Separation between Read-once Oblivious Algebraic Branching Programs (ROABPs) and Multilinear Depth Three Circuits

Neeraj Kayal
Microsoft Research India
neeraka@microsoft.com

Vineet Nair
Indian Institute of Science
vineet.nair@csa.iisc.ernet.in

Chandan Saha
Indian Institute of Science
chandan@csa.iisc.ernet.in

September 22, 2015

Abstract

We show an exponential separation between two well-studied models of algebraic computation, namely read-once oblivious algebraic branching programs (ROABPs) and multilinear depth three circuits. In particular we show the following:

1. There exists an explicit n-variate polynomial computable by linear sized multilinear depth three circuits (with only two product gates) such that every ROABP computing it requires $2^{\Omega(n)}$ size.
2. Any multilinear depth three circuit computing $\mathrm{IMM}_{n,d}$ (the iterated matrix multiplication polynomial formed by multiplying d, $n \times n$ symbolic matrices) has $n^{\Omega(d)}$ size. $\mathrm{IMM}_{n,d}$ can be easily computed by a $\mathrm{poly}(n, d)$ sized ROABP.
3. Further, the proof of 2 yields an exponential separation between multilinear depth four and multilinear depth three circuits: There is an explicit n-variate, degree d polynomial computable by a $\mathrm{poly}(n)$ sized multilinear depth four circuit such that any multilinear depth three circuit computing it has size $n^{\Omega(d)}$. This improves upon the quasi-polynomial separation of [RY09] between these two models.

The hard polynomial in 1 is constructed using a novel application of expander graphs in conjunction with the evaluation dimension measure [Raz09, Raz06, RY09, FS13], while 2 is proved via a new adaptation of the dimension of the partial derivatives measure of [NW97]. Our lower bounds hold over any field.

1 Introduction

Proving lower bounds and separating complexity classes lie at the heart of complexity theory. In algebraic complexity, separating classes VP and VNP (the algebraic analogues of P and NP) equates to proving super-polynomial lower bounds for arithmetic circuits. Another prominent and pertinent problem is polynomial identity testing (PIT)[1], one of the very few natural problems in BPP (in fact, in co-RP) not known to be in P. Showing arithmetic circuit lower bounds and derandomizing PIT[2] are closely related: [KI04] showed that a polynomial time PIT over integers implies a super-polynomial arithmetic circuit lower bound for the family of permanent polynomials or NEXP $\not\subseteq$ P/poly. [HS80, Agr05] showed that a polynomial time blackbox[3] PIT implies exponential lower bounds for circuits computing polynomials whose coefficients can be computed in PSPACE. Conversely, [KI04] also showed that a super-polynomial (exponential) circuit lower bound[4] implies a sub-exponential (quasi-polynomial) time algorithm for PIT, in fact blackbox PIT, using Nisan-Wigderson generators [NW94] and Kaltofen's [Kal89] polynomial factorization algorithm. [DSY09] showed a similar connection between lower bounds and PIT for low depth circuits[5]. So, in this certain sense the complexity of proving strong lower bounds and devising efficient PIT algorithms are quite similar. Derandomizing PIT is also interesting in its own right. It is well-known that such a derandomization would imply the problem of checking existence of a perfect matching in a given graph is in NC [Tut47].

Research over the the past several years has made notable progress[6] on both lower bounds and PIT for interesting special cases of arithmetic circuits and helped identify the frontiers of our current knowledge. In particular, we understand better the reason why super-polynomial lower bounds and poly-time PIT have remained elusive even for depth three circuits: An exponential lower bound (similarly, a poly-time blackbox PIT) for depth three circuits over fields of characteristic zero implies an exponential lower bound (similarly, quasi-polynomial-time PIT) for general circuits [GKKS13].

A potentially useful and interesting restriction to consider at depth three is multilinearity[7] (meaning, every product gate computes a multilinear polynomial). We do know of strong lower bounds for multilinear depth three circuits due to [RY09] and also this paper (theorem 4), but as yet no efficient (meaning, quasi-polynomial) PIT is known for this model. One reason for this is the absence of hardness versus randomness tradeoff results for bounded depth multilinear circuits. Recently, [dOSV15] has given a sub-exponential time blackbox PIT algorithm for multilinear depth three circuits using recently found quasi-polynomial blackbox PIT for another model, namely read-once oblivious algebraic branching programs (ROABPs) [FS13, AGKS15] (Definition 2), thereby con-

[1] PIT is the following problem: given an arithmetic circuit computing a multivariate polynomial over some field, determine whether the polynomial is identically zero.

[2] a polynomial time randomized PIT follows easily from [DL78, Sch80, Zip79].

[3] meaning, we are only allowed to evaluate the circuit at points from $\mathbb{F}^n$, where n is the number of inputs and $\mathbb{F}$ the underlying field

[4] for any family of exponential-time computable multilinear polynomials

[5] lower bounds for bounded depth circuits imply efficient PIT for bounded depth circuits computing polynomials with low individual degree [DSY09]

[6] refer to the surveys [SY10], [CKW11], [Sap14, KS14], [Sax09, Sax13]

[7] Most of the hard polynomials used in the literature are multilinear, e.g. determinant, permanent, iterated matrix multiplication, Nisan-Wigderson polynomials etc. So, it is worthwhile to develop a fuller understanding of multilinear models [Raz09, Raz06, RY08, RY09, DMPY12].

necting these two interesting models of computation. Could there be a more efficient reduction from multilinear depth three circuits to ROABPs? If so then that would readily imply an efficient PIT algorithm for multilinear depth three circuits. This question has lead us to this work.

Related work and motivation. The model ROABP (see Definition 2) has been studied intensely in the recent years in the context of black-box PIT, equivalently *hitting-set generators*[8] (Definition 11). This has resulted in deterministic, quasi-polynomial time hitting-set generators for ROABPs [AGKS15, FS13] and other associated models like set-multilinear algebraic branching programs [FSS14, FS13] (a special case of which is set-multilinear depth three circuits [ASS13, FS13]), non-commutative algebraic branching programs [FS13] and diagonal depth-3 circuits [ASS13, FS13]. Quite recently, [dOSV15] has given a $2^{\tilde{O}(n^{\frac{2}{3}(1+\delta)})}$ time hitting-set generator for multilinear depth three circuits of size at most 2^{n^δ} by 'reducing' a multilinear depth three circuit to a collection of ROABPs and 'putting together' the hitting-sets of the ROABPs. This 'putting together' process raises the hitting-set complexity from quasi-polynomial (for a single ROABP) to sub-exponential (for a composition of several ROABPs). Had it been the case that a multilinear depth three circuit can be directly reduced to a single small size ROABP, an efficient hitting set for the former would have ensued immediately from [AGKS15, FS13]. One of the results in the paper (theorem 2), rules out this possibility. In fact, theorem 2 shows something stronger as described below.

A closer look at [AGKS15] and [dOSV15] reveals an interesting, and potentially useful, intermediate model that we call *superposition of (two or more) set-multilinear depth three circuits* (see Definition 5). An example of superposition of two set-multilinear depth three circuits is,

$$C(X,Y) = (1+3x_1+5y_2)(4+x_2+y_1) + (6+9x_1+4y_1)(2+5x_2+3y_2).$$

The variable sets $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2\}$ are completely disjoint and are called the *base sets* of $C(X,Y)$. When projected on X variables (i.e after putting the Y variables to zero), $C(X,Y)$ is a set-multilinear depth three circuit in the X variables. A similar thing is true for the Y variables. Thus, every base set is associated with a set-multilinear depth three circuit and vice versa. Any multilinear depth three circuit can be trivially viewed as a superposition of n set-multilinear depth three circuits with single variable in every base set, where n is the number of variables. A crucial observation in [dOSV15] is that every multilinear depth three circuit is *"almost"* a superposition of n^ε set-multilinear depth three circuits for some $\varepsilon < 1$, and further the associated n^ε base sets can be found in sub-exponential time using k-wise independent hash functions. Once we know the $r = n^\varepsilon$ base sets corresponding to r set-multilinear depth three circuits whose superposition forms a circuit of size s, finding a hitting set for the circuit in time $s^{r.\log s}$ follows easily by taking a direct product of hitting sets for r many ROABPs (in fact, set-multilinear depth three circuits). We think a useful model to consider at this juncture is superposition of constantly many set-multilinear depth three circuits with <u>*unknown*</u> base sets. In this case knowing the $r = O(1)$ base sets readily gives us a quasi-polynomial time hitting set, but finding these base sets from a given circuit is NP-hard for $r \geq 3$ (as we show in observation 1), which rules out the possibility of knowing the base-sets even if we are allowed to see the circuit (as in the *white-box* case). Indeed, even in this special case where the given multilinear depth three circuit is promised to be a superposition of

[8] A hitting-set generator for a class of circuits takes a size parameter s as input and outputs a set of points in $\mathbb{F}^n$ such that every circuit (from the class) of size bounded by s and computing a nonzero polynomial, evaluates to nonzero at one of the points

constantly many (say, 2) set-multilinear depth three circuits, the algorithm in [dOSV15] finds and works with many base sets and the resulting hitting set complexity grows to roughly $\exp(\sqrt{n})$. Could it be that superposition of constantly many set-multilinear depth three circuits efficiently reduce to ROABPs? Unfortunately, the answer to this also turns out to be negative as theorem 2 gives an explicit example of a superposition of *two* set-multilinear depth three circuit computing an n-variate polynomial f such that any ROABP computing f has width $2^{\Omega(n)}$.

While comparing two models (here multilinear depth three circuits and ROABPs), it is desirable to show a separation in both directions whenever an efficient reduction from one to the other seems infeasible. In this sense, we show a complete seperation between the models under consideration by giving an explicit polynomial computable by a polynomial sized ROABP such that every multilinear depth three circuit computing it requires exponential size. In fact, this explicit polynomial is simply the Iterated Matrix Multiplication $\mathrm{IMM}_{n,d}$ - the $(1,1)$-th entry of a product of d $n \times n$ symbolic matrices (theorem 4). $\mathrm{IMM}_{n,d}$ can be easily computed by a polynomial-sized ROABP (see observation 5). Although, a $2^{\Omega(d)}$ lower bound for multilinear depth three circuit computing Det_d is known [RY09], this does not imply a lower bound for $\mathrm{IMM}_{n,d}$ (despite the fact that Det and IMM are both complete for algebraic branching programs (ABPs) [MV97]) as the projection from IMM to Det can make the circuit non-multilinear. Another related work by [DMPY12] showed a separation between multilinear ABPs and multilinear formulas by exhibiting an explicit polynomial (namely, an *arc-full-rank* polynomial) that is computable by a linear size multilinear ABP but requires super-polynomial size multilinear formulas. But again multilinearity of a circuit can be lost when IMM is projected to arc-full-rank polynomials, and hence this result too does not imply a lower bound for IMM. An extension of theorem 4 to a super-polynomial lower bound for multilinear formulas computing IMM will have interesting consequences in separating noncommutative formulas and noncommutative ABPs. In a contemporary work [KST15], some of the authors of this work and Sébastien Tavenas have been able to show an $n^{\Omega(\sqrt{d})}$ lower bound for multilinear depth four circuits computing $\mathrm{IMM}_{n,d}$ by significantly extending a few of the ideas present in this work and building upon (thereby improving) the work of [FLMS14]. Thus, in summary the models poly-sized ROABPs and poly-sized multilinear depth three ciruits have provably different computational powers, although they share a non-trivial intersection as poly-sized set-multilinear depth three circuits is harbored in both.

An interesting outcome of the proof of the lower bound for multilinear depth three circuits computing IMM is an exponential separation between multilinear depth three and multilinear depth four circuits. Previously, [RY09] showed a super-polynomial separation between multilinear constant depth h and depth $h+1$ circuits, which when applied to the depth three versus four setting gives a quasi-polynomial seperation between the two models. In comparison, theorem 6 gives an exponential separation.

The models and our results. We define the relevant models and state our results now.

Definition 1 (Algebraic Branching Program)**.** *An Algebraic Branching Program(ABP) in the variables $X = \{x_1, x_2, ..., x_n\}$ is a directed acyclic graph with a source vertex s and a sink vertex t. It has $(d+1)$ sets or layers of vertices $V_1, V_2, ..., V_{d+1}$, where V_1 and V_{d+1} contain only s and t respectively. The width of an ABP is the maximum number of vertices in any of the $(d+1)$ layers. All the edges in an ABP are such that an edge starts from a vertex in V_i and is directed to a vertex in*

V_{i+1}, where V_i belongs to the set $\{V_1, V_2, ..., V_d\}$. The edges in an ABP are labelled by polynomials[9] *over a base field $\mathbb{F}$. The weight of the path between any two vertices u and v in an ABP is computed by taking the product of the edge labels on the path from u to v. An ABP computes the sum of the weights of all the paths from s to t.*

A special kind of ABP, namely ROABP, is defined in [FS13].

Definition 2 (Read-Once Oblivious Algebraic Branching Program)**.** *A Read-Once Oblivious Algebraic Branching Program(ROABP) over a field $\mathbb{F}$ has an associated permutation $\pi : [n] \to [n]$ of the variables in X. The number of variables is equal to the number of layers of vertices minus one, i.e. $n = (d+1) - 1 = d$. The label associated with an edge from a vertex in V_i to a vertex in V_{i+1} is an univariate polynomial over $\mathbb{F}$ in the variable $x_{\pi(i)}$.*

Definition 3 (Multilinear depth four and depth three circuits)**.** *A circuit $C = \sum_{i=1}^{s} \prod_{j=1}^{d_i} Q_{ij}(X_j^i)$ is a multilinear depth four ($\Sigma\Pi\Sigma\Pi$) circuit in X variables over a field $\mathbb{F}$, if $X = \uplus_{j=1}^{d_i} X_j^i$ and $Q_{ij} \in \mathbb{F}[X_j^i]$ is a multilinear polynomial for every $i \in [s]$ and $j \in [d_i]$. If Q_{ij}'s are linear polynomials then C is a multilinear depth three ($\Sigma\Pi\Sigma$) circuit. The parameter s is the* top fan-in *of C.*

Definition 4 (Set-multilinear depth three circuit)**.** *A circuit $C = \sum_{i=1}^{s} \prod_{j=1}^{d} l_{ij}(X_j)$ is a set-multilinear depth three ($\Sigma\Pi\Sigma$) circuit in X variables over a field $\mathbb{F}$, if $X = \uplus_{j=1}^{d} X_j$ and $l_{ij} \in \mathbb{F}[X_j]$ is a linear polynomial for every $i \in [s]$ and $j \in [d]$. The sets $X_1, X_2, ..., X_d$ are called the* colors *of X. If $|X_j| = 1$ for every $j \in [d]$ then we say X has singleton colors and C is a set-multilinear depth three circuit with singleton colors.*

As a bridge between multilinear and set-multilinear depth three circuits we define a model called superposition of set-multilinear depth three circuits.

Definition 5 (Superposition of set-multilinear depth three circuits)**.** *A multilinear depth three ($\Sigma\Pi\Sigma$) circuit C over a field $\mathbb{F}$ is a superposition of t set-multilinear depth three circuits over variables $X = \uplus_{i=1}^{t} Y_i$, if for every $i \in [t]$, C is a set-multilinear depth three circuit in Y_i variables over the field $\mathbb{F}(X \setminus Y_i)$. The sets $Y_1, ..., Y_t$ are called the* base sets *of C. Further, we restrict the Y_i to have singleton colors for every $i \in [t]$* [10]*.*

We make the following initial observation for superposition of set-multilinear depth three circuits.

Observation 1. *Given a circuit C which is a superposition of t set-multilinear circuits on <u>unknown</u> base sets $Y_1, Y_2, ..., Y_t$, finding t base sets $Y_1', Y_2', ..., Y_t'$ such that C is a superposition of t set-multilinear circuits on base sets $Y_1', Y_2', ..., Y_t'$ is NP-hard when $t > 2$.*

The proof of the observation appears in appendix A. We now state the main results of this paper.

Theorem 2 (Main Theorem 1)**.** *1. There is an explicit family of $2n$-variate polynomials $\{g_n\}_{n \geq 1}$ over any field $\mathbb{F}$ such that the following hold: g_n is computable by a multilinear depth three circuit C over $\mathbb{F}$ with top fanin <u>three</u> and C is also a superposition of <u>two</u> set-multilinear depth three circuits. Any ROABP over $\mathbb{F}$ computing g_n has width $2^{\Omega(n)}$.*

[9] in a standard definition of an ABP, the edges are labeled by linear polynomials

[10] although the notion of superposition makes sense even if Y_i's do not have singleton colors, we restrict to singletons as this model itself captures multilinear depth three circuits

2. *There is an explicit family of $3n$-variate polynomials $\{g_n\}_{n\geq 1}$ over any field $\mathbb{F}$ such that the following hold: g_n is computable by a multilinear depth three circuit C over $\mathbb{F}$ with top fanin two and C is also a superposition of three set-multilinear depth three circuits. Any ROABP over $\mathbb{F}$ computing f_n has width $2^{\Omega(n)}$.*

We prove theorem 2 in section 3. The tightness of the theorem is exhibited by this observation.

Observation 3. *A polynomial computed by a multilinear $\Sigma\Pi\Sigma$ circuit with top fan-in two and at most two variables per linear polynomial can also be computed by an ROABP with constant width.*

The proof of observation 3 is in appendix A. Thus, it follows from theorem 2 that if we increase either the top fan-in or the number of variables per linear polynomial from two to three in multilinear depth three circuits then there exist polynomials computed by such circuits such that ROABPs computing these polynomials have exponential width. We now state the "converse" of theorem 2.

Theorem 4 (Main Theorem 2)**.** *Any multilinear depth three circuit (over any field) computing $\mathrm{IMM}_{n,d}$, the $(1,1)$-th entry of a product of d $n \times n$ symbolic matrices, has top fan-in $n^{\Omega(d)}$ for $n \geq 11$. (Note: This also implies a lower bound for determinant, see corollary 27.)*

We prove theorem 4 in section 4. It is not hard to observe the following.

Observation 5. $\mathrm{IMM}_{n,d}$ *can be computed by an n^2 width ROABP.*

The proof of observation 5 is given in appendix A. Thus, theorem 2, theorem 4 and observation 5 together imply a complete separation between multilinear depth three circuits and ROABPs. As a consequence of the proof of theorem 4 we also get an exponential separation between multilinear depth three and multilinear depth four circuits (proof in section 4).

Theorem 6. *There is an explicit family of $O(n^2d)$-variate polynomials of degree d, $\{f_d\}_{d\geq 1}$, such that f_d is computable by a $O(n^2d)$-sized multilinear depth four circuit with top fan-in one (i.e. a $\Pi\Sigma\Pi$ circuit) and every multilinear depth three circuit computing f_d has top fan-in $n^{\Omega(d)}$ for $n \geq 11$.*

Observe that the hard polynomials used in theorem 2 belong to a special class of multilinear depth three circuits - they are both superpositions of constantly many set-multilinear depth three circuits and simultaneously a sum of constantly many set-multilinear depth three circuits. Here is an example of a circuit from this class.

$$C(X,Y) = (1+3x_1+5y_2)(4+x_2+y_1) + (9+6x_1+4y_2)(3+2x_2+5y_1)$$

$$+(6+9x_1+4y_1)(2+5x_2+3y_2) + (3+6x_1+9y_1)(5+8x_2+2y_2)$$

$C(X,Y)$ is a superposition of two set-multilinear depth three circuits with base sets $X = \{x_1\} \cup \{x_2\}$ and $Y = \{y_1\} \cup \{y_2\}$. But $C(X,Y)$ is also a sum of two set-multilinear depth three circuits with $\{x_1, y_2\}$, $\{x_2, y_1\}$ being the colors in the first set-multilinear depth three circuit (corresponding to the first two products) and $\{x_1, y_1\}$, $\{x_2, y_2\}$ the colors in the second set-multilinear depth three circuit (corresponding to the last two products). For such a subclass of multilinear depth three circuits, we give a quasi-polynomial time hitting set by extending the proof technique of [ASS13].

Theorem 7. *Let $\mathcal{C}_{n,m,l,s}$ be a subclass of multilinear depth three circuits computing n-variate polynomials such that every circuit in $\mathcal{C}_{n,m,l,s}$ is a superposition of at most m set-multilinear depth three circuits and simultaneously a sum of at most l set-multilinear depth three circuits, and has top fan-in bounded by s. There is a hitting-set generator for $\mathcal{C}_{n,m,l,s}$ running in $(ns)^{O(lm\log s)}$ time.*

When m and l are bounded by $\mathsf{poly}(\log ns)$, we get quasi-polynomial time hitting sets. The proof of theorem 7, which extends the shift and rank concentration technique of [ASS13], is given in appendix E. To our understanding, even if m and l are constants, [dOSV15]'s algorithm yields an $exp(\sqrt{n})$ hitting set complexity. Also, [GKST15] has recently given a $(ndw)^{O(l2^l \log(ndw))}$ time hitting set generator for n-variate, individual (variable) degree d polynomials computed by sum of l ROABPs each of width less than w. Sum of l set-multilinear depth three circuits reduces to sum of l ROABPs as set-multilinear depth three circuits readily reduce to poly-sized ROABPs. But, observe the doubly exponential dependence on l in their result. On the contrary, in theorem 7 the dependence is singly exponential in l. So, the hitting-set complexity remains quasi-polynomial for $l = (\log n)^{O(1)}$ whereas [GKST15] gives an exponential time hitting-set generator when applied to the model in theorem 7. However, it is also important to note that the model considered in theorem 7 is somewhat weaker than the sum of ROABPs model in [GKST15] because of the additional restriction that our model is also a superposition of m set-multilinear depth three circuits.

Proof ideas for theorem 2 and 4. Theorem 2 is proved by connecting the notion of edge expansion (definition 8) with the evaluation dimension measure (definition 6). Starting with an explicit 3-regular bipartite expander G, we associate distinct variables with distinct vertices. Every edge now corresponds to a linear polynomial – it is the sum of the variables associated with the vertices on which the edge is incident upon. A multilinear depth three circuit C is derived from the expander G as follows: C has three product terms, each term formed by taking product of the linear polynomials associated with the edges of a matching in G. Now, edge expansion of G can be used to argue that for every subset S of variables of a certain size there exists of a product term in C that has high evaluation dimension with respect to S. Further, one can show that high evaluation dimension of a product term implies high evaluation dimension of C with respect to S by restricting the circuit modulo two linear polynomials to nullify the other two product terms. On the other hand, for every ROABP there is a set S (of any size) such that the evaluation dimesion of the ROABP with respect to S is bounded by its width. This gives a lower bound on the width of the ROABP computing the same polynomial as C thereby proving part 1 of theorem 2. Part 2 is proved similarly, but now we associate edges and vertices of a bipartite expander G with variables and linear polynomials respectively. Circuit C is formed by adding two product terms, each term formed by multiplying the linear polynomials associated with the left or the right vertex set of G. As before, edge expansion of C implies for every set S of variables of a certain size there is a product term of C with high evaluation dimension and this in turn implies high evaluation dimension of C.

While writing this article, we came to know about a recent work by Jukna [Juk15] that uses Ramanujan graphs to give an alternate proof of a known exponential lower bound for monotone arithmetic circuits. To our understanding, it does seem that Jukna's proof also implictly relates expansion with evaluation dimension, but the argument in [Juk15] is directed towards monotone circuits and it does not seem to imply any of the lower bounds shown in this work. In particular, the hard polynomial in [Juk15] could have any complexity, whereas in our case we need the hard polynomial to be computable by a small multilinear depth three circuit.

Theorem 4 is proved by introducing a new variant of the dimension of the space of partial derivatives measure that is inspired by both [NW97] and [Raz09]. At a high level, the idea is to consider a polynomial f in two sets of variables X and Y such that $|Y| >> |X|$ is large. If we take derivatives

of f with respect to all degree k monomials in Y-variables and set all the Y-variables to zero after taking derivatives then we do expect to get a 'large' space of derivatives (especially, when f is a 'hard' polynomial) simply because $|Y|$ is large. On the other hand, in any depth three multilinear circuit C computing f, the dimension of the space of derivatives of a product term is influenced only by the number of linear polynomials containing the X-variables as all the Y-variables are set to zero subsequently. Thus, the measure is somewhat small for a product term of C as $|X| << |Y|$. By subadditivity of the measure (lemma 8), this implies high top fan-in of C computing f. A notable difference with [Raz09, RY09] is that the variable sets X and Y are fixed deterministically, a priori, and not by random partitioning of the entire set of variables.

2 Preliminaries

Measures. We have used two complexity measures, namely evaluation dimension and a novel variant of the dimension of the space of partial derivatives, to prove theorem 2 and 4 respectively. Evaluation dimension was first defined in [FS13][11]. Let X be a set of variables.

Definition 6 (Evaluation Dimension)**.** *The evaluation dimension of a polynomial $g \in \mathbb{F}[X]$ with respect to a set $S \subseteq X$, denoted as* $\text{Evaldim}_S(g)$*, is defined as*

$$\dim(\text{span}_{\mathbb{F}}\{g(X)|_{\forall x_j \in S \ x_j=\alpha_j} : \forall x_j \in S \ \alpha_j \in \mathbb{F}\}).$$

Evaluation dimension is a nearly equivalent variant of another measure, the *rank of the partial derivatives matrix*, defined and used earlier in [Raz09, Raz06, RY08, RY09, DMPY12] to prove lower bounds and separations for several multilinear models. These two measures are identical over fields of characteristic zero (or sufficiently large), but the former is well defined over any field.

The partial derivatives measure was introduced in [NW97]. The following is a simple variant of this measure that is also inspired by the measure used in [Raz09].

Definition 7 ("Skewed" partial derivatives)**.** *Let $f \in \mathbb{F}[X, Y]$, where X and Y are disjoint sets of variables, and $\mathcal{Y}_k$ be the set of all monomials in Y variables of degree $k \in \mathbb{N}$. Define the measure* $\mathsf{PD}_{\mathcal{Y}_k}(f)$ *as*

$$\dim\left(\text{span}_{\mathbb{F}}\left\{\left[\frac{\partial f(X,Y)}{\partial m}\right]_{\forall y \in Y \ y=0} : m \in \mathcal{Y}_k\right\}\right).$$

In proving theorem 4, we apply the above measure with a significant difference (or *skew*) between the number of X and Y variables – it is this imbalance that plays a crucial role in the proof. Both the above measures obey the property of subadditivity (proof in appendix B).

Lemma 8 (Subadditivity)**.** *1. Let $g_1, g_2 \in \mathbb{F}[X]$ and $S \subseteq X$. Then*

$$\text{Evaldim}_S(g_1 + g_2) \leq \text{Evaldim}_S(g_1) + \text{Evaldim}_S(g_2).$$

2. Let $f_1, f_2 \in \mathbb{F}[X, Y]$. Then $\mathsf{PD}_{\mathcal{Y}_k}(f_1 + f_2) \leq \mathsf{PD}_{\mathcal{Y}_k}(f_1) + \mathsf{PD}_{\mathcal{Y}_k}(f_2)$.

[11] they attributed the notion to Ramprasad Saptharishi

Expander Graphs. A vital ingredient that helps us construct the hard polynomials in theorem 2 is a family of explicit 3-regular expanders. We recall a few basic definitions from [HLW06].

Definition 8 (Edge expansion and family of expanders)**.** *Let $G = (V, E)$ be an undirected d-regular graph. For $S \subseteq V$, let $E(S, \overline{S})$ be the set of edges with one end incident on a vertex in S and the other incident on a vertex in $\overline{S} = V \backslash S$. The* edge expansion *of G denoted $h(G)$ is defined as:*

$$h(G) = \min_{S:\ |S| \leq \frac{|V|}{2}} \frac{|E(S, \overline{S})|}{|S|}.$$

A sequence of d-regular graphs $\{G_i\}_{i \in \mathbb{N}}$ of size increasing with i is a family of d-regular expanders *if there exists an $\varepsilon > 0$ such that $h(G_i) > \varepsilon$ for every i.*

Definition 9 (Mildly explicit expanders)**.** *Let $\mathcal{G} = \{G_i\}_{i \in \mathbb{N}}$ be a family of d-regular expanders such that the number of vertices in G_i is bounded by a polynomial in i. $\mathcal{G}$ is* mildly explicit *if there exists an algorithm that takes input i and constructs G_i in time polynomial in the size of G_i.*

A family of mildly explicit expanders. [HLW06] mentions a family of mildly explicit 3-regular p-vertex expanders $\{G_p\}_{p \text{ prime}}$ such that for every graph G_p in the family: $h(G_p) > \frac{2+10^{-4}}{2}$. The vertices of G_p correspond to elements in $\mathbb{Z}_p$. A vertex x in G_p is connected to $x + 1$, $x - 1$ and to its inverse x^{-1} (operations are modulo p and inverse of 0 is defined as 0). We refer the reader to [HLW06], section 11.1.2, for more details. Denote this family of 3-regular p-vertex expanders by $\mathcal{S}$.

Double Cover. The proof of theorem 2 works with bipartite expanders. It is standard to transform a d-regular expander graph to a d-regular bipartite expander graph by taking its double cover.

Definition 10 (Double Cover)**.** *The double cover of a graph $G = (V, E)$ is the bipartite graph $\tilde{G} = (L \uplus R, \tilde{E})$ where $|L| = |R| = |V|$. Corresponding to a vertex $u \in V$ we have two vertices $u_L \in L$ and $u_R \in R$. Edges (u_L, v_R) and $(u_R, v_L) \in \tilde{E}$ if and only if there is an edge $(u, v) \in E$.*

Lemma 9. *Let $\mathcal{S} = \{G_p\}_{p \text{ prime}}$ be the family of expanders as described above, and $\tilde{\mathcal{S}} = \{\tilde{G}_p\}_p$ the family of double covers of graphs in $\mathcal{S}$. Then $h(\tilde{G}_p) > \frac{2+10^{-4}}{2}$ for every p.* [Proof in appendix B.]

Hitting-set generators. In theorem 7, we give a quasi-polynomial time hitting-set generator for a subclass of multilinear depth three circuits.

Definition 11 (Hitting-set generators)**.** *A hitting-set generator for a class of circuits $\mathcal{C}$ is a Turing machine $\mathcal{H}$ that takes $(1^n, 1^s)$ as input and outputs a set $\{a_1, \ldots, a_m\} \subseteq \mathbb{Z}^n$ such that for every circuit $C \in \mathcal{C}$ of size bounded by s and computing a nonzero n-variate polynomial over a field $\mathbb{F} \supset \mathbb{Z}$, there is an $i \in [m]$ for which $C(a_i) \neq 0$. Complexity of $\mathcal{H}$ is its running time.* [12]

Technical Lemmas. The following lemmas are used in theorem 2. Lemma 10 follows from Hall's marriage theorem [Hal35]. The proofs of lemmas 11 and 12 are given in appendix B.

Lemma 10. *A d-regular graph can be split into d edge disjoint perfect matchings.*

[12] Hitting-set generators can be defined similarly over finite fields by considering field extensions

Lemma 11. *Suppose $g_1(X), g_2(X), ..., g_m(X) \in \mathbb{F}[X]$ are $\mathbb{F}$-linearly independent polynomials in the variables $X = \{x_1, x_2, ..., x_n\}$ where $m = 2^n$. If $Y = \{y_1, y_2, ..., y_n\}$ are n variables different from X then (by identifying an $i \in [m]$ with an $S \subseteq [n]$),*

$$\text{Evaldim}_Y(\sum_{S\subseteq[n]} y_S \cdot g_S(X)) = m, \quad \text{where for } S \subseteq [n],\ y_S := \prod_{j\in S} y_j.$$

Lemma 12. *If R is a width-k ROABP that computes $g(X)$ then for every $i \in [0, |X|]$ there exists a set $S \subseteq X$ of size i such that* $\text{Evaldim}_S(g) \leq k$.

3 Lower bounds for ROABP: Proof of theorem 2

Proof of part 1

Construction of the polynomial family. We construct a family of $2n$-variate multilinear polynomials $\{g_n\}_{n\geq 1}$ from the explicit family of 3-regular expander graphs $\mathcal{S}$ (described in section 2). From an n-vertex graph $G = (V, E)$ in $\mathcal{S}$, construct a polynomial $g(X, Y)$ in variables $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_n\}$ as follows: Let $\tilde{G} = (L \uplus R, \tilde{E})$ be the double cover of G. By lemma 9, $h(\tilde{G}) > \frac{2+10^{-4}}{2}$. With every vertex in L (similarly, R) associate a unique variable in X (respectively, Y), thus vertices in L and R are identified with the X and Y variables respectively. An edge between x_i and y_j is associated with the linear polynomial $(1 + x_i + y_j)$. By lemma 10, $\tilde{G}$ can be split into three edge disjoint perfect matchings. Polynomial $g(X, Y)$ is a sum of three product terms corresponding to the three edge disjoint perfect matchings of $\tilde{G}$; a product term is formed by taking product of the linear polynomials associated with the edges of the corresponding matching. It is easy to show the following claim (proof given in appendix C).

Claim 13. *Polynomial g (constructed above) is computed by a multilinear depth three circuit C of size $\Theta(n)$ and top fan-in three, and C is a superposition of two set-multilinear depth three circuits.*

High evaluation dimension of $g(X, Y)$. It turns out that the evaluation dimension of $g(X, Y)$ with respect to any subset of variables of size $n/10$ is large.

Lemma 14. *For any set $S \subseteq X \uplus Y$ of size $n/10$,* $\text{Evaldim}_S(g) \geq 2^{\varepsilon n}$ *where $\varepsilon > 0$ is a constant.*

Proof. Consider any subset S of $n/10$ variables from $X \uplus Y$. With respect to set S we can classify the linear polynomials in the product terms of $g(X, Y)$ into three types: *untouched* - if none of the two variables in the linear polynomial belong to S, *partially touched* - if exactly one of the variables in the linear polynomial belongs to S, and *completely touched* - if both variables belong to S. Call the three product terms of g – P_1, P_2 and P_3. Proof of the next claim appears in appendix C.

Claim 15. *There exists a set $X_0 \subseteq X$ of $\left(\frac{7n}{10} - 4\right)$ X-variables such that every $x \in X_0$ appears in an untouched linear polynomial in every P_i (for $i \in [3]$), and further if $(1+x+y_{j_1}), (1+x+y_{j_2})$ and $(1+x+y_{j_3})$ are the linear polynomials occurring in P_1, P_2 and P_3 respectively then $y_{j_1} \neq y_{j_2} \neq y_{j_3}$.*

For $i \in [3]$, let B_i be the set of partially touched linear polynomials in term P_i.

Claim 16. *There is an $i \in [3]$ such that $|B_i| \geq \varepsilon n$ where $\varepsilon = 0.01$.*

Proof. Let T be such that, for all $i \in [3]$, $|B_i| \leq T$. Recall that g has been constructed from the bipartite expander $\tilde{G}$, and vertices in $\tilde{G}$ identified with the variable set $X \uplus Y$. We denote the vertices in $\tilde{G}$ corresponding to the variables in S also by S, and denote the set of edges going out from S to $\overline{S} = L \uplus R \backslash S$ in $\tilde{G}$ by $\tilde{E}(S, \overline{S})$. Using the expansion property of $\tilde{G}$,

$$|\tilde{E}(S, \overline{S})| \geq h(\tilde{G}) \cdot |S| \geq \frac{2 + 10^{-4}}{2} \cdot \left(\frac{n}{10}\right).$$

Every edge in $\tilde{E}(S, \overline{S})$ corresponds to a partially touched linear polynomial. Since $\tilde{G}$ is 3-regular, at least $\frac{|\tilde{E}(S,\overline{S})|}{3}$ of the edges correspond to distinct partially touched linear polynomials. By assumption, the number of such partially touched linear polynomials is at most $3T$; and so $T \geq 0.01n$. □

The next claim completes the proof of lemma 14.

Claim 17. *If there exists an $i \in [3]$ such that $|B_i| \geq \varepsilon n$ for $\varepsilon > 0$, then* $\text{Evaldim}_S(g) \geq 2^{\varepsilon n}$.

Proof. Without loss of generality, assume $|B_1| \geq \varepsilon n$. Pick two variables, say x and x', from the set X_0 (as described in claim 15). Let $(1 + x + y_{j_2})$ and $(1 + x' + y'_{j_3})$ be the linear polynomials appearing in P_2 and P_3 respectively. By substituting $x = -(1 + y_{j_2})$ and $x' = -(1 + y'_{j_3})$ in g, the terms P_2 and P_3 vanish but P_1 does not (by claim 15). Let $\hat{g}$ be the polynomial g after the substitution. Polynomial $\hat{g}$ has only one product term $\hat{P}_1$ (i.e. P_1 under the substitution), and $\hat{P}_1$ has as many partially touched linear polynomials as P_1. At this point, it is not difficult to prove the following observation. (Proof given in appendix C.)

Observation 18. $\text{Evaldim}_S(g) \geq \text{Evaldim}_S(\hat{g}) = \text{Evaldim}_S(\hat{P}_1) \geq 2^{\varepsilon n}$.

This completes the proof of claim 17. □

From lemma 12 and 14 we conclude that any ROABP computing $g(X, Y)$ has width at least $2^{\varepsilon n}$. □

Proof of part 2

Construction of the polynomial family. Similar to part 1, we construct a family of $3n$-variate multilinear polynomials $\{g_n\}_{n \geq 1}$ from the explicit family of 3-regular expanders $\mathcal{S}$ – but this time edges will be associated with variables and vertices with linear polynomials. From an n-vertex graph $G = (V, E)$ in $\mathcal{S}$, construct a polynomial $g(X, Y, Z)$ in variables $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_n\}$ and $Z = \{z_1, \ldots, z_n\}$ as follows: Let $\tilde{G} = (L \uplus R, \tilde{E})$ be the double cover of G, and as before $h(\tilde{G}) > \frac{2+10^{-4}}{2}$. Edges of $\tilde{G}$ can be split into three edge disjoint perfect matchings (by lemma 10). Label the edges of the first perfect matching by distinct X-variables, the edges of the second matching by distinct Y-variables, and the edges of the third by distinct Z-variables. Vertices of $\tilde{G}$ now correspond to linear polynomials naturally – if the three edges incident on a vertex are labelled x_i, y_j and z_k then associate the linear polynomial $(1 + x_i + y_j + z_k)$ with the vertex. Let P_1 be the product of the linear polynomials associated with the vertices of L, and P_2 the product of linear polynomials associated with the vertices of R. Polynomial $g(X, Y, Z)$ is the sum of P_1 and P_2. The following claim is easy to show (just like claim 13).

Claim 19. *Polynomial g (constructed above) is computed by a multilinear depth three circuit C of size $\Theta(n)$ and top fan-in two, and C is a superposition of three set-multilinear depth three circuits.*

High evaluation dimension of $g(X, Y)$. The proof of the following lemma is similar to that of lemma 14, differences arise only due to the 'dual' nature of g.

Lemma 20. *For any $S \subseteq X \uplus Y \uplus Z$ of size $n/10$, $\text{Evaldim}_S(g) \geq 2^{\varepsilon n}$ where $\varepsilon > 0$ is a constant.*

Proof. Let S be any set of $\frac{n}{10}$ variables from $X \uplus Y \uplus Z$. The definitions of untouched, partially touched and completely touched linear polynomials are almost the same as in the proof of lemma 14. The difference is we have three variables instead of two in a linear polynomial in g. So, a linear polynomial is partially touched if at most two of the three variables belong to S. For $i \in [2]$, let B_i be the set of partially touched linear polynomials and C_i the set of completely touched linear polynomials in product term P_i of g.

Claim 21. *There is an $i \in [2]$ such that $|B_i| \geq \varepsilon n$ where $\varepsilon = 0.01$.*

Proof. Let T be such that, for all $i \in [2]$, $|B_i| \leq T$. We show the next observation in appendix C.

Observation 22. *$|C_1| + |C_2|$ is at least $\frac{n}{15} - \frac{8T}{3}$.*

Let C be the set of vertices in $\tilde{G}$ corresponding to the completely touched linear polynomials in both the product gates, thus $|C| = |C_1| + |C_2| \geq \frac{n}{15} - \frac{8T}{3}$. Each edge in $\tilde{E}(C, \overline{C})$ connects a vertex which corresponds to a completely touched linear polynomial to a vertex which corresponds to a partially touched linear polynomial. Using expansion of $\tilde{G}$,

$$|\tilde{E}(C, \overline{C})| \geq h(\tilde{G}) \cdot |C| \geq \frac{2 + 10^{-4}}{2} \cdot \left(\frac{n}{15} - \frac{8T}{3} \right).$$

Since edges in $\tilde{E}(C, \overline{C})$ are associated with variables in S, a vertex corresponding to a partially touched linear polynomial has at most two edges from $\tilde{E}(C, \overline{C})$ incident on it. Hence the number of vertices corresponding to partially touched linear polynomials is at least $\frac{\tilde{E}(C, \overline{C})}{2}$. But, by assumption, the number of such vertices is at most $2T$. Thus,

$$2T \geq \frac{|\tilde{E}(C, \overline{C})|}{2} \geq \frac{2 + 10^{-4}}{4} \cdot \left(\frac{n}{15} - \frac{8T}{3} \right) \quad \Rightarrow \quad T \geq 0.01n.$$

□

The proof of the next claim is much like that of claim 17 and is given in appendix C.

Claim 23. *If there exists an $i \in [2]$ such that $|B_i| \geq \varepsilon n$ for $\varepsilon > 0$, then $\text{Evaldim}_S(g) \geq 2^{\varepsilon n}$.*

This completes the proof of lemma 20. From lemma 12 and 20 we conclude that any ROABP computing g has width at least $2^{\varepsilon n}$. □

4 Lower bounds for multilinear depth three circuits

The proofs of theorems 4 and 6 are inspired by a particular kind of *projection* of $\text{IMM}_{n,d}$ considered in [FLMS14]. We say a polynomial f is a *simple projection* of another polynomial g if f is obtained by simply setting some variables to field constants in g.

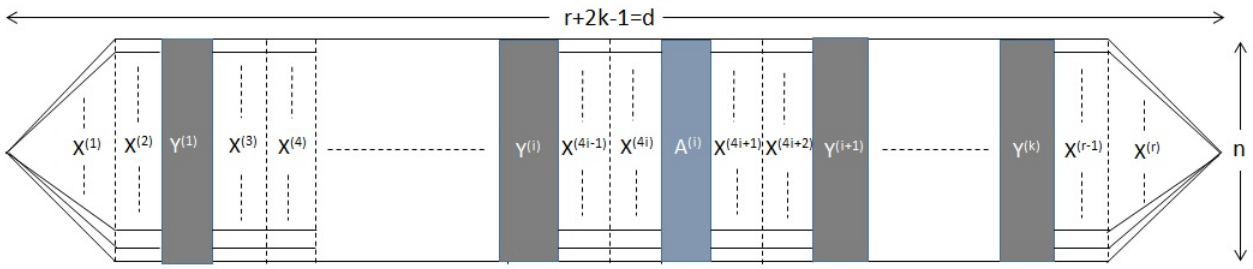

Figure 1: ABP $\mathcal{M}$

Proof of theorem 4: The proof proceeds by constructing an ABP $\mathcal{M}$ of width n and with $d+1$ layers of vertices such that (a) the polynomial computed by $\mathcal{M}$, say f, is a simple projection of $\text{IMM}_{n,d}$, and (b) any multilinear depth three circuit computing f has top fan-in $n^{\Omega(d)}$. Since an ABP can be viewed equivalently as a product of matrices, we will describe $\mathcal{M}$ using matrices. Figure 1 depicts the ABP $\mathcal{M}$.

Description of $\mathcal{M}$. The polynomial f, computed by $\mathcal{M}$, is defined over two disjoint sets of variables, X and Y. The Y variables are contained in k matrices, $\{Y^{(1)}, ..., Y^{(k)}\}$; the (u,v)-th entry in $Y^{(i)}$ is a formal variable $y^{(i)}_{u,v}$. There are $(k-1)$ matrices $\{A^{(1)}, ..., A^{(k-1)}\}$, such that all the entries in these matrices are ones. The X variables are contained in r matrices, $\{X^{(1)}, ..., X^{(r)}\}$. Matrices $X^{(1)}$ and $X^{(r)}$ are row and column vectors of size n respectively. The u-th entry in $X^{(1)}$ (similarly, $X^{(r)}$) is $x^{(1)}_u$ (respectively, $x^{(r)}_u$). All the remaining matrices $\{X^{(2)}, ..., X^{(r-1)}\}$ are diagonal matrices in the X variables, i.e. the (u,u)-th entry in $X^{(i)}$ is $x^{(i)}_u$ and all other entries are zero for $i \in [2, r-1]$. The matrices are placed as follows: Between two adjacent Y matrices, $Y^{(i)}$ and $Y^{(i+1)}$, we have five matrices ordered from left to right as $X^{(4i-1)}, X^{(4i)}, A^{(i)}, X^{(4i+1)}$ and $X^{(4i+2)}$ for every $i \in [1, k-1]$. Ordered from left to right, $X^{(1)}$ and $X^{(2)}$ are on the left of $Y^{(1)}$ and $X^{(r-1)}$ and $X^{(r)}$ are on the right of $Y^{(k)}$. Naturally, we have the following relation among k, r and d: $r = 4k$ and $d = r+2k-1$, i.e. $k = \frac{d+1}{6}$. Thus $|X| = nr = 4nk$ and $|Y| = n^2k$. This imbalance between the X and Y variables plays a vital role in the proof. As before, call the polynomial computed by this ABP $\mathcal{M}$ as $f(X,Y)$.

The following claim is easy to verify as f is a simple projection of $IMM_{n,d}$.

Claim 24. *If* $\text{IMM}_{n,d}$ *is computed by a multilinear depth three circuit having top fan-in* s *then* f *is also computed by a multilinear depth three circuit having top fan-in* s. [Proof in appendix D.]

We show every multilinear depth three circuit computing f has top fan-in $n^{\Omega(d)}$ for $n \geq 11$.

Lower bounding $\mathsf{PD}_{\mathcal{Y}_k}(f)$. Let $\tilde{\mathcal{Y}}_k \subseteq \mathcal{Y}_k$ be the set of monomials formed by picking exactly one Y-variable from each of the matrices $Y^{(1)}, ..., Y^{(k)}$ and taking their product. Then, $|\tilde{\mathcal{Y}}_k| = n^{2k}$.

Claim 25. $\mathsf{PD}_{\mathcal{Y}_k}(f(X,Y)) = |\tilde{\mathcal{Y}}_k| = n^{2k}$.

Proof. The derivative of f with respect to a monomial $m \in \mathcal{Y}_k$ is nonzero if and only if $m \in \tilde{\mathcal{Y}}_k$. Also, such a derivative $\frac{\partial f}{\partial m}$ is a multilinear degree-r monomial in X-variables. The derivatives of f with respect to two distinct monomials m and m' in $\tilde{\mathcal{Y}}_k$ give two distinct multilinear degree-r monomials in X-variables. Hence, $\mathsf{PD}_{\mathcal{Y}_k}(f) = |\tilde{\mathcal{Y}}_k|$. □

Upper bounding $\mathsf{PD}_{\mathcal{Y}_k}$ of a multilinear depth three circuit.

Lemma 26. *Let C be a multilinear depth three circuit having top fan-in s computing a polynomial in X and Y variables. Then $\mathsf{PD}_{\mathcal{Y}_k}(C) \leq s \cdot (k+1) \cdot \binom{|X|}{k}$ if $k \leq \frac{|X|}{2}$.*

Proof. Let $C = \sum_{i=1}^{s} T_i$, where each T_i is a product of linear polynomials on disjoint sets of variables. From lemma 8, $\mathsf{PD}_{\mathcal{Y}_k}(C) \leq s \cdot \max_{i\in[s]} \mathsf{PD}_{\mathcal{Y}_k}(T_i)$. We need to upper bound the dimension of the "skewed" partial derivatives of a term $T_i = T$ (say). Let $T = \prod_{j=1}^{q} l_j$, where l_j is a linear polynomial. Among the q linear polynomials at most $|X|$ of them contain the X variables. Without loss of generality, assume the linear polynomials $l_1, \ldots, l_p$ contain X-variables and the remaining $l_{p+1}, \ldots, l_q$ are X-free (here $p \leq |X|$). Let $Q = \prod_{j=p+1}^{q} l_j$. Then, $T = Q \cdot \prod_{j=1}^{p} l_j$. We take the derivative of T with respect to a monomial $m \in \mathcal{Y}_k$ and then substitute the Y variables to zero. Applying the product rule of differentiation and observing that the derivative of a linear polynomial with respect to a variable makes it a constant we have the following:

$$\left[\frac{\partial T}{\partial m}\right]_{Y=\bar{0}} = \sum_{\substack{S\subseteq[p]\\|S|\leq k}} \alpha_S \prod_{j\in[p]\setminus S} [l_j]_{Y=\bar{0}}$$

where α_S's are constants from the field. Here m is a representative element of the set $\mathcal{Y}_k$. Hence every such derivative can be expressed as a linear combination of $\sum_{t=0}^{k}\binom{p}{t} \leq (k+1)\cdot\binom{|X|}{k}$ polynomials, where the last inequality is due to $k \leq \frac{|X|}{2}$ (if $t > p$ then $\binom{p}{t} \stackrel{\text{def}}{=} 0$). Therefore, $\mathsf{PD}_{\mathcal{Y}_k}(T) \leq (k+1)\cdot\binom{|X|}{k}$ and $\mathsf{PD}_{\mathcal{Y}_k}(C) \leq s\cdot(k+1)\cdot\binom{|X|}{k}$. □

It follows from claim 25 and lemma 26 that the top fan-in s of any multilinear depth three circuit computing $f(X,Y)$ is such that

$$s \geq \frac{n^{2k}}{(k+1)\cdot\binom{4nk}{k}} \geq \frac{n^{2k}}{(k+1)\cdot(4ne)^k} = n^{\Omega(d)},$$

as $n \geq 11$ and $k \leq |X|/2$ (required in lemma 26). Claim 24 now completes the proof of theorem 4.

Theorem 4 implies the following corollary (already known due to [RY09]) as $IMM_{n,d}$ is a simple projection of $\mathsf{Det}_{nd\times nd}$, the determinant of an $nd \times nd$ symbolic matrix [Val79].

Corollary 27 ([RY09])**.** *Any multilinear depth three circuit (over any field) computing Det_d, the determinant of a $d \times d$ symbolic matrix, has top fan-in $2^{\Omega(d)}$.*

Proof of theorem 6: We now show that the polynomial $f(X,Y)$, computed by the ABP $\mathcal{M}$, can also be computed a multilinear depth four circuit of size $O(n^2 d)$ and having top fan-in just one. ABP $\mathcal{M}$ has k matrices, $Y^{(1)}, \ldots, Y^{(k)}$, containing the Y-variables. Associate with each matrix $Y^{(i)}$ four matrices containing the X-variables, two on the immediate left $X^{(4i-3)}$ and $X^{(4i-2)}$, and two on the immediate right $X^{(4i-1)}$ and $X^{(4i)}$. Every monomial in f is formed by picking exactly one variable from every matrix and taking their product. Once we pick $y^{(i)}_{u,v}$ from $Y^{(i)}$, this automatically fixes the variables picked from $X^{(4i-3)}, X^{(4i-2)}, X^{(4i-1)}$ and $X^{(4i)}$, as these are diagonal matrices. Moreover, any variable can be picked from $Y^{(i)}$ irrespective of which other Y-variables are picked from $Y^{(1)}, \ldots, Y^{(i-1)}, Y^{(i+1)}, \ldots, Y^{(k)}$. This observation can be easily formalized to show that

$$f = \prod_{i=1}^{k} \sum_{u,v\in[n]} x^{(4i-3)}_u x^{(4i-2)}_u \cdot y^{(i)}_{u,v} \cdot x^{(4i-1)}_v x^{(4i)}_v.$$

The size of this multilinear $\Pi\Sigma\Pi$ circuit is $O(n^2k) = O(n^2d)$.

Acknowledgements

VN and CS would like to thank Rohit Gurjar for helpful discussions on observations 1 and 3. Thanks also to Nitin Saxena and Arpita Korwar for some early discussions at the onset of this work. NK and CS would like to thank Sébastien Tavenas for attending a presentation of this work and giving us some useful feedback.

References

[AGKS15] Manindra Agrawal, Rohit Gurjar, Arpita Korwar, and Nitin Saxena. Hitting-sets for ROABP and sum of set-multilinear circuits. *SIAM J. Comput.*, 44(3):669–697, 2015.

[Agr05] Manindra Agrawal. Proving lower bounds via pseudo-random generators. In *FSTTCS 2005: Foundations of Software Technology and Theoretical Computer Science, 25th International Conference, Hyderabad, India, December 15-18, 2005, Proceedings*, pages 92–105, 2005.

[ASS13] Manindra Agrawal, Chandan Saha, and Nitin Saxena. Quasi-polynomial hitting-set for set-depth-Δ formulas. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 321–330, 2013.

[CKW11] Xi Chen, Neeraj Kayal, and Avi Wigderson. Partial Derivatives in Arithmetic Complexity and Beyond. *Foundations and Trends in Theoretical Computer Science*, 6(1-2):1–138, 2011.

[DL78] Richard A. DeMillo and Richard J. Lipton. A probabilistic remark on algebraic program testing. *Inf. Process. Lett.*, 7(4):193–195, 1978.

[DMPY12] Zeev Dvir, Guillaume Malod, Sylvain Perifel, and Amir Yehudayoff. Separating multilinear branching programs and formulas. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 615–624, 2012.

[dOSV15] Rafael Mendes de Oliveira, Amir Shpilka, and Ben Lee Volk. Subexponential size hitting sets for bounded depth multilinear formulas. In *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, pages 304–322, 2015.

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

[FLMS14] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *STOC*, pages 128–135, 2014.

[FS13] Michael A. Forbes and Amir Shpilka. Quasipolynomial-time identity testing of non-commutative and read-once oblivious algebraic branching programs. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 243–252, 2013.

[FSS14] Michael A. Forbes, Ramprasad Saptharishi, and Amir Shpilka. Hitting sets for multilinear read-once algebraic branching programs, in any order. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 867–875, 2014.

[GKKS13] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth three. In *Foundations of Computer Science (FOCS)*, pages 578–587, 2013.

[GKST15] Rohit Gurjar, Arpita Korwar, Nitin Saxena, and Thomas Thierauf. Deterministic identity testing for sum of read-once oblivious arithmetic branching programs. In *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, pages 323–346, 2015.

[Hal35] Philip Hall. On representatives of subsets. *J. London Math. Soc.*, 10(1):26–30, 1935.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[HS80] Joos Heintz and Claus-Peter Schnorr. Testing polynomials which are easy to compute (extended abstract). In *Proceedings of the 12th Annual ACM Symposium on Theory of Computing, April 28-30, 1980, Los Angeles, California, USA*, pages 262–272, 1980.

[Juk15] Stasys Jukna. Lower bounds for tropical circuits and dynamic programs. *Theory Comput. Syst.*, 57(1):160–194, 2015.

[Kal89] Erich Kaltofen. Factorization of Polynomials Given by Straight-Line Programs. In *Randomness and Computation*, pages 375–412. JAI Press, 1989.

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

[KS01] Adam Klivans and Daniel A. Spielman. Randomness efficient identity testing of multivariate polynomials. In *Proceedings on 33rd Annual ACM Symposium on Theory of Computing, July 6-8, 2001, Heraklion, Crete, Greece*, pages 216–223, 2001.

[KS14] Neeraj Kayal and Ramprasad Saptharishi. A selection of lower bounds for arithmetic circuits. *Perspectives in Computational Complexity*, 2014.

[KST15] Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. Formulas having low individual degree. *Manuscript*, 2015.

[MV97] Meena Mahajan and V. Vinay. Determinant: Combinatorics, algorithms, and complexity. *Chicago J. Theor. Comput. Sci.*, 1997, 1997.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994.

[NW97] Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997.

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

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

[RY08] Ran Raz and Amir Yehudayoff. Balancing syntactically multilinear arithmetic circuits. *Computational Complexity*, 17(4):515–535, 2008.

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

[Sap14] Ramprasad Saptharishi. Recent progress on arithmetic circuit lower bounds. *Bulletin of the EATCS*, 114, 2014.

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

[Sax13] Nitin Saxena. Progress on polynomial identity testing - II. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:186, 2013.

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

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

[Tut47] W.T. Tutte. The Factorization of Linear Graphs. *J. London Math. Soc.*, 22:107–111, 1947.

[Val79] L. G. Valiant. Completeness Classes in Algebra. In *STOC '79: Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 249–261, New York, NY, USA, 1979. ACM Press.

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Symbolic and Algebraic Computation, EUROSAM '79, An International Symposiumon Symbolic and Algebraic Computation, Marseille, France, June 1979, Proceedings*, pages 216–226, 1979.

A Proofs of observations in section 1

Observation 1 (restated): *Given a circuit C which is a superposition of t set-multilinear circuits on unknown base sets $Y_1, Y_2, ..., Y_t$, finding t base sets $Y_1^{'}, Y_2^{'}, ..., Y_t^{'}$ such that C is a superposition of t set-multilinear circuits on base sets $Y_1^{'}, Y_2^{'}, ..., Y_t^{'}$ is NP-hard when $t > 2$.*

Proof. We will reduce the t-coloring problem to this problem. Given a graph G, the t-coloring problem asks to color the vertices of G with t colors such that no two adjacent vertices of G have the same color. Suppose we are given a graph $G = (V, E)$. From G construct a circuit C as follows. Let $V = \{u_1, ..., u_n\}$, identify these vertices with n-variables. C contains a product gate P multiplying the n variables $(u_1) \cdots (u_n)$. If there exists an edge between two vertices u_1 and u_2 in G then add a product gate P_{u_1,u_2} in C having a single linear polynomial $(u_1 + u_2)$. Claim: circuit C is a superposition of t set-multilinear depth three circuits if and only if graph G is t-colorable. Suppose G is t-colorable. The t sets of vertices of G with t different colors correspond to t valid base sets in C and thus C is a superposition of t set-multilinear depth three circuits. In the reverse direction, say C is a superposition of t set-multilinear depth three circuits. This implies C has t base sets. These t base sets can be readily used to give a t-coloring of G: Every base set gets a unique color. Say two variables u_i and u_j belong to the same base set, then u_i and u_j do not appear in the same linear polynomial. But this implies there is no edge between u_i and u_j in G else there would have been a product gate P_{u_i,u_j} having a single linear polynomial $(u_i + u_j)$ in C. □

Observation 3 (restated): *A polynomial computed by a multilinear $\Sigma\Pi\Sigma$ circuit with top fan-in two and at most two variables per linear polynomial can also be computed by an ROABP with constant width.*

Proof. Let C be a multilinear depth three circuit with top fan-in two and at most two variables per linear polynomial computing the polynomial $f(X)$ in n variables $\{x_1, \ldots x_n\}$. Let $\sigma : [n] \to [n]$ be a permutation function. Then without loss of generality $f(X)$ can be expressed as

$$f(X) = \prod_{i\in[n],i \text{ odd}} (1 + x_i + x_{i+1}) \; + \prod_{i\in[n],i \text{ odd}} (1 + x_{\sigma(i)} + x_{\sigma(i+1)}).$$

We have assumed that the coefficients of x_i's and the constant term in every linear polynomial is 1, and n is even. These are without any loss of generality and the argument holds even otherwise. Let $P_1 = \prod_{i\in[n],i \text{ odd}}(1 + x_i + x_{i+1})$ and $P_2 = \prod_{i\in[n],i \text{ odd}}(1 + x_{\sigma(i)} + x_{\sigma(i+1)})$. Product gates P_1 and P_2 can be easily computed individually by ROABPs of width two but with different variable orderings. We express the two ROABPs in the same variable ordering and add the polynomials computed by them (P_1 and P_2) to get an ROABP computing f.

We partition the linear polynomials in P_1 and P_2 into sets $\{L_{11}, L_{12}, \ldots, L_{1k}\}$ and $\{L_{21}, L_{22}, \ldots, L_{2k}\}$ respectively such that the sets of variables appearing in the linear polynomials in L_{1t} and L_{2t}, where $t \in [k]$, are equal and this set is completely disjoint from the set of variables appearing in the linear polynomials in L_{mr}, where $m \in [2]$ and $r \in [k] \setminus \{t\}$. We give a 'greedy' partition procedure below. Mark all the linear polynomials in P_1 and P_2 as unpicked. Initialize $t = 1$ and $i = 1$:

1. Pick an unpicked linear polynomial $l_p = (1 + x_i + x_{i+1})$ in P_1 and put it in L_{1t}. Mark l_p as picked. Store the value i in temp: temp=i.
2. Let the linear polynomial in which the variable x_{i+1} appears in P_2 be $l_q = (1 + x_{i+1} + x_j)$. Put l_q in L_{2t} and mark l_q as picked.
3. If j is equal to temp then increment t and start from step 1.
4. Else set $i = j$ and let the linear polynomial in which the variable x_i appears in P_1 be $l_r = (1 + x_i + x_{i+1})$. Put l_r in L_{1t} and mark l_r as picked.

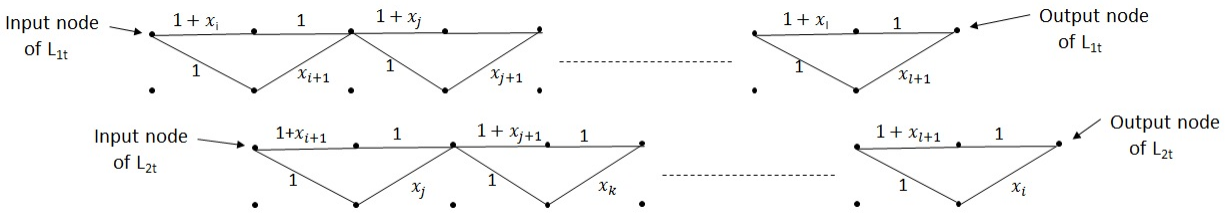

Figure 2: ROABPs corresponding to L_{1t} and L_{2t}

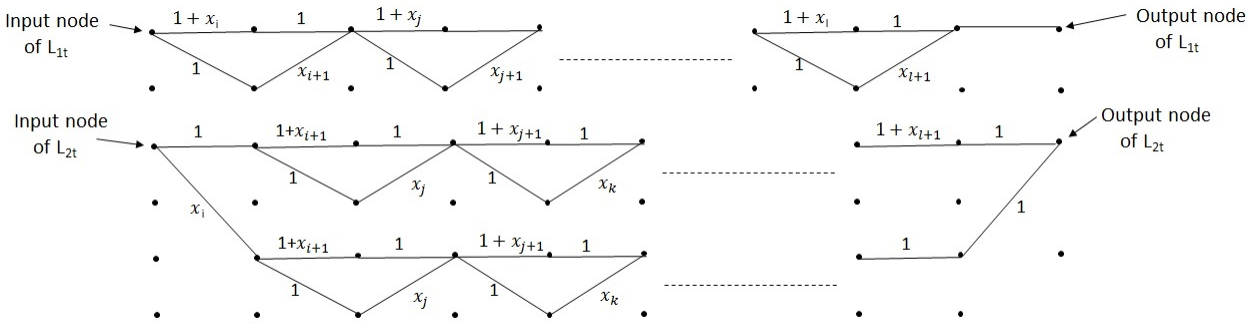

Figure 3: ROABPs (with same variable ordering) corresponding to L_{1t} and L_{2t}

5. Repeat from step 2.

Clearly, the sets of variables appearing in the linear polynomials in L_{1t} and L_{2t}, where $t \in [k]$, are equal and this set is disjoint from the set of variables appearing in the linear polynomials in L_{mr}, for $m \in [2]$ and $r \in [k] \setminus \{t\}$. We express the two ROABPs computing P_1 and P_2 in the same variable ordering as a sequence of k parts. In part t, we compute the product of linear polynomials in L_{1t} and L_{2t} separately using two ROABPs such that the variable orderings in these two ROABPs are the same. Finally, we connect the ROABPs from these k parts to give a single ROABP of width six.

We arrange the linear polynomials in L_{1t} and L_{2t} in the order they are picked during the partition process. Suppose after this arrangement we have $L_{1t} = \{(1 + x_i + x_{i+1}), (1 + x_j + x_{j+1}), ..., (1 + x_l + x_{l+1})\}$ and $L_{2t} = \{(1 + x_{i+1} + x_j), (1 + x_{j+1} + x_k), ..., (1 + x_{l+1} + x_i)\}$. Figure 2 shows the two ROABPs computing the product of linear polynomials in L_{1t} and L_{2t} respectively. Consider the input and output nodes of L_{1t} and L_{2t} marked in figure 2 as the sources and sinks of these two ROABPs respectively. The variables are arranged such that except x_i all variables are in the same order in the two ROABPs. We order x_i by breaking the second ROABP in two parts as shown in figure 3. The first part computes the polynomial in which x_i does not appear and the second part brings x_i to the beginning and computes the polynomial in which x_i appears. Finally we add these two parts by adding an extra layer. We get similar directed acyclic graphs each having two ROABPs with consistent variable ordering from all the k pairs of sets of linear polynomials. We connect these k graphs by adding weight 1 edges between the input nodes of L_{1r}, L_{2r} and the output nodes of $L_{1(r+1)}, L_{2(r+1)}$ respectively, where $r \in [k-1]$. The resulting graph is an ROABP

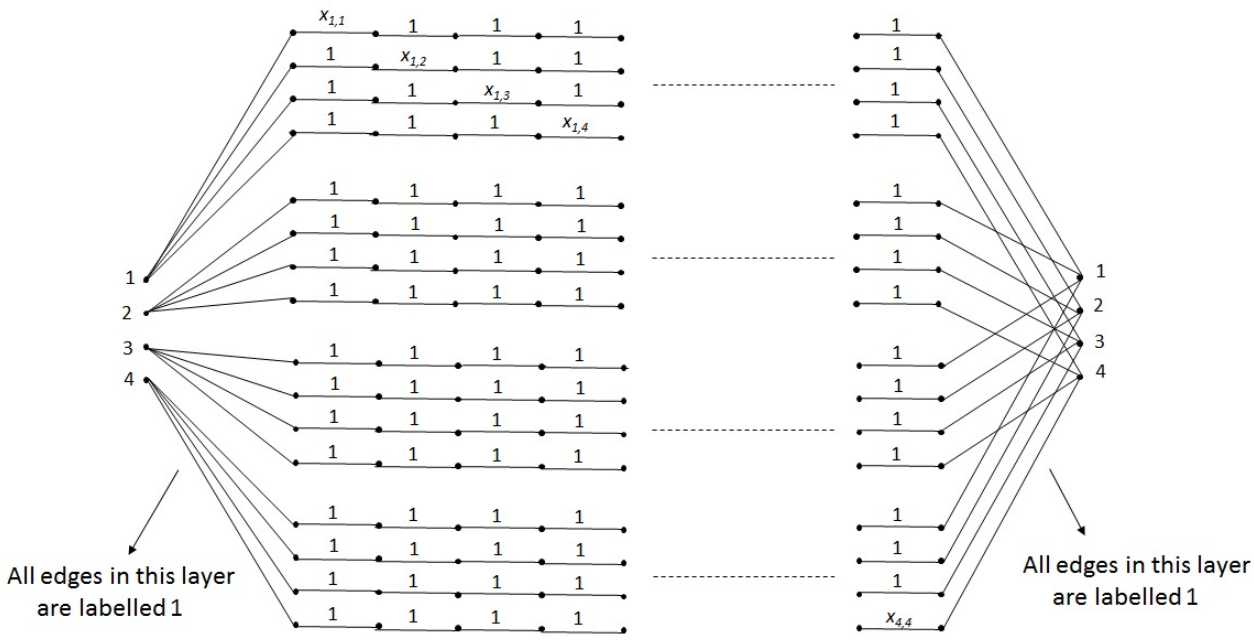

Figure 4: ROABP corresponding to the split of a matrix X

of width six computing f. □

Observation 5 (restated): $\text{IMM}_{n,d}$ *can be computed by an* n^2 *width ROABP.*

Proof. We transform the width n ABP computing $\text{IMM}_{n,d}$ to a width n^2 ROABP computing the same. Let $\{X^{(1)}, X^{(2)}, ..., X^{(d)}\}$ be the d matrices in $\text{IMM}_{n,d}$. The (j,k)-th entry in $X^{(i)}$ is $x^{(i)}_{j,k}$. We replace matrix $X^{(i)}$ by n^2+2 matrices: $A^{(i,1)}$, $A^{(i,2)}$ and $A^{(i,j,k)}$ where $j,k \in [n]$. $A^{(i,1)}$ and $A^{(i,2)}$ are rectangular matrices of dimension $n \times n^2$ and $n^2 \times n$ respectively. For $j,k \in [n]$, $A^{(i,j,k)}$ are diagonal matrices of dimension n^2. Ordered from left to right $A^{(i,1)}$ and $A^{(i,2)}$ are first and last respectively and $A^{(i,j_1,k_1)}$ comes before $A^{(i,j_2,k_2)}$ if $j_1 < j_2$ or if $j_1 = j_2$ and $k_1 < k_2$. The (a,a)-th entry of $A^{(i,j,k)}$ is $x^{(i)}_{j,k}$ if $a = n \cdot (j-1) + k$ and 1 otherwise. The (a,b)-th entry of $A^{(i,1)}$ is 1 if $(a-1) \cdot n + 1 \leq b \leq a \cdot n$ and 0 otherwise. Similarly, the (a,b)-th entry of $A^{(i,2)}$ is 1 if $b \equiv a \mod n$ and 0 otherwise. Figure 4 shows the part of ROABP corresponding to the split of a matrix X into n^2+2 matrices, when $n=4$, as explained above. When we split $X^{(i)}$ into n^2+2 matrices as above the corresponding part of ROABP computing the product of these n^2+2 matrices has n vertices in both the leftmost and rightmost layers of vertices. There is a unique path from the j-th vertex in leftmost layer to the k-th vertex in rightmost layer with weight $x^{(i)}_{j,k}$. Hence the product of the n^2+2 matrices arranged as above is $X^{(i)}$.

To transform the ABP computing $\text{IMM}_{n,d}$ to an ROABP we have introduced between every pair of adjacent layers of vertices in the ABP, n^2 layers with n^2 vertices in each layer, hence the width of the ROABP is n^2. □

B Proofs of lemmas in section 2

Lemma 8 (restated).

1. *Let* $g_1, g_2 \in \mathbb{F}[X]$ *and* $S \subseteq X$. *Then*

$$\text{Evaldim}_S(g_1 + g_2) \leq \text{Evaldim}_S(g_1) + \text{Evaldim}_S(g_2).$$

2. *Let* $f_1, f_2 \in \mathbb{F}[X, Y]$. *Then* $\mathsf{PD}_{\mathcal{Y}_k}(f_1 + f_2) \leq \mathsf{PD}_{\mathcal{Y}_k}(f_1) + \mathsf{PD}_{\mathcal{Y}_k}(f_2)$.

Proof. For $i \in \{1, 2\}$, let

$$\begin{aligned} V_i &= \text{span}_{\mathbb{F}}\{g_i(X)|_{\forall x_j \in S \ x_j = \alpha_j} : \forall x_j \in S \ \alpha_j \in \mathbb{F}\} \text{ and} \\ W &= \text{span}_{\mathbb{F}}\{(g_1 + g_2)(X)|_{\forall x_j \in S \ x_j = \alpha_j} : \forall x_j \in S \ \alpha_j \in \mathbb{F}\}. \end{aligned}$$

Every polynomial in W belongs to $V_1 + V_2$, where $V_1 + V_2 = \{f_1 + f_2 | f_1 \ \in V_1, f_2 \in V_2\}$. Hence, $\text{Evaldim}_S(g_1 + g_2) = \dim(W) \leq \dim(V_1 + V_2) \leq \dim(V_1) + \dim(V_2) = \text{Evaldim}_S(g_1) + \text{Evaldim}_S(g_2)$.

Proving part two is similar to part one. For $i \in \{1, 2\}$ let

$$\begin{aligned} A_i &= \text{span}_{\mathbb{F}}\left\{ \left[\frac{\partial f_i(X, Y)}{\partial m} \right]_{\forall y \in Y \ y=0} : m \in \mathcal{Y}_k \right\} \text{ and} \\ B &= \text{span}_{\mathbb{F}}\left\{ \left[\frac{\partial (f_1 + f_2)(X, Y)}{\partial m} \right]_{\forall y \in Y \ y=0} : m \in \mathcal{Y}_k \right\} \end{aligned}$$

Again observing B is a subspace of $A_1 + A_2$, where $A_1 + A_2 = \{g_1 + g_2 | g_1 \in A_1, g_2 \in A_2\}$, part two follows. □

Definition 12. *Let* $G = (V, E)$ *be a* d*-regular graph with* $|V| = n$. *Let* A_G *be the adjacency matrix of* G *and* $d = \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ *the* n *eigenvalues of* A_G. *Then* $\lambda(G) \stackrel{def}{=} \max\{|\lambda_2|, |\lambda_n|\}$. *The ordered set of eigenvalues* $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ *is called the spectrum of* G.

Theorem 2.4 in [HLW06]. (Cheeger's inequality) *Let* G *be a* d*-regular graph with spectrum* $(\lambda_1, \lambda_2, \ldots, \lambda_n)$. *Then*

$$\frac{d - \lambda_2}{2} \leq h(G) \leq \sqrt{2d(d - \lambda_2)}.$$

Lemma 9 (restated). *Let* $\mathcal{S} = \{G_p\}_{p \text{ prime}}$ *be the family of expander graphs as described in section 2, and* $\tilde{\mathcal{S}} = \{\tilde{G}_p\}_p$ *the family of double covers of graphs in* $\mathcal{S}$. *Then* $h(\tilde{G}_p) > \frac{2 + 10^{-4}}{2}$ *for every* p.

Proof. The family $\mathcal{S} = \{G_p\}_{p \text{ prime}}$ is such that $\lambda(G_p) < 1 - 10^{-4}$ for every p (argued in section 11.1.2 of [HLW06]). Observe that if $(\lambda_1, \ldots, \lambda_p)$ is the spectrum of G_p then $\{\pm\lambda_1, \ldots, \pm\lambda_p\}$ are exactly the eigenvalues of the adjacency matrix of the bipartite graph $\tilde{G}_p$. Hence, $\lambda(G_p)$ is the second largest eigenvalue of $A_{\tilde{G}_p}$. By applying Cheeger's inequality, $h(\tilde{G}_p) > \frac{2 + 10^{-4}}{2}$ for every p as $\tilde{G}_p$ is 3-regular. □

Lemma 11 (restated). *Suppose $g_1(X), g_2(X), ..., g_m(X) \in \mathbb{F}[X]$ are $\mathbb{F}$-linearly independent polynomials in the variables $X = \{x_1, x_2, ..., x_n\}$ where $m = 2^n$. If $Y = \{y_1, y_2, ..., y_n\}$ are n variables different from X then (by identifying an $i \in [m]$ with an $S \subseteq [n]$),*

$$\text{Evaldim}_Y(\sum_{S\subseteq[n]} y_S \cdot g_S(X)) = m, \quad \text{where for } S \subseteq [n],\ y_S := \prod_{j\in S} y_j.$$

Proof. Consider the following $\mathbb{F}$-evaluations of $\{y_1, y_2, ..., y_n\}$: for every $S \subseteq [n]$, if $j \in S$ set $y_j = 1$ else set $y_j = 0$. There are $m = 2^n$ such evaluations. By taking appropriate $\mathbb{F}$-linear combinations of these evaluations of the polynomial $\sum_{S\subseteq[n]} y_S \cdot g_S$, one can get the m polynomials $\{g_S\}_{S\subseteq[n]}$. Since these m polynomials are given to be $\mathbb{F}$-linearly independent, $\text{Evaldim}_Y(\sum_{S\subseteq[n]} y_S g_S(X)) \geq m$. On the other hand, any $\mathbb{F}$-evaluation of the Y-variables of the polynomial $\sum_{S\subseteq[n]} y_S \cdot g_S(X)$ is a $\mathbb{F}$-linear combination of the m polynomials $\{g_S\}_{S\subseteq[n]}$ and hence $\text{Evaldim}_Y(\sum_{S\subseteq[n]} y_S \cdot g_S(X)) \leq m$. □

Lemma 12 (restated). *If R is a width-k ROABP that computes $g(X)$ then for every $i \in [0, |X|]$ there exists a set $S \subseteq X$ of size i such that $\text{Evaldim}_S(g(X)) \leq k$.*

Proof. Without loss of generality, assume the permutation π associated with the ROABP R is the identity permutation. Hence R can be equivalently viewed as a product of n matrices $M_1, \ldots, M_n$ computing $g(X) = M_1 \cdot M_2 \cdots M_n$, where M_1 is a $1 \times k$ matrix with entries from $\mathbb{F}[x_1]$, M_n is a $k \times 1$ matrix with entries from $\mathbb{F}[x_n]$, and M_j is a $k \times k$ matrix with entries from $\mathbb{F}[x_j]$ for every $j \in [2, n-1]$. Let $S = \{x_1, x_2, ..., x_i\}$. Consider any $\mathbb{F}$-evaluation of the S variables in $g(X)$. Denote the resulting polynomial by $g_1(X \setminus S) \in \mathbb{F}[x_{i+1}, ..., x_n]$. Observe that $g_1(X \setminus S) = M_{eval} \cdot M_{i+1} \cdots M_n$ where $M_{eval} \in \mathbb{F}^{1\times k}$. Let $M = M_{i+1} \cdots M_n$ be the $k \times 1$ column vector with entries from $\mathbb{F}[x_{i+1}, \ldots, x_n]$. Thus, $g_1(X \setminus S) = M_{eval} \cdot M$ is an $\mathbb{F}$-linear combination of k polynomials in $\mathbb{F}[x_{i+1}, \ldots, x_n]$ that do not depend on which evaluation of the $\{x_1, \ldots, x_i\}$-variables we began with. Hence, evaluation dimension of $g(X)$ with respect to S is upper bounded by k. □

C Proofs of observations and claims in section 3

Claim 13 (restated). *Polynomial g (as constructed in section 3, proof of part 1) is computed by a multilinear depth three circuit C of size $\Theta(n)$ and top fan-in three, and C is a superposition of two set-multilinear depth three circuits.*

Proof. Since g is a sum of three product terms, where each product term is a product of linear polynomials on disjoint sets of variables, it can be computed by a multilinear depth three circuit C with top fan-in three. The bottom fan-in (fan-in of the sum gates at layer 3) is three since there are two variables and the field constant 1 per linear polynomial. The fan-in of every product gate is n. As there are three product gates, the total number of edges in C is $3+3(n(1+3)) = 3+12n = \Theta(n)$. Every linear polynomial of a product gate has two variables, an X and a Y variable. Hence the circuit is a superposition of two set-multilinear depth three circuits on base sets X and Y. □

Claim 15 (restated). *There exists a set $X_0 \subseteq X$ of $\left(\frac{7n}{10} - 4\right)$ X-variables such that every $x \in X_0$ appears in an untouched linear polynomial in every P_i (for $i \in [3]$), and further if $(1+x+y_{j_1}), (1+x+y_{j_2})$ and $(1+x+y_{j_3})$ are the linear polynomials occurring in P_1, P_2 and P_3 respectively then $y_{j_1} \neq y_{j_2} \neq y_{j_3}$.*

Proof. Recall $|S| = \frac{n}{10}$. A linear polynomial in a product gate is called touched if it is either a partially touched or a completely touched linear polynomial. For every $i \in [3]$, let D_i represent the set of touched linear polynomials in product gate i. Hence $|D_1| + |D_2| + |D_3| \leq \frac{3n}{10}$. Thus the number of X-variables that are part of these touched linear polynomials is at most $\frac{3n}{10}$ as every linear polynomial has exactly one X-variable. This implies at least $\frac{7n}{10}$ X-variables are part of untouched linear polynomials in every product gate. As $g(X, Y)$ is constructed from $\tilde{G}$, two product gates contain the same linear polynomial l if and only if there is a double edge between the endpoints of the edge corresponding to the linear polynomial l in $\tilde{G}$. Graph $\tilde{G}$ is the double cover of the n-vertex graph $G \in \mathcal{S}$ where $n > 2$ is a prime. A double edge between vertices u_L and v_R in $\tilde{G}$ implies existence of a double edge between vertices u and v in G. Vertices of G are identified with elements of $\mathbb{Z}_n$. A vertex a in G_n is connected to $a + 1, a - 1$ and a^{-1} (operations are modulo n and inverse of 0 is 0). Thus, there is a double edge incident on a vertex a in G if and only if any two of the vertices $a + 1, a - 1$ and a^{-1} are the same. If $a + 1 = a - 1 \mod n$, then $2 = 0 \mod n$ which cannot be true as $n > 2$. Hence if there is a double edge incident on a then either $a + 1 = a^{-1} \mod n$, or $a - 1 = a^{-1} \mod n$. This means G has exactly two sets of double edges – between $\frac{(-1-\sqrt{5})}{2}$ and $\frac{1-\sqrt{5}}{2}$, and between $\frac{(-1+\sqrt{5})}{2}$ and $\frac{1+\sqrt{5}}{2}$ – if 5 is a square in $\mathbb{Z}_n$; otherwise G has no double edge. As a double edge in G gives rise to two double edges in $\tilde{G}$, the latter has at most four double edges. Thus at most four out of the $\frac{7n}{10}$ X-variables are part of untouched linear polynomials that appear in more than one product gate. We remove these four variables. X_0 is the set of the remaining X-variables of size at least $\left(\frac{7n}{10} - 4\right)$. Naturally, every variable in X_0 has the desired property as stated in the claim. □

Observation 18 (restated). $\text{Evaldim}_S(g) \geq \text{Evaldim}_S(\hat{g}) = \text{Evaldim}_S(\hat{P}_1) \geq 2^{\varepsilon n}$.

Proof. It is easy to see $\text{Evaldim}_S(g) \geq \text{Evaldim}_S(\hat{g}) = \text{Evaldim}_S(\hat{P}_1)$ as follows. Let

$$V = \text{span}_{\mathbb{F}}\{g(X, Y)|_{\forall u_j \in S \ u_j = \alpha_j} : \forall u_j \in S \ \alpha_j \in \mathbb{F}\},$$

$$\hat{V} = \text{span}_{\mathbb{F}}\{\hat{P}_1(X, Y)|_{\forall u_j \in S \ u_j = \alpha_j} : \forall u_j \in S \ \alpha_j \in \mathbb{F}\}$$

and $t = \text{Evaldim}_S(g)$. Let $\{h_1, \ldots, h_t\}$ be a basis of V. Since the linear polynomials $(1+x+y_{j_2})$ and $(1+x'+y'_{j_3})$ are untouched, the variables x, x', y_{j_2}, y_{j_3} do not belong to S and hence the polynomials $\{\hat{h}_1, \ldots, \hat{h}_t\}$ span the space $\hat{V}$, where $\hat{h}_i$ is polynomial h_i under the substitution $x = -(1 + y_{j_2})$ and $x' = -(1 + y_{j_3})$. Below we show $\text{Evaldim}_S(\hat{P}_1) \geq 2^{\varepsilon n}$.

Suppose $\hat{P}_1$ has $T \geq \varepsilon n$ partially touched linear polynomials $\{l_1, l_2, ..., l_T\}$. For every $r \in [T]$, let $l_r = 1 + z_r + u_r$ where $z_r \in S$ and $u_r \in (X \cup Y) \backslash S$. Let $Z = \{z_1, z_2, ..., z_T\}$. Make the following two kinds of substitutions in $\hat{P}_1$: first, substitute all variables in $S \backslash Z$ by 1; second, substitute $u_r = u_r - 1$ for every $r \in [T]$. Let $\tilde{P}_1$ correspond to $\hat{P}_1$ after these substitutions. It follows easily that $\text{Evaldim}_Z(\tilde{P}_1) \leq \text{Evaldim}_S(\hat{P}_1)$ as $Z \subseteq S$.

Let f be the polynomial formed by multiplying all linear polynomials in $\tilde{P}_1$ that are free of variables in Z. Then,

$$\tilde{P}_1 = \left(\sum_{\nu \subseteq [T]} z_\nu u_{[T] \backslash \nu} \right) \cdot f,$$

where $z_\nu = \prod_{j\in\nu} z_j$ and $u_{[T]\setminus\nu} = \prod_{j\in[T]\setminus\nu} u_j$. Since f is Z-free,

$$\text{Evaldim}_Z(\tilde{P}_1) = \text{Evaldim}_Z\left(\sum_{\nu\subseteq[T]} z_\nu u_{[T]\setminus\nu}\right) = 2^T \quad \text{(by lemma 11)}.$$

□

Observation 22 (restated). *$|C_1| + |C_2|$ is at least $\frac{n}{15} - \frac{8T}{3}$.*

Proof. Recall for all $i \in [2]$, $|B_i| \leq T$. The number of variables in S that are part of partially touched linear polynomials in either of the product gates is at most $4T$; $2T$ from each product gate. Hence at least $\frac{n}{10} - 4T$ variables in S are part of completely touched linear polynomials in both the product gates. Since the number of variables per linear polynomial is 3, the number of completely touched linear polynomials in each of the product gates is at least $\left(\frac{n}{30} - \frac{4T}{3}\right)$. Hence $|C_1| + |C_2| \geq \left(\frac{n}{15} - \frac{8T}{3}\right)$. □

Claim 23 (restated). *If there exists an $i \in [2]$ such that $|B_i| \geq \varepsilon n$ for $\varepsilon > 0$, then* $\text{Evaldim}_S(g) \geq 2^{\varepsilon n}$.

Proof. Without loss of generality assume $|B_1| \geq \varepsilon n$. Since no two vertices in $\tilde{G}$ have all the three edges in common, the linear polynomial l is unique to a product gate, i.e if l is a linear factor of P_2 then l is not a linear factor of P_1. Pick an untouched linear polynomial: $(1 + x + y + z)$ in P_2 such that x also appears in an untouched linear polynomial in P_1 – we know there are at least $n - \frac{2n}{10} = \frac{4n}{5}$ such X-variables. By substituting $x = -(1 + y + z)$, P_2 vanishes but P_1 remains non zero. Let $\hat{g}$ be the polynomial we get after this substitution. $\hat{g}$ has just one product term $\hat{P}_1$ (corresponding to P_1 after substitution). $\hat{P}_1$ has as many partially touched linear polynomials as P_1. From here on a similar argument used to prove observation 18 above can be used to show $\text{Evaldim}_S(g) \geq \text{Evaldim}_S(\hat{g}) = \text{Evaldim}_S(\hat{P}_1) \geq 2^{\varepsilon n}$. □

D Proof of claim in section 4

Claim 24 (restated). *If $\text{IMM}_{n,d}$ is computed by a multilinear depth three circuit having top fan-in s then f is also computed by a multilinear depth three circuit having top fan-in s.*

Proof. f is computed by the ABP $\mathcal{M}$ of width n and length d as described in section 4. Each edge in $\mathcal{M}$ is labelled by a distinct variable or 1. Let $\text{IMM}_{n,d}$ be the $(1,1)$-th entry of a product of d $n \times n$ symbolic matrices $\{Z^{(1)}, Z^{(2)}, ..., Z^{(d)}\}$ ordered from left to right. The (j,k)-th entry in $Z^{(i)}$ is the formal variable $z^{(i)}_{j,k}$. We project $\text{IMM}_{n,d}$ to f as follows. Recall $\mathcal{M}$ has three kinds of matrices: X,Y and A. The matrices $\{Z^{(1)}, Z^{(2)}, ..., Z^{(d)}\}$ would correspond to the X, Y and A matrices in the same order as they appear in the ABP $\mathcal{M}$. $Z^{(1)}$ corresponds to the row vector $X^{(1)}$, so $z^{(1)}_{1,l}$ maps to $x^{(1)}_l$ and $z^{(1)}_{m,l}$ to 0 for $m \in [2,n]$. Similarly, $Z^{(d)}$ corresponds to the column vector $X^{(r)}$, so $z^{(d)}_{m,1}$ maps to $x^{(r)}_m$ and $z^{(d)}_{m,l}$ to 0 for $l \in [2,n]$. If $Z^{(i)}$ corresponds to $X^{(j)}$ for $j \in [2, r-2]$ then we map $z^{(i)}_{m,m}$ to $x^{(j)}_m$ and $z^{(i)}_{m,l}$ to 0 if $m \neq l$. If $Z^{(i)}$ corresponds to $Y^{(j)}$ then we map $z^{(i)}_{m,l}$ to $y^{(j)}_{m,l}$. If $Z^{(i)}$ corresponds to $A^{(j)}$ then we map all the variables in $Z^{(i)}$ to 1. Such a projection of

$\text{IMM}_{n,d}$ equates to f. Suppose $\text{IMM}_{n,d}$ is computed by a multilinear depth three circuit C. Then by applying the map on the variables of $\text{IMM}_{n,d}$ and C, we get that the image of C computes f. Two distinct variables are not mapped to the same variable under this map. Hence image of C is still a multilinear depth three circuit having top fan-in same as that of C. □

E Proof of theorem 7

We prove theorem 7 in this section. In particular we use the shift and rank concentration technique used in [ASS13] to give a quasi-polynomial time hitting set for a restricted class of multilinear depth three circuits. The model we consider is a multilinear depth three circuit that is both a superposition of m set-multilinear depth three circuits and simultaneously a sum of l set-multilinear depth three circuits, where m and l are constants. Before we prove 7 we briefly review the shift and rank concentration technique from [ASS13].

Shift and rank concentration. Suppose we wish to check whether a polynomial computed by a set-multilinear depth three circuit is identically zero. Let the given circuit be $C(X) = \sum_{i=1}^{k}\prod_{j=1}^{d} l_{ij}(X_j)$, where $X = \uplus_j^d X_j$, $X_j = \{x_{j1}, x_{j2}, ..., x_{jn}\}$ and l_{ij}'s are linear polynomials in variables X_j. We view the polynomial C as a k component vector where the i-th component is the polynomial computed by the i-th product gate. A dot product with the all ones vector $\overline{1}$, would give us the polynomial C. In shift and rank concentration, we shift each variable x_{jr} to $x_{jr} = x_{jr} + t_{jr}$, where t_{jr}'s are formal variables. Let $T_j = \{t_{j1}, t_{j1}, ..., t_{jn}\}$, $T = \uplus_j^d T_j$, $S \subseteq X$, $\nu_S = \prod_{x_{jr}\in S} x_{jr}$ and Z_{ν_S} be the coefficient vector over $\mathbb{F}(T)$ corresponding to the monomial ν_S. We use a map $\tau : t_{jr} \to t^{\omega_{jr}}$ such that

$$\text{span}_{\mathbb{F}(t)}\{Z_{\nu_S} : |S| \leq \lceil \log k \rceil\} \;=\; \text{span}_{\mathbb{F}(t)}\{Z_{\nu_S}\},$$

where $\text{span}_{\mathbb{F}(t)}\{Z_{\nu_S}\}$ denotes the span of the coefficient vectors over $\mathbb{F}(t)$ corresponding to the different monomials in the shifted polynomial and $|S|$ equals the support[13] of the monomial ν_S. [ASS13] showed that it is sufficient to try $n^{O(\log k)}$ many maps to find the desired one such that the ω_{jr}'s are bounded by a polynomial in nd, the number of variables. After such a shift using the desired map, the polynomial C is non zero if and only if there a exists a monomial in the shifted polynomial with support less than or equal to $\lceil \log k \rceil$ and a non-zero coefficient in $\mathbb{F}(t)$. Thus we check whether the shifted polynomial has a non-zero monomial with support less than or equal to $\lceil \log k \rceil$, by projecting over all possible choices of $\lceil \log k \rceil$ variables and test if the shifted polynomial is non zero using [KS01] in $n^{O(\log k)}$ time. Now we prove theorem 7.

Theorem 7 (restated) *Let $\mathcal{C}_{n,m,l,s}$ be a subclass of multilinear depth three circuits computing n-variate polynomials such that every circuit in $\mathcal{C}_{n,m,l,s}$ is a superposition of at most m set-multilinear depth three circuits and simultaneously a sum of at most l set-multilinear depth three circuits, and has top fan-in s. There is a hitting-set generator for $\mathcal{C}_{n,m,l,s}$ running in $(ns)^{O(lm\log s)}$ time.*

Proof. Let $X = \uplus_i^m X_i$, where $X_1, \ldots, X_m$ are m base sets of variables, and C be a superposition of m set-multilinear depth three circuits in these base sets. Assume for all $i \in [m]$ $|X_i| = a$. This is without loss of generality and our argument holds even when the size of base sets are not equal.

[13] Support of a monomial is the number of variables in the monomial with degree at least 1.

So, let $X_i = \{x_{i1}, x_{i2}, ..., x_{ia}\}$. By assumption, C is a sum of l set-multilinear depth three circuits, say $\{C_1, C_2, ..., C_l\}$. Let s_k be the top fan-in of C_k. Naturally $\sum_{k=1}^{l} s_k \leq s$. We can therefore represent C_k as follows.

$$C_k(X) = \sum_{i=1}^{s_k} \prod_{j=1}^{a} (\alpha_{ij} + z_{i_{1j}} x_{1j} + z_{i_{2\sigma_{k_2}(j)}} x_{2\sigma_{k_2}(j)} + ... + z_{i_{m\sigma_{k_m}(j)}} x_{m\sigma_{k_m}(j)}),$$

where for $q \in \{2, ..., m\}$, σ_{k_q} represents the permutation function $[a] \to [a]$ corresponding to the q-th base set, and the z's are constants from $\mathbb{F}$. Without loss of generality we can assume σ_{k_1} is the identity permutation corresponding to the first base set X_1, and $\alpha_{ij} \neq 0$. Circuit $C(X) = \sum_{k=1}^{l} C_k$. Also let $U_r = \uplus_{j=1}^{r} X_j$ and $W_r = X \setminus U_r$.

Proof outline. Let $T_i = \{t_{i1}, t_{i2}, ..., t_{ia}\}$, we have m such sets $T_1, T_2, ..., T_m$. Let $T = \uplus_i^m T_i$ and $Y_r = \uplus_{j=1}^{r} T_j$ and $Z_r = T \setminus Y_r$. We shift a variable x_{ij} to $(x_{ij} + t_{ij})$. For now consider each of the t_{ij} variables as formal variables, finally we will substitute $t_{ij} = t^{\omega_{ij}}$, where t is a fresh variable and w_{ij} is an appropriate small constant. We analyse the shift in m steps. In the r-th step we analyse the shift of the X_r variables and show that there exists a monomial in U_r variables of support less than $r \log s$ that has a non-zero coefficient polynomial in $\mathbb{F}[Y_r \cup W_r]$ if $C \neq 0$. Further, to keep this coefficient polynomial non-zero under a substitution $t_{ij} = t^{\omega_{ij}}$ we only need to preserve the 'non-zeroness' of a small collection of polynomials in a few variables in Y_r. Finally, by using [KS01] to construct a good substitution $t_{ij} = t^{\omega_{ij}}$, we can keep all polynomials (in the T-variables) collected over m-steps non-zero, thereby showing that the original polynomial C is non-zero if and only if there exists a monomial in the shifted polynomial of support less than $m \log s$ that has a non-zero coefficient over the field $\mathbb{F}(t)$. Once we show this, finding a hitting set is easy: project over all possible choices of $(m \log s)$ variables and test if the shifted polynomial is non-zero over $\mathbb{F}(t)$ using sparse PIT [KS01]. We stress that the algorithm shifts all variables simultaneously, only the analysis proceeds in steps. We explain step 1 and then generalize the argument to the r-th step.

Step 1: We view $C(X)$ as a polynomial in X_1 variables with coefficients in $\mathbb{F}(W_1)$. In step 1 we shift x_{1j} to $x_{1j} + t_{1j}$. Since C is a set-multilinear depth three circuit in X_1 variables, from [ASS13] we know that, C is non-zero if and only if there exists a monomial in X_1 variables (of support at most $\log s$) that has a non-zero coefficient polynomial in $\mathbb{F}[Y_1 \cup W_1]$. For this step to work under a substitution $t_{ij} = t^{\omega_{ij}}$, we need to keep $a^{O(\log s)}$ multilinear polynomials in T_1-variables non-zero where each of the multilinear polynomials involves $O(\log s)$ T_1-variables. Let $\nu = \prod_{j=1}^{\log s} x_{1j}$ be such a monomial with a non-zero coefficient polynomial. The coefficient polynomial $C^{(2)}(W_1)$ of ν is a superposition of $(m-1)$ set-multilinear depth three circuits on base sets $X_2, ..., X_m$ and sum of l set-multilinear depth three circuits $\{C_1^{(2)}, C_2^{(2)}, ..., C_l^{(2)}\}$ over $\mathbb{F}(Y_1)$. The representation of $C_k^{(2)}$ is as follows:

$$C_k^{(2)}(W_1) = \sum_{i=1}^{s_k} \alpha_i \cdot \prod_{j=\log s+1}^{a} (1 + z_{i_{1j}} t_{1j} + z_{i_{2\sigma_{k_2}(j)}} x_{2\sigma_{k_2}(j)} + ... + z_{i_{m\sigma_{k_m}(j)}} x_{m\sigma_{k_m}(j)}),$$

where $\alpha_i \in F$ and (reusing symbols) the z's are also in $\mathbb{F}$. We can rewrite $C_k^{(2)}$ as follows by defining

$z_{i_{1j}} = z_{i_{q\sigma_{k_q}(j)}} = 0$ for $j \in [\log s]$ and $q \in [2, m]$,

$$C_k^{(2)}(W_1) = \sum_{i=1}^{s_k} \alpha_i \cdot \prod_{j=1}^{a} (1 + z_{i_{1j}} t_{1j} + z_{i_{2\sigma_{k_2}(j)}} x_{2\sigma_{k_2}(j)} + ... + z_{i_{m\sigma_{k_m}(j)}} x_{m\sigma_{k_m}(j)}).$$

Suppose in the $(r-1)$-th step we have a monomial ν in U_{r-1} variables with support less than $(r-1)\log s$ that has a non-zero coefficient polynomial in $\mathbb{F}[Y_{r-1} \cup W_{r-1}]$. This coefficient polynomial $C^{(r)}(W_{r-1}) = C^{(r)}$ is a superposition of $m - r + 1$ base sets $X_r, ..., X_m$ and sum of l set-multilinear depth three circuits $\{C_1^{(r)}, C_2^{(r)}, ..., C_l^{(r)}\}$. Without loss of generality assume $C_k^{(r)}$ can be represented as follows (like before for $r = 2$ which is the base case):

$$C_k^{(r)}(W_{r-1}) = \sum_{i=1}^{s_k} \prod_{j=1}^{a} (1 + z_{i_{1j}} t_{1j} + z_{i_{2\sigma_{k_2}(j)}} t_{2\sigma_{k_2}(j)} + ... + z_{i_{(r-1)\sigma_{k_{(r-1)}}(j)}} t_{(r-1)\sigma_{k_{(r-1)}}(j)}$$

$$+ z_{i_{r\sigma_{k_r}(j)}} x_{r\sigma_{k_r}(j)} + ... + z_{i_{m\sigma_{k_m}(j)}} x_{m\sigma_{k_m}(j)}).$$

Step r: In step r we shift the X_r variables by T_r. We show that after this shift there exists in the (shifted) circuit $C^{(r)}(W_{r-1})$ a monomial in X_r variables with support less than $\log s$ such that it has a non-zero coefficient polynomial in $\mathbb{F}[Y_r \cup W_r]$. For this coefficient polynomial to remain non-zero under a substitution $t_{ij} = t^{\omega_{ij}}$, we need to keep a small collection of polynomials in Y_r-variables non-zero where each polynomial has few Y_r-variables – this is explained below. Thus, after the r-th step we have a monomial in U_r variables of support less than $r \log s$, in the (shifted) circuit C, that has a non-zero coefficient polynomial in $\mathbb{F}[Y_r \cup W_r]$. Let us see the details now.

For all $k \in [l]$, we rewrite $C_k^{(r)}(W_{r-1})$, such that, we can associate the identity permutation with the base set X_r.

$$C_k^{(r)}(W_{r-1}) = \sum_{i=1}^{s_k} \prod_{j=1}^{a} (1 + z_{i_{1\pi_{k_1}(j)}} t_{1\pi_{k_1}(j)} + z_{i_{2\pi_{k_2}(j)}} t_{2\pi_{k_2}(j)} + ... + z_{i_{(r-1)\pi_{k_{(r-1)}}(j)}} t_{(r-1)\pi_{k_{(r-1)}}(j)}$$

$$+ z_{i_{rj}} x_{rj} + ... + z_{i_{m\pi_{k_m}(j)}} x_{m\pi_{k_m}(j)}).$$

Here again π_{k_q} represents the permutation function $[a] \to [a]$ corresponding to the q-th base set. We view $C^{(r)}(W_{r-1})$ as a polynomial in X_r variables over $\mathbb{F}(Y_{r-1} \uplus W_r)$. For $J \subseteq [a]$, $X_{rJ} := \prod_{j \in J} x_{rj}$. The coefficient vector of a monomial X_{rJ} is an s component vector $Z_{X_{rJ}}$, where the i-th component is the coefficient of the monomial in the i-th product gate of $C^{(r)}(W_{r-1})$. Assume the i-th product gate is a part of $C_k^{(r)}(W_{r-1})$. Then the coefficient of the monomial X_{rJ} in the i-th product gate is

$$\prod_{j \in J} z_{i_{rj}} \prod_{j \in [a] \setminus J} (1 + z_{i_{1\pi_{k_1}(j)}} t_{1\pi_{k_1}(j)} + ... + z_{i_{(r-1)\pi_{k_{(r-1)}}(j)}} t_{(r-1)\pi_{k_{(r-1)}}(j)}$$

$$+ z_{i_{(r+1)\pi_{k_{(r+1)}}(j)}} x_{(r+1)\pi_{k_{r+1}}(j)} + ... + z_{i_{m\pi_{k_m}(j)}} x_{m\pi_{k_m}(j)}).$$

Pick a monomial $X_{rJ'}$ whose support is exactly $\log s + 1$[14]. Say we pick $X_{rJ'}$, corresponding to $J^{'} = [\log s + 1]$. Consider the monomial $X_{rJ'}$ and all its subset monomials X_{rJ} corresponding to $J \subseteq J^{'}$; there are exactly $2^{\log s+1} > s$ many such monomials. Hence we have a linear dependency among the coefficient vectors of these monomials, i.e.

$$\sum_{J \subseteq [\log s+1]} b_J Z_{X_{rJ}} = 0,$$

where $\forall\ J \subseteq [\log s + 1],\ b_J \in \mathbb{F}(Y_{r-1} \uplus W_r)$ and $\exists\ J \subseteq [\log s + 1]$ such that $b_J \neq 0$. Now we shift the variables in X_r, i.e. $x_{rj} = x_{rj} + t_{rj}$. Let $\tilde{C}^{(r)}$ correspond to $C^{(r)}$ and for all $k \in [l]$, $\tilde{C}^{(r)}_k$ correspond to $C^{(r)}_k$ after this shift. Thus for all $k \in [l]$ we have

$$\tilde{C}^{(r)}_k(W_{r-1}) = \sum_{i=1}^{s_k} \prod_{j=1}^{a} (1 + z_{i_{1\pi_{k_1}(j)}} t_{1\pi_{k_1}(j)} + z_{i_{2\pi_{k_2}(j)}} t_{2\pi_{k_2}(j)} + ... + z_{i_{(r-1)\pi_{k_{(r-1)}}(j)}} t_{(r-1)\pi_{k_{(r-1)}}(j)}$$

$$+ z_{i_{rj}}(x_{rj} + t_{rj}) + ... + z_{i_{m\pi_{k_m}(j)}} x_{m\pi_{k_m}(j)}).$$

For $i \in [s_k]$ and $j \in [a]$ let

$$\rho_{ij} := (z_{i_{1\pi_{k_1}(j)}} t_{1\pi_{k_1}(j)} + z_{i_{2\pi_{k_2}(j)}} t_{2\pi_{k_2}(j)} + ... + z_{i_{(r-1)\pi_{k_{(r-1)}}(j)}} t_{(r-1)\pi_{k_{(r-1)}}(j)}$$

$$+ z_{i_{(r+1)\pi_{k_{(r+1)}}(j)}} x_{(r+1)\pi_{k_{(r+1)}}(j)} + ... + z_{i_{m\pi_{k_m}(j)}} x_{m\pi_{k_m}(j)}).$$

Then

$$\tilde{C}^{(r)}_k(W_{r-1}) = \sum_{i=1}^{s_k} \prod_{j=1}^{a} (1 + \rho_{ij} + z_{i_{rj}}(x_{rj} + t_{rj})).$$

In particular we have

$$\tilde{C}^{(r)}(W_{r-1}) = \sum_{i=1}^{s} \prod_{j=1}^{a} (1 + \rho_{ij} + z_{i_{rj}}(x_{rj} + t_{rj}))$$

$$\Rightarrow \tilde{C}^{(r)}(W_{r-1}) = \sum_{i=1}^{s} \prod_{j=1}^{a} (1 + \rho_{ij} + z_{i_{rj}} t_{rj}) \cdot \prod_{j=1}^{a} (1 + \frac{z_{i_{rj}} x_{rj}}{1 + \rho_{ij} + z_{i_{rj}} t_{rj}}).$$

Let

$$z^{'}_{i_{rj}} = \frac{z_{i_{rj}}}{1 + \rho_{ij} + z_{i_{rj}} t_{rj}} \quad \Rightarrow \quad z_{i_{rj}} = \frac{z^{'}_{i_{rj}}(1 + \rho_{ij})}{1 - z^{'}_{i_{rj}} t_{rj}}.$$

$C^{(r)}(W_{r-1})$ is non-zero if and only if $\tilde{C}^{(r)}(W_{r-1})$ is non-zero, as shifting is an invertible operation. We intend to show that $\tilde{C}^{(r)}(W_{r-1})$ is non-zero if and only if there exists a low support monomial of X_r variables in $\tilde{C}^{(r)}(W_{r-1})$ that has a non-zero coefficient polynomial in $\mathbb{F}[Y_r \uplus W_r]$. To prove this let $Z^{'}_{X_{rJ}}$ be an s-component vector whose i-th component is $\prod_{j \in J} z^{'}_{i_{rj}}$ where $J \subseteq [a]$. Observe that the coefficient of X_{rJ} in $\tilde{C}^{(r)}$ is the dot product of $Z^{'}_{X_{rJ}}$ with another vector whose i-th component is $\prod_{j=1}^{a}(1 + \rho_{ij} + z_{i_{rj}} t_{rj})$. We show the following next,

$$\text{span}_{\mathbb{F}(Y_r \uplus W_r)}\{Z^{'}_{X_{rJ}} | J \subseteq [a]\} \ = \ \text{span}_{\mathbb{F}(Y_r \uplus W_r)}\{Z^{'}_{X_{2J}} | J \subseteq [a], |J| \leq \log s\}.$$

[14]We avoid ceil, floor notation for ease of exposition.

This immediately implies if $\tilde{C}^{(r)} \neq 0$ then $\tilde{C}^{(r)}$ has a monomial in X_r-variables of support at most $\log s$ with its coefficient polynomial in $\mathbb{F}[Y_r \uplus W_r]$. Recall

$$\sum_{J \subseteq [\log s+1]} b_J Z_{X_{rJ}} = 0, \quad \text{where } b_J \in \mathbb{F}(Y_{r-1} \uplus W_r).$$

We write the above equation for the i-th component,

$$\sum_{J \subseteq [\log s+1]} b_J \prod_{j \in J} z_{i_{rj}} \prod_{j \in [a] \setminus J} (1 + \rho_{ij}) = 0.$$

Since $\prod_{j=\log s+1}^{a}(1 + \rho_{ij}) \neq 0$ we have

$$\sum_{J \subseteq [\log s+1]} b_J \prod_{j \in J} z_{i_{rj}} \prod_{j \in [\log s+1] \setminus J} (1 + \rho_{ij}) = 0$$

$$\Rightarrow \sum_{J \subseteq [\log s+1]} b_J \prod_{j \in J} \frac{z'_{i_{rj}}(1 + \rho_{ij})}{1 - z'_{i_{rj}} t_{rj}} \prod_{j \in [\log s+1] \setminus J} (1 + \rho_{ij}) = 0$$

$$\Rightarrow \sum_{J \subseteq [\log s+1]} b_J \prod_{j \in J} \frac{z'_{i_{rj}}}{1 - z'_{i_{rj}} t_{rj}} = 0.$$

Multiplying both sides by $\prod_{j \in [\log s+1]}(1 - z'_{i_{rj}} t_{rj})$

$$\sum_{J \subseteq [\log s+1]} b_J \prod_{j \in J} z'_{i_{rj}} \prod_{j \in [\log s+1] \setminus J} (1 - z'_{i_{rj}} t_{rj}) = 0.$$

Since this is true for any $i \in [s]$, we have

$$\sum_{J \subseteq [\log s+1]} b_J \cdot Z'_{rJ} \prod_{j \in [\log s+1] \setminus J} (1 - Z'_{r\{j\}} t_{rj}) = 0$$

$$\Rightarrow \underbrace{\Big(\sum_{J \subseteq [\log s+1]} b_J \cdot (-1)^{\log s+1-|J|} \prod_{j \in [\log s+1] \setminus J} t_{rj} \Big)}_{g_{[\log s+1]}(Y_r \uplus W_r)} \cdot Z'_{r[\log s+1]} + \sum_{J \subset [\log s+1]} g_J(Y_r \uplus W_r) \cdot Z'_{rJ} = 0$$

Since $\forall\ J \subseteq [\log s + 1]$, $b_J \in \mathbb{F}(Y_{r-1} \uplus W_r)$, $g_{[\log s+1]}(Y_r \uplus W_r)$ is non-zero, and hence $Z'_{r[\log s+1]}$ is $\mathbb{F}(Y_r \uplus W_r)$ linearly dependent on vectors Z'_{rJ} for $J \subset [\log s + 1]$. The set $[\log s + 1]$ is just a representative case. By the same argument we have that Z'_{rJ}, where $J \subseteq [a]$ and $|J| = \log s + 1$ is $\mathbb{F}(Y_r \uplus W_r)$ linearly dependent on vectors $Z'_{rJ'}$ where $J' \subset J$. Thus, for every $J \subseteq [a]$, Z'_{rJ} can be inductively expressed as $\mathbb{F}(Y_r \uplus W_r)$ linear combinations of $Z'_{rJ'}$, where $J' \subseteq [a]$ and $|J'| \leq \log s$.

The crux of the argument is to show that $g_{[\log s+1]}(Y_r \uplus W_r)$ remains non-zero even under an efficient map $t_{ij} = t^{\omega_{ij}}$ which we argue next. If $C(X)$ computes a non-zero polynomial then after the r-th step we have a monomial in U_r variables with support less than $r \log s$ such that it has a non-zero coefficient polynomial over $\mathbb{F}(Y_r \uplus W_r)$. Hence after m steps there exists a monomial in X

variables with support less than $m \log s$, such that it has a non-zero coefficient polynomial over $\mathbb{F}(T)$. We apply a map ψ that maps t_{ij} to $t^{\omega_{ij}}$, where for all $i \in [m]$ and $j \in [a]$, $\omega_{ij} \in \mathbb{N}$ such that ω_{ij}'s are bounded by a polynomial in $n = m \cdot a$ and the non-zero coefficient polynomial over $\mathbb{F}(T)$ corresponding to the monomial in X variables with support less than $m(\log s + 1)$, continues to be non-zero over $\mathbb{F}(t)$, as ψ preserves the non-zeroness of $g_J(Y_r \uplus W_r)$ at the r-th step for every $J \subseteq [a]$ and $|J| \leq \log s + 1$. Map ψ can be constructed in time $(sn)^{O(lm \log s)}$. To see how we try to understand the structure of $g_{[\log s+1]}(Y_r \uplus W_r)$ in the following claim.

Claim 28. *The term $g_{[\log s+1]}(Y_r \uplus W_r)$ in the above argument is a rational function in $\mathbb{F}(Y_r \uplus W_r)$ with at most $O(ml \log s)$ distinct Y_r and W_r variables appearing in it. Further, the degree of the polynomials in the numerator and denominator of $g_{[\log s+1]}(Y_r \uplus W_r)$ is bounded by $O(s^2 \log s)$.*

Proof. We show that there are $O(ml \log s)$ variables in $Y_{r-1} \cup W_r$ such that every b_J in the expression for $g_{[\log s+1]}(Y_r \uplus W_r)$ is a rational function in these variables. Further, the degree of the polynomials in the numerator and denominator of b_J is bounded by $O(s \log s)$. Recall that

$$\sum_{J \subseteq [\log s+1]} b_J Z_{X_{rJ}} = 0.$$

Focusing on the i-th component and assuming the i-th product gate us part of $C_k^{(r)}$,

$$\sum_{J \subseteq [\log s+1]} b_J \prod_{j \in J} z_{i_{rj}} \prod_{j \in [a] \setminus J} (1 + \rho_{ij}) = 0$$

$$\Rightarrow \sum_{J \subseteq [\log s+1]} b_J \prod_{j \in J} z_{i_{rj}} \prod_{j \in [\log s+1] \setminus J} (1 + \rho_{ij}) = 0.$$

Define a vector $\tilde{Z}_{X_{rJ}}$ whose i-th component is $\prod_{j \in J} z_{i_{rj}} \cdot \prod_{j \in [\log s+1] \setminus J} (1 + \rho_{ij})$. Then

$$\sum_{J \subseteq [\log s+1]} b_J \tilde{Z}_{X_{rJ}} = 0.$$

Observe that ρ_{ij} is a linear polynomial in $(m-1)$ variables from $Y_{r-1} \cup W_r$ and so $\prod_{j \in [\log s+1] \setminus J} (1 + \rho_{ij})$ is a polynomial in at most $m \log s$ variables from $Y_{r-1} \cup W_r$. Further, for $i \neq i'$ the set of variables in $\prod_{j \in [\log s+1] \setminus J} (1 + \rho_{ij})$ is the same as the set of variables in $\prod_{j \in [\log s+1] \setminus J} (1 + \rho_{i'j})$ if both the i-th and the i'-th product gates are part of the same circuit $C_k^{(r)}$. Hence, there is a set of at most $lm \log s$ variables from $Y_{r-1} \cup W_r$ such that every entry in every $\tilde{Z}_{X_{rJ}}$ (for $J \subseteq [\log s + 1]$) is a polynomial in these $lm \log s$ variables of degree bounded by $\log s + 1$. Applying Cramer's rule, every b_J is also a rational function in the same $lm \log s$ variables from $Y_{r-1} \cup W_r$ with degree of the numerator and the denominator bounded by $O(s \log s)$. Finally, from the expression for $g_{[\log s+1]}(Y_r \uplus W_r)$ one can readily infer that it is also a rational function in $O(ml \log s)$ variables from $Y_r \cup W_r$ with degree of the numerator and the denominator bounded by $O(s^2 \log s)$.

□

Since the numerator and denominator of $g_{[\log s+1]}(Y_r \uplus W_r)$ can have at most $s^{O(ml \log s)}$ monomials, it is fairly standard to construct a map $\psi : t_{ij} \to t^{\omega_{ij}}$ in time $s^{O(ml \log s)}$ that keeps $g_{[\log s+1]}(Y_r \uplus W_r)$ non-zero. The overall complexity would be $(ns)^{O(ml \log s)}$ as one has to ensure non-zeroness of $g_J(Y_r \uplus W_r)$ for every $J \subseteq [a]$ and $|J| = \log s + 1$.

□