get $D'(\overline{x}) = D(\overline{x} + \phi(\bar{t}))$ then $\det(D'_{i\mathbf{0}}) = \det(D|_{\overline{x}=\phi(\bar{t})}) \neq 0$.

For sparse polynomials, Agrawal et al. [ASS13, Lemma 16] have given an efficient shift to achieve low-support concentration. Here, we rewrite their lemma. The map $\phi_{\ell'}\colon \bar{t} \to \{t^i\}_{i=0}^{\infty}$ is said to be separating ℓ'-support monomials of degree δ, if for any two monomials $\bar{t}^{e_1}$ and $\bar{t}^{e_2}$ which have support bounded by ℓ' and degree bounded by δ, $\phi_{\ell'}(\bar{t}^{e_1}) \neq \phi_{\ell'}(\bar{t}^{e_2})$. For a polynomial $D(\overline{x})$, let $\mu(D)$ be the maximum support of a monomial in D, i.e. $\mu(D) := \max_{\overline{x}^e \in \mathsf{S}(D)} \text{s}(e)$.

Lemma 38 ([ASS13]). *Let V be a $\mathbb{F}$-vector space of dimension k. Let $D(\overline{x}) \in V[\overline{x}]$ be a polynomial with degree bound δ. Let $\ell := 1 + 2\min\{\lceil\log(k \cdot \mathsf{s}(D))\rceil, \mu(D)\}$ and ϕ_ℓ be a monomial map separating ℓ-support monomials of degree δ. Then $D(\overline{x} + \phi_\ell(\bar{t}))$ has ℓ-concentration over $\mathbb{F}(t)$.*

The [ASS13] version of the Lemma 38 gave a concentration result about sparse polynomials over $\mathbb{H}_k(\mathbb{F})$. But observe that the process of shifting and the definition of concentration only deal with the additive structure of $\mathbb{H}_k(\mathbb{F})$, and the multiplication structure is irrelevant. Hence, the result is true over any $\mathbb{F}$-vector space, in particular, over the matrix algebra. By combining these observations, we have the following.

Lemma 39. *Let $D(\overline{x}) = \prod_{i=1}^{d} D_i(\overline{x}_i)$ be a polynomial in $\mathbb{F}^{w\times w}[\overline{x}]$ with $\det(D) \neq 0$ such that for all $i \in [d]$, D_i has degree bounded by δ, $\mathsf{s}(D_i) \le s$ and $\mu(D_i) \le \mu$. Let $\ell := 1 + 2\min\{\lceil\log(w^2 \cdot s)\rceil, \mu\}$ and $M := \text{poly}(s^w(n\delta)^\ell)$. Then there is a set of M monomial maps with degree bounded by $M \log M$ such that for at least one of the maps ϕ, $D' := D(\overline{x} + \phi(\bar{t}))$ has ℓw^2-concentration.*

Proof. Let $\phi\colon \bar{t} \to \{t^i\}_{i=0}^{\infty}$ be a map such that it separates all the monomials in $\mathsf{S}(\det(D_i(\bar{t}_i)))$, for all $i \in [d]$. There are ds^{2w} such monomial pairs. Also assume that ϕ separates all monomials of support bounded by ℓ. There are $(n\delta)^{O(\ell)}$ such monomials. Hence, total number of monomial pairs which need to be separated are $s^{O(w)} + (n\delta)^{O(\ell)}$. From Lemma 25, we know that there is a set of M monomial maps with highest degree $M \log M$ such that at least one of the maps ϕ separates the desired monomials, where $M = \text{poly}(s^w(n\delta)^\ell)$. As the map ϕ separates all the monomials in $\mathsf{S}(\det(D_i(\bar{t}_i)))$, $\det(D_i(\phi(\bar{t}_i))) \neq 0$ and hence, $D'_{i\mathbf{0}}$ is invertible for all $i \in [d]$. So, $D'(\overline{x})$ has w^2-block-support concentration from Lemma 35.

From Lemma 38, $D'_i(\overline{x}_i)$ has ℓ-concentration for all $i \in [d]$. Hence, from Lemma 37, $D'(\overline{x})$ has ℓw^2-concentration.

Now, we come back to the proof of Theorem 3 (restated in Section 5). We want to find a hitting set for $C(\overline{x}) = D_0^T D D_{d+1}$, where $D \in \mathbb{F}^{w\times w}[\overline{x}]$ is the polynomial as described in Lemma 39 and $D_0 \in \mathbb{F}^w[\overline{x}_0]$, $D_{d+1} \in \mathbb{F}^w[\overline{x}_{d+1}]$. Using (w^2+2)-block-support concentration of $C(\overline{x})$ from Corollary 36, and arguing as in the proof of Lemma 39, we show $\ell(w^2+2)$-concentration of $C(\overline{x} + \phi(\bar{t}))$, where ϕ is a map which needs to separate $s^{O(w)}+(n\delta)^{O(\ell)}$-many monomials pairs. From Lemma 25, there is set of $\text{poly}(s^w(n\delta)^\ell)$-many shifting maps with highest degree $\text{poly}(s^w(n\delta)^\ell)$ such that one of them is the desired map. Similar to Lemma 2, we can show that if $C(\overline{x} + \phi(\bar{t}))$ has $\ell(w^2+2)$-concentration and has degree bound $\delta^{O(1)}$ then there is a hitting-set of size $(n\delta)^{O(\ell w^2)}$. Each of these evaluations will be a polynomial in t with highest degree $\text{poly}(s^w(n\delta)^\ell)$. Hence, total time complexity becomes $\text{poly}(s^w(n\delta)^{\ell w^2})$.

Note that for constant width ROABP, when $\mu(D_i)$ is bounded by a constant for each $0 \leq i \leq d+1$, in particular when each D_i is univariate, the parameter ℓ becomes constant and the hitting-set becomes polynomial-time.

E Width-2 Read Once ABP

In Section 5, the crucial part in finding a hitting-set for an ROABP, is the assumption that the matrix product $D(\overline{x})$ is invertible. Now, we will show that for width-2 ROABP this assumption is not required. Via a factorization property of 2×2 matrices, we will show that PIT for width-2 sparse-factor ROABP reduces to PIT for width-2 sparse-invertible-factor ROABP.

Lemma 40 (2×2 invertibility). *Let $C(\overline{x}) = D_0^T \left(\prod_{i=1}^d D_i\right) D_{d+1}$ be a polynomial computed by a width-2 sparse-factor ROABP. Then we can write $\alpha(\overline{x})C(\overline{x}) = C_1(\overline{x})C_2(\overline{x})\cdots C_{m+1}(\overline{x})$, for some nonzero $\alpha \in \mathbb{F}[\overline{x}]$ and some $m \leq d$, where $C_i(\overline{x})$ is a polynomial computed by a width-2 sparse-invertible-factor ROABP, for all $i \in [m+1]$.*

Proof. Let us say, for some $i \in [d]$, $D_i(\overline{x}_i)$ is not invertible. Let $D_i = \left[\begin{smallmatrix} a_i & b_i \\ c_i & d_i \end{smallmatrix}\right]$ with $a_i, b_i, c_i, d_i \in \mathbb{F}[\overline{x}_i]$ and $a_i d_i = b_i c_i$. Without loss of generality, at least one of $\{a_i, b_i, c_i, d_i\}$ is nonzero. Let us say $a_i \neq 0$ (other cases are similar). Then we can write,

$$\begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} = \frac{1}{a_i} \begin{bmatrix} a_i \\ c_i \end{bmatrix} \begin{bmatrix} a_i & b_i \end{bmatrix}.$$

In other words, we can write $\alpha_i D_i = A_i B_i^T$, where $A_i, B_i \in \mathbb{F}^2[\overline{x}_i]$ and $0 \neq \alpha_i \in \{a_i, b_i, c_i, d_i\}$. Note that $\mathsf{s}(\alpha_i), \mathsf{s}(A_i), \mathsf{s}(B_i) \leq \mathsf{s}(D_i)$. Let us say the set of non-invertible D_is is $\{D_{i_1}, D_{i_2}, \ldots, D_{i_m}\}$. Writing all of them in the above form we get,

$$C(\overline{x}) \prod_{j=1}^{m} \alpha_{i_j} = \prod_{j=1}^{m+1} C_j,$$

where

$$C_j := \begin{cases} D_0^T \left(\prod_{i=1}^{i_1-1} D_i\right) A_{i_1} & \text{if } j = 1, \\ B_{i_{j-1}}^T \left(\prod_{i=i_{j-1}+1}^{i_j-1} D_i\right) A_{i_j} & \text{if } 2 \leq j \leq m, \\ B_{i_m}^T \left(\prod_{i=i_m+1}^{d} D_i\right) D_{d+1} & \text{if } j = m+1. \end{cases}$$

Clearly, for all $j \in [m+1]$, C_j can be computed by a sparse-invertible-factor ROABP.

Now, from the above lemma it is easy to construct a hitting-set. First we write a general result about hitting-sets for a product of polynomials from some class [SY10, Observation 4.1].

Lemma 41 (Lagrange interpolation). *Suppose $\mathcal{H}$ is a hitting-set for a class of polynomials $\mathcal{C}$. Let $C(\overline{x}) = C_1(\overline{x})C_2(\overline{x})\cdots C_m(\overline{x})$, where $C_i \in \mathcal{C}$ and has degree bounded by δ, for all $i \in [m]$. There is a hitting-set of size $m\delta|\mathcal{H}|+1$ for $C(\overline{x})$.*

Proof. Let $h = |\mathcal{H}|$ and $\mathcal{H} = \{\overline{\alpha}_1, \overline{\alpha}_2, \ldots, \overline{\alpha}_h\}$. Let $B := \{\beta_i\}_{i=1}^h$ be a set of constants. The Lagrange interpolation $\overline{\alpha}(u)$ of the points in $\mathcal{H}$ is defined as follows

$$\overline{\alpha}(u) := \sum_{i=1}^{h} \frac{\prod_{j\neq i}(u-\beta_j)}{\prod_{j\neq i}(\beta_i-\beta_j)}\overline{\alpha}_i.$$

The key property of the interpolation is that when we put $u = \beta_i$, $\overline{\alpha}(\beta_i) = \overline{\alpha}_i$ for all $i \in [h]$. For any $a \in [m]$, we know that $C_a(\overline{\alpha}_i) \neq 0$, for some $i \in [h]$. Hence, $C_a(\overline{\alpha}(u))$ as a polynomial in u is nonzero because $C_a(\overline{\alpha}(\beta_i)) = C_a(\overline{\alpha}_i) \neq 0$. So, we can say $C(\overline{\alpha}(u)) \neq 0$ as a polynomial in u. Degree of $\overline{\alpha}(u)$ is h. So, degree of $C(\overline{\alpha}(u))$ in u is bounded by $m\delta h$. We can put $(m\delta h+1)$-many distinct values of u to get a hitting-set for $C(\overline{\alpha}(u))$.

Note that a hitting-set for $\alpha(\overline{x})C(\overline{x})$ is also a hitting-set for $C(\overline{x})$ if α is a nonzero polynomial. Recall that we get a hitting-set for invertible ROABP from Theorem 3 (Section 5). Lemma 40 tells us how to write a width-2 ROABP as a product of width-2 invertible ROABPs. Combining these results with Lemma 41 we directly get the following.

Theorem 4. *Let $C(\overline{x}) = D_0^T(\overline{x}_0)(\prod_{i=1}^{d} D_i(\overline{x}_i))D_{d+1}(\overline{x}_{d+1})$ be a polynomial in $\mathbb{F}[\overline{x}]$ computed by a width-2 ROABP such that for all $0 \leq i \leq d+1$, D_i has degree bounded by δ, $\mathsf{s}(D_i) \leq s$ and $\mu(D_i) \leq \mu$. Let $\ell := 1 + 2\min\{\lceil \log(4\cdot s)\rceil, \mu\}$ Then there is a hitting-set of size $\mathrm{poly}((n\delta s)^{\ell})$.*

We remark again that when all D_is are constant-variate or linear polynomials, the hitting-set is polynomial-time.

F Sum of set-multilinear circuits

In this section, we will reduce the PIT for sum of constantly many set-multilinear depth-3 circuits, to the PIT for depth-3 circuits with m base sets having δ distance, where $m\delta = o(n)$. Thus, we get a subexponential time whitebox algorithm for this class (from Theorem 2). Note that a sum of constantly many set-multilinear depth-3 circuits is equivalent to a depth-3 multilinear circuit such that the number of distinct partitions, induced by its product gates, is constant.

We first look at the case of two partitions. For a partition $\mathbb{P}$ of $[n]$, let $\mathbb{P}|_B$ denote the restriction of $\mathbb{P}$ on a base set $B \subseteq [n]$. For example, if $\mathbb{P} = \{\{1,2\},\{3,4\},\{5,6,\ldots,n\}\}$ and $B = \{1,3,4\}$ then $\mathbb{P}|_B = \{\{1\},\{3,4\}\}$. Recall that $d(\mathbb{P}_1,\mathbb{P}_2,\ldots,\mathbb{P}_c)$ denotes the *distance* of the partition sequence $(\mathbb{P}_1,\mathbb{P}_2,\ldots,\mathbb{P}_c)$ (Definition 4). For a partition sequence $(\mathbb{P}_1,\mathbb{P}_2,\ldots\mathbb{P}_c)$, and a base set $B \subseteq [n]$, let $d_B(\mathbb{P}_1,\mathbb{P}_2,\ldots,\mathbb{P}_c)$ denote the distance of the partition sequence when restricted to the base set B, i.e. $d(\mathbb{P}_1|_B,\mathbb{P}_2|_B,\ldots,\mathbb{P}_c|_B)$.

Lemma 42. *For any two partitions $\{\mathbb{P}_1,\mathbb{P}_2\}$ of the set $[n]$, there exists a partition of $[n]$, into at most $2\sqrt{n}$ base sets $\{B_1,B_2,\ldots,B_m\}$ $(m < 2\sqrt{n})$, such that for any $i \in [m]$, either $d_{B_i}(\mathbb{P}_1,\mathbb{P}_2) = 1$ or $d_{B_i}(\mathbb{P}_2,\mathbb{P}_1) = 1$.*

Proof. Let us divide the set of colors in the partition $\mathbb{P}_1$, into two types of colors: One with at least $\sqrt{n}$ elements and the other with less than $\sqrt{n}$ elements. In other words, $\mathbb{P}_1 = \{X_1,X_2,\ldots,X_r\} \cup \{Y_1,Y_2,\ldots,Y_q\}$ such that $|X_i| \geq \sqrt{n}$ and $|Y_j| < \sqrt{n}$, for all $i \in [r]$, $j \in [q]$. Let us make each X_i a base set, i.e. $B_i = X_i$, $\forall i \in [r]$. As $|X_i| \geq \sqrt{n}$, $\forall i \in [r]$, we get $r \leq \sqrt{n}$. Now, for any $i \in [r]$, $\mathbb{P}_1|_{B_i}$ has only one color. Hence, irrespective of what colors $\mathbb{P}_2|_{B_i}$ has, $d_{B_i}(\mathbb{P}_2,\mathbb{P}_1) = 1$, for all $i \in [r]$.

Now, for the other kind of colors, we will make base sets which have exactly one element from each color Y_j. More formally, let $Y_j = \{y_{j,1}, y_{j,2}, \ldots, y_{j,r_j}\}$, for all $j \in [q]$. Let $r' = \max\{r_1, r_2, \ldots, r_q\}$ ($r' < \sqrt{n}$). Now define base sets $B'_1, B'_2, \ldots, B'_{r'}$ such that for any $a \in [r']$, $B'_a = \{y_{j,a} \mid j \in [q],\ |Y_j| \geq a\}$. In other words, all those Y_js which have at least a elements, contribute their a-th element to B'_a. Now for any $a \in [r']$, $\mathbb{P}_1|_{B'_a} = \{\{y_{j,a}\} \mid j \in [q],\ |Y_j| \geq a\}$, i.e. it has exactly one element in each color. Clearly, irrespective of what colors $\mathbb{P}_2|_{B'_a}$ has, $d_{B'_a}(\mathbb{P}_1, \mathbb{P}_2) = 1$, for all $a \in [r']$.

$\{B_1, B_2, \ldots, B_r\} \cup \{B'_1, B'_2, \ldots, B'_{r'}\}$ is our final set of base sets. Clearly, they form a partition of $[n]$. The total number of base sets, $m = r + r' < 2\sqrt{n}$.

Now, we generalize Lemma 42 to any constant number of partitions, by induction.

Lemma 43. *For any c partitions $\{\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_c\}$ of the set $[n]$, there exists a partition of $[n]$, into m base sets $\{B_1, B_2, \ldots, B_m\}$ with $m < 2^{c-1} \cdot n^{1-(1/2^{c-1})}$ such that for any $i \in [m]$, there exists a permutation of the partitions, $(\mathbb{P}_{i_1}, \mathbb{P}_{i_2}, \ldots, \mathbb{P}_{i_c})$ with $d_{B_i}(\mathbb{P}_{i_1}, \mathbb{P}_{i_2}, \ldots, \mathbb{P}_{i_c}) = 1$.*

Proof. Let $f(c, n) := 2^{c-1} \cdot n^{1-(1/2^{c-1})}$. The proof is by induction on the number of partitions.

Base case: For $c = 2$, $f(c, n)$ becomes $2\sqrt{n}$. Hence, the statement follows from Lemma 42.

Induction hypothesis: The statement is true for any $c - 1$ partitions.

Induction step: Like in Lemma 42, we divide the set of colors in $\mathbb{P}_1$ into two types of colors. Let $\mathbb{P}_1 = \{X_1, X_2, \ldots, X_r\} \cup \{Y_1, Y_2, \ldots, Y_q\}$ such that $|X_i| \geq \sqrt{n}$ and $|Y_j| < \sqrt{n}$, for all $i \in [r],\ j \in [q]$. Let us set $B_i = X_i$ and let $n_i := |B_i|$, $\forall i \in [r]$. Our base sets will be further subsets of these B_is. For a fixed $i \in [r]$, let us define $\mathbb{P}'_h = \mathbb{P}_h|_{B_i}$, as a partition of the set B_i, for all $h \in [c]$. Clearly, $\mathbb{P}'_1$ has only one color. Now, we focus on the partition sequence $(\mathbb{P}'_2, \mathbb{P}'_3, \ldots, \mathbb{P}'_c)$. From the inductive hypothesis, there exists a partition of B_i into m_i base sets $\{B_{i,1}, B_{i,2}, \ldots, B_{i,m_i}\}$ ($m_i \leq f(c-1, n_i)$) such that for any $u \in [m_i]$, there exists a permutation of $(\mathbb{P}'_2, \mathbb{P}'_3, \ldots, \mathbb{P}'_c)$, given by $(\mathbb{P}'_{i_2}, \mathbb{P}'_{i_3}, \ldots, \mathbb{P}'_{i_c})$, with $d_{B_{i,u}}(\mathbb{P}'_{i_2}, \mathbb{P}'_{i_3}, \ldots, \mathbb{P}'_{i_c}) = 1$. As $\mathbb{P}'_1$ has only one color, so does $\mathbb{P}'_1|_{B_{i,u}}$. Hence, $d_{B_{i,u}}(\mathbb{P}'_{i_2}, \mathbb{P}'_{i_3}, \ldots, \mathbb{P}'_{i_c}, \mathbb{P}'_1)$ is also 1. From this, we easily get $d_{B_{i,u}}(\mathbb{P}_{i_2}, \mathbb{P}_{i_3}, \ldots, \mathbb{P}_{i_c}, \mathbb{P}_1) = 1$. The above argument can be made for all $i \in [r]$.

Now, for the other colors, we proceed as in Lemma 42. Let $Y_j = \{y_{j,1}, y_{j,2}, \ldots, y_{j,r_j}\}$, for all $j \in [q]$. Let $r' = \max\{r_1, r_2, \ldots, r_q\}$ ($r' < \sqrt{n}$). Now define sets $B'_1, B'_2, \ldots, B'_{r'}$ such that for any $a \in [r']$, $B'_a = \{y_{j,a} \mid j \in [q],\ |Y_j| \geq a\}$. In other words, all those Y_js which have at least a elements, contribute their a-th element to B'_a. Let $n'_a := |B'_a|$, for all $a \in [r']$. Our base sets will be further subsets of these B'_as. For a fixed $a \in [r']$, let us define $\mathbb{P}'_h = \mathbb{P}_h|_{B'_a}$, as a partition of the set B'_a, for all $h \in [c]$. Clearly, $\mathbb{P}'_1$ has exactly one element in each of its colors. Now, we focus on the partition sequence $(\mathbb{P}'_2, \mathbb{P}'_3, \ldots, \mathbb{P}'_c)$. From the inductive hypothesis, there exists a partition of B'_a into m'_a base sets $\{B'_{a,1}, B'_{a,2}, \ldots, B'_{a,m'_a}\}$ ($m'_a \leq f(c-1, n'_a)$) such that for any $u \in [m'_a]$, there exists a permutation of $(\mathbb{P}'_2, \mathbb{P}'_3, \ldots, \mathbb{P}'_c)$, given by $(\mathbb{P}'_{i_2}, \mathbb{P}'_{i_3}, \ldots, \mathbb{P}'_{i_c})$, with $d_{B'_{a,u}}(\mathbb{P}'_{i_2}, \mathbb{P}'_{i_3}, \ldots, \mathbb{P}'_{i_c}) = 1$. As $\mathbb{P}'_1$ has exactly one element in each of its colors, so does $\mathbb{P}'_1|_{B'_{a,u}}$. Hence, $d_{B'_{a,u}}(\mathbb{P}'_1, \mathbb{P}'_{i_2}, \mathbb{P}'_{i_3}, \ldots, \mathbb{P}'_{i_c})$ is also 1. From this, we easily get $d_{B'_{a,u}}(\mathbb{P}_1, \mathbb{P}_{i_2}, \mathbb{P}_{i_3}, \ldots, \mathbb{P}_{i_c}) = 1$. The above argument can be made for all $a \in [r']$.

Our final set of base sets will be $\{B_{i,u} \mid i \in [r],\ u \in [m_i]\} \cup \{B'_{a,u} \mid a \in [r'],\ u \in [m'_a]\}$. As argued above, when restricted to any of these base sets, the given partitions have a sequence, which has distance 1. Now, we need to bound the number of these base sets,

$$m = \sum_{i \in [r]} m_i + \sum_{a \in [r']} m'_a.$$

From the bounds on m_i and m'_a, we get

$$m \leq \sum_{i \in [r]} f(c-1, n_i) + \sum_{a \in [r']} f(c-1, n'_a).$$

Recall that $n_i \geq \sqrt{n}$. We break the second sum, in the above equation, into two parts. Let $R_1 = \{a \in [r'] \mid n'_a \geq \sqrt{n}\}$ and $R_2 = \{a \in [r'] \mid n'_a < \sqrt{n}\}$.

$$m \leq \sum_{i \in [r]} f(c-1, n_i) + \sum_{a \in R_1} f(c-1, n'_a) + \sum_{a \in R_2} f(c-1, n'_a). \tag{20}$$

Let us first focus on the third sum. Note that $|R_2| \leq r' < \sqrt{n}$. For $a \in R_2$, $n'_a < \sqrt{n}$ and hence $f(c-1, n'_a) < f(c-1, \sqrt{n}) = 2^{c-2} \cdot n^{1/2-(1/2^{c-1})}$. So,

$$\sum_{a \in R_2} f(c-1, n'_a) < \sqrt{n} \cdot 2^{c-2} \cdot n^{1/2-(1/2^{c-1})} = 2^{c-2} \cdot n^{1-(1/2^{c-1})}. \tag{21}$$

Now, we focus on first two sums in Equation 20. As, $n_i \geq \sqrt{n}$, $\forall i \in [r]$ and $n'_a \geq \sqrt{n}$, $\forall a \in R_1$, we combine these two sums (with an abuse of notation) and write the sum as follows,

$$\sum_{i \in [r'']} f(c-1, n_i),$$

where $r'' = r + |R_1|$, and $n_i \geq \sqrt{n}$, $\forall i \in [r'']$. As each $n_i \geq \sqrt{n}$, we know $r'' < \sqrt{n}$ (as $\sum n_i \leq n$).

Observe that $f(c-1, z)$, as a function of z, is a concave function (its derivative is monotonically decreasing, when $z > 0$). From the properties of a concave function, we know,

$$\frac{1}{r''} \sum_{i \in [r'']} f(c-1, n_i) \leq f\left(c-1, \frac{1}{r''} \sum_{i \in [r'']} n_i\right).$$

Now, $\sum_{i \in [r'']} n_i \leq n$ and $f(c-1, z)$ is an increasing function (when $z > 0$). Hence,

$$\frac{1}{r''} \sum_{i \in [r'']} f(c-1, n_i) \leq f\left(c-1, \frac{1}{r''} n\right).$$

Equivalently,

$$\begin{aligned}
\sum_{i \in [r'']} f(c-1, n_i) &\leq r'' \cdot 2^{c-2} \cdot (n/r'')^{1-(1/2^{c-2})} \\
&= 2^{c-2} \cdot n^{1-(1/2^{c-2})} \cdot (r'')^{1/2^{c-2}} \\
&< 2^{c-2} \cdot n^{1-(1/2^{c-2})} \cdot n^{1/2^{c-1}} \\
&= 2^{c-2} \cdot n^{1-(1/2^{c-1})}
\end{aligned}$$

Using this with Equation 21 and substituting in Equation 20, we get

$$m < 2^{c-1} \cdot n^{1-(1/2^{c-1})}.$$

Now, we combine these results with our hitting-sets for depth-3 circuits having m base sets with δ-distance.

Theorem 5. *Let $C(\overline{x})$ be a n-variate polynomial, which is a sum of c set-multinear depth-3 circuits, each having top fan-in k. Then there is a $n^{O(2^{c-1} n^{1-\epsilon} \log k)}$-time whitebox test for C, where $\epsilon := 1/2^{c-1}$.*

Proof. As mentioned earlier, the polynomial $C(\overline{x})$ can be viewed as computed by a depth-3 multilinear circuit, such that its product gates induce at most c-many distinct partitions. From Lemma 43, we can partition the variable set into m base sets, such that for each of these base sets, the partitions can be sequenced to have distance 1, where $m := 2^{c-1} n^{1-\epsilon}$. Hence, the polynomial C has m base sets with 1-distance and top fan-in ck. Moreover, from the proof of Lemma 43, it is clear that such base sets can be computed in $n^{O(c)}$-time. From Theorem 2, we know there is $n^{O(m \log(ck))}$-time whitebox test for such a circuit. Substituting the value of m, we get the result.

Tightness of this result

Lemma 42 can be put in other words as, any two partitions have m-base-sets-δ-distance with $m\delta = O(\sqrt{n})$. We can, in fact, show that this result is tight.

Showing the lower bound: Let $d(\mathbb{P}_1, \mathbb{P}_2) = \delta$. Then each color of $\mathbb{P}_2$ has a friendly neighborhood (of at most δ colors) which is exactly partitioned in $\mathbb{P}_1$. Now construct δ base sets such that i-th base set takes the variables of i-th color from every neighborhood of $\mathbb{P}_2$. Clearly, when restricted to one of these bases sets, $d(\mathbb{P}_1, \mathbb{P}_2)$ is 1. In other words $\mathbb{P}_1$ and $\mathbb{P}_2$ have δ-base-sets-1-distance. Similarly, one can argue that if $\mathbb{P}_1$ and $\mathbb{P}_2$ have m-base-sets-δ-distance then they also have $m\delta$-base-sets-1-distance. Now, we will show that if we want m-base-sets-1-distance for two partitions then $m = \Omega(\sqrt{n})$.

Consider the following example (assuming n is a square):

$\mathbb{P}_1 = \{\{1, 2, \ldots \sqrt{n}\}, \{\sqrt{n}+1, \sqrt{n}+2, \ldots, 2\sqrt{n}\}, \ldots, \{\sqrt{n}(\sqrt{n}-1)+1, \sqrt{n}(\sqrt{n}-1)+2, \ldots, n\}\}$ and

$\mathbb{P}_2 = \{\{1, \sqrt{n}+1, \ldots, \sqrt{n}(\sqrt{n}-1)+1\}, \{2, \sqrt{n}+2, \ldots, \sqrt{n}(\sqrt{n}-1)+2\}, \ldots, \{\sqrt{n}, 2\sqrt{n}, \ldots, n\}\}$. Basically, $\mathbb{P}_2$ has the residue classes (mod $\sqrt{n}$).

Observation 44 *A base set B, such that $d_B(\mathbb{P}_1, \mathbb{P}_2) = 1$, has at most $\sqrt{n}$ variables.*

Proof. Suppose it has more than $\sqrt{n}$ variables. Then, there is at least one color in $\mathbb{P}_1$ which contributes two variables to B. These two variables have to be in two different colors of $\mathbb{P}_2$ (because of our design of $\mathbb{P}_1$ and $\mathbb{P}_2$). So, $d_B(\mathbb{P}_1, \mathbb{P}_2)$ is at least 2. We get a contradiction.

The number of such base sets has to be at least $\sqrt{n}$. Combining this with the reduction from m-base-sets-δ-distance to $m\delta$-base-sets-1-distance, we get $m\delta = \Omega(\sqrt{n})$.

It is not clear if Lemma 43 is tight. We conjecture that for any set of partitions, $m\delta = O(\sqrt{n})$ can be achieved.