Subexponential Size Hitting Sets for Bounded Depth Multilinear Formulas

Rafael Oliveira* Amir Shpilka† Ben Lee Volk‡

Abstract

In this paper we give subexponential size hitting sets for bounded depth multilinear arithmetic formulas. Using the known relation between black-box PIT and lower bounds we obtain lower bounds for these models.

For depth-3 multilinear formulas, of size $\exp(n^{\delta})$, we give a hitting set of size $\exp\left(\tilde{O}\left(n^{2/3+2\delta/3}\right)\right)$. This implies a lower bound of $\exp(\tilde{\Omega}(n^{1/2}))$ for depth-3 multilinear formulas, for some explicit polynomial.

For depth-4 multilinear formulas, of size $\exp(n^{\delta})$, we give a hitting set of size $\exp\left(\tilde{O}\left(n^{2/3+4\delta/3}\right)\right)$. This implies a lower bound of $\exp(\tilde{\Omega}(n^{1/4}))$ for depth-4 multilinear formulas, for some explicit polynomial.

A regular formula consists of alternating layers of $+, \times$ gates, where all gates at layer i have the same fan-in. We give a hitting set of size (roughly) $\exp\left(n^{1-\delta}\right)$, for regular depth-$d$ multilinear formulas with formal degree at most n and size $\exp(n^{\delta})$, where $\delta = O(1/\sqrt{5}^{d})$. This result implies a lower bound of roughly $\exp(\tilde{\Omega}(n^{1/\sqrt{5}^{d}}))$ for such formulas.

We note that better lower bounds are known for these models, but also that none of these bounds was achieved via construction of a hitting set. Moreover, no lower bound that implies such PIT results, even in the white-box model, is currently known.

Our results are combinatorial in nature and rely on reducing the underlying formula, first to a depth-4 formula, and then to a read-once algebraic branching program (from depth-3 formulas we go straight to read-once algebraic branching programs).

*Department of Computer Science, Princeton University. Research supported by NSF grant CCF-1217416 and by the Sloan fellowship. Email: rmo@cs.princeton.edu.

†Department of Computer Science, Tel Aviv University, Tel Aviv, Israel, shpilka@post.tau.ac.il. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 257575, and from the Israel Science Foundation (grant number 339/10).

‡Department of Computer Science, Tel Aviv University, Tel Aviv, Israel, benleevolk@gmail.com. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 257575.

1 Introduction

Arithmetic circuits are the standard model for computing polynomials. Roughly speaking, given a set of variables $X = \{x_1, ..., x_n\}$, an arithmetic circuit uses additions and multiplications to compute a polynomial f in the set of variables X. An arithmetic formula is an arithmetic circuit whose computation graph is a tree. An arithmetic circuit (or formula) is multilinear if the polynomial computed at each of its gates is multilinear (as a formal polynomial), that is, in each of its monomials the power of every input variable is at most one (see Section 1.1 for definition of the models studied in this paper).

Two outstanding open problems in complexity theory are to prove exponential lower bounds on the size of arithmetic circuits, i.e., to prove a lower bound on the number of operations required to compute some polynomial f, and to give efficient deterministic polynomial identity testing (PIT for short) algorithms for them. The PIT problem for arithmetic circuits asks the following question: given an arithmetic circuit Φ computing a polynomial f, determine, *efficiently and deterministically*, whether "$f \equiv 0$". The black-box version of the PIT problem asks to construct a small *hitting set*, i.e., a set of evaluation points $\mathcal{H}$, for which any such non-zero f does not vanish on all the points in $\mathcal{H}$.

It is known that solving any one of the problems (proving a lower bound or deterministic PIT), with appropriate parameters, for small depth (multilinear) formulas, is equivalent to solving it in the general (multilinear) case [VSBR83, AV08, Koi10, GKKS13, Tav13]. It is also known that these two problems are tightly connected and that solving one would imply a solution to the other, both in the general case [HS80, KI03, Agr05] and in the bounded depth case[1] [DSY09]. We note that in the multilinear case it is only known that hitting sets imply circuit lower bounds but not vice versa.

In this work we study the PIT problem for several models of bounded depth multilinear formulas. Our main results are subexponential size hitting sets for depth-3 and depth-4 multilinear formulas of subexponential size and for *regular* depth-d multilinear formulas of subexponential size (with construction size deteriorating among the different models). Using the connection between explicit hitting sets and circuit lower bounds we get, as corollaries, subexponential lower bounds for these models.

1.1 Models for Computing Multilinear Polynomials

An arithmetic circuit Φ over the field $\mathbb{F}$ and over the set of variables X is a directed acyclic graph as follows. Every vertex in Φ of in-degree 0 is labelled by either a variable in X or a field element in $\mathbb{F}$. Every other vertex in Φ is labelled by either $\times$ or $+$. An arithmetic circuit is called a formula if it is a directed tree (whose edges are directed from the leaves to the root). The vertices of Φ are also called gates. Every gate of in-degree 0 is called an input gate. Every gate of out-degree 0 is called an output gate. Every gate labelled by $\times$ is called a product gate. Every gate labelled by $+$ is called a sum gate. An arithmetic circuit computes a polynomial in a natural way. An input gate labelled by $y \in \mathbb{F} \cup X$ computes the polynomial y. A product

[1] The result of [DSY09] is more restricted than the results for circuits with no depth restrictions.

gate computes the product of the polynomials computed by its children. A sum gate computes the sum of the polynomials computed by its children.

A polynomial $f \in \mathbb{F}[X]$ is called multilinear if the degree of each variable in f is at most one. An arithmetic circuit (formula) Φ is called multilinear if every gate in Φ computes a multilinear polynomial.

In this work we are interested in small depth multilinear formulas. A depth-3 $\Sigma\Pi\Sigma$ formula is a formula composed of three layers of alternating sum and product gates. Thus, every polynomial computed by a $\Sigma\Pi\Sigma$ formula of size s has the following form

$$f = \sum_{i=1}^{s} \prod_{j=1}^{d_i} \ell_{i,j},$$

where the $\ell_{i,j}$ are linear functions. In a $\Sigma\Pi\Sigma$ multilinear formula, it holds that in every product gate $\prod_{j=1}^{d_i} \ell_{i,j}$, the linear functions $\ell_{i,1}, \ldots, \ell_{i,d_i}$ are supported on disjoint sets of variables.

Similarly, a depth-4 $\Sigma\Pi\Sigma\Pi$ formula is a formula composed of four layers of alternating sum and product gates. Thus, every polynomial computed by a $\Sigma\Pi\Sigma\Pi$ formula of size s has the following form

$$f = \sum_{i=1}^{s} \prod_{j=1}^{d_i} Q_{i,j},$$

where the $Q_{i,j}$ are computed at the bottom $\Sigma\Pi$ layers and are s-sparse polynomials, i.e., polynomials that have at most s monomials. As in the depth-3 case, we have that at every product gate the polynomials $Q_{i,1}, \ldots, Q_{i,d_i}$ are supported on disjoint sets of variables.

Another important definition for us is that of a regular depth-d formula. A regular depth-d formula is specified by a list of d integers $(a_1, p_1, a_2, p_2, \ldots)$. It has d layers of alternating sum and product gates. The fan-in of every sum gate at the $(2i-1)$'th layer is a_i and, similarly, the fan-in of every product gate at the $(2i)$'th layer is p_i. For example, a depth-4 formula that is specified by the list (a_1, p_1, a_2, p_2) has the following form:

$$f = \sum_{i=1}^{a_1} \prod_{j=1}^{p_1} Q_{i,j},$$

where each $Q_{i,j}$ is a polynomial of degree p_2 that has (at most) a_2 monomials. As before, a regular depth-d multilinear formula is a regular depth-d formula in which every gate computes a multilinear polynomial. For such formulas, we also impose the restriction that their formal degree is at most n, i.e., $\prod_i p_i \leq n$.

Regular formulas were first defined by Kayal, Saha and Saptharishi [KSS14], who proved quasi-polynomial lower bounds for logarithmic-depth regular formulas. It is interesting to note that in the reductions from general (multilinear) circuits/formulas to depth-d (multilinear) formulas, one gets a regular depth-d (multilinear) formula [VSBR83, AV08, Koi10, Tav13].

Finally, we also need to consider the model of Read-Once Algebraic Branching Programs (ROABPs) as our construction is based on a reduction to this model. Algebraic Branching

Programs (or ABPs, for short) were first defined in the work of Nisan [Nis91], who proved exponential lower bounds on the size of non-commutative ABPs computing the determinant or permanent polynomials. Roughly, an ABP consists of a layered graph with edges going from the i'th layer to the $(i+1)$'th layer. The first layer consists of a single source node and the last layer contains a single sink. The edges of the graph are labeled with polynomials (in our case we only consider linear functions as labels). The weight of a path is the product of the weights of the edges in the path. The polynomial computed by the ABP is the sum of the weights of all the paths from the source to the sink. An ABP is called a read-once ABP (ROABP) if the only variable appearing on edges that connect the i'th and the $(i+1)$'th layer is x_i. It is clear that a ROABP whose edges are labeled with linear functions computes a multilinear polynomial.

1.2 Polynomial Identity Testing

In the PIT problem we are given an arithmetic circuit or formula Φ, computing some polynomial f, and we have to determine whether "$f \equiv 0$". That is, we are asking if f is the zero polynomial in $\mathbb{F}[x_1, \ldots, x_n]$. By the Schwartz-Zippel-DeMillo-Lipton lemma [Zip79, Sch80, DL78], if $0 \neq f \in \mathbb{F}[x_1, \ldots, x_n]$ is a polynomial of degree $\leq d$, and $\alpha_1, \ldots, \alpha_n \in A \subseteq \mathbb{F}$ are chosen uniformly at random, then $f(\alpha_1, \ldots, \alpha_n) = 0$ with probability at most[2] $d/|A|$. Thus, given Φ, we can perform these evaluations efficiently, giving an efficient randomized procedure for answering "$f \equiv 0$?". It is an important open problem to find a derandomization of this algorithm, that is, to find a *deterministic* procedure for PIT that runs in polynomial time (in the size of Φ).

One interesting property of the above randomized algorithm of Schwartz-Zippel-DeMillo-Lipton is that the algorithm does not need to "see" the circuit Φ. Namely, the algorithm only uses the circuit to compute the evaluation $f(\alpha_1, \ldots, \alpha_n)$. Such an algorithm is called a *black-box* algorithm. In contrast, an algorithm that can access the internal structure of the circuit Φ is called a *white-box* algorithm. Clearly, the designer of the algorithm has more resources in the white-box model and so one can expect that solving PIT in this model should be a simpler task than in the black-box model.

The problem of derandomizing PIT has received a lot of attention in the past few years. In particular, many works examine a specific class of circuits $\mathcal{C}$, and design PIT algorithms only for circuits in that class. One reason for this attention is the strong connection between deterministic PIT algorithms for a class $\mathcal{C}$ and lower bounds for $\mathcal{C}$. This connection was first observed by Heintz and Schnorr [HS80] (and later also by Agrawal [Agr05]) for the black-box model and by Kabanets and Impagliazzo [KI04] for the white-box model (see also [DSY09] for a similar result for bounded depth circuits). Another motivation for studying the problem is its relation to algorithmic questions. Indeed, the famous deterministic primality testing algorithm of Agrawal, Kayal and Saxena [AKS04] is based on derandomizing a specific polynomial identity. Finally, the PIT problem is, in some sense, the most general problem that we know today for which we have randomized coRP algorithms but no polynomial time algorithms, thus studying it is a natural step towards a better understanding of the relation between RP and P. For more on the PIT problem we refer to the survey by Shpilka and Yehudayoff [SY10].

[2]Note that this is meaningful only if $d < |A| \leq |\mathbb{F}|$, which in particular implies that f is not the zero function.

Among the most studied circuit classes we find Read-Once Algebraic Branching Programs [FS13, FSS14, AGKS15, GKST15], set-multilinear formulas [RS05, FS12, ASS13], depth-3 formulas [DS06, KS07, KS11, KS09, SS11], depth-4 formulas [KMSV13, SV11, ASSS12, Gup14] and bounded-read multilinear formulas [SV08, SV09, AvMV11, ASSS12]. We note that none of these results follow from a reduction a la [KI04] (or the reduction of [DSY09] for bounded depth circuits) from PIT to lower bounds. Indeed, this reduction does not work for the restricted classes mentioned here. In particular, for the multilinear model no reduction from PIT to lower bounds is known. That is, even given lower bounds for multilinear circuits/formulas (e.g., the subexponential lower bound of [RY09] for constant depth multilinear formulas) we do not know how to construct a PIT algorithm for a related model.

The works on depth-3 and multilinear depth-4 formulas gave polynomial time algorithms only when the fan-in of the top gate (the output gate) is constant, and became exponential time when the top fan-in was $\Omega(n)$, both in the white-box and black-box models [KS07, SS11, SV11, Gup14]. [RS05] gave a polynomial time PIT for set-multilinear depth-3 circuits and [FS13] and [ASS13] gave a quasi-polynomial size hitting set for this model. Recall that in a depth-3 set-multilinear formula, the variables are partitioned to sets, and each linear function at the bottom layer only involves variables from a single set.

Recently, [AGKS15] gave a subexponential white-box algorithm for a depth-3 formula that computes the sum of c set-multilinear formulas, each of size s, with respect to different partitions of the variables. The running time of their algorithm is $n^{O(2^c n^{1-2/2^c} \log s)}$. In particular, for $c = \Omega(\log\log(n))$ the running time is $\exp(n)$. Shortly after, [GKST15] gave a white box algorithm for sums of c ROABPs, that runs in time $(ndw^{2^c})^{O(c)}$ for ROABPs with n variables, individual degree d and width w. They also give a black box algorithm that runs in time $(ndw)^{O(c2^c \log(ndw))}$. Thus, while for constant c this gives a significant improvement over [AGKS15] for this model, the doubly exponential dependence on c still trivializes those constructions if $c = \Omega(\log n)$.

To conclude, prior to this work there were no subexponential PIT algorithms, even for depth-3 multilinear formulas with top fan-in n.

1.3 Our Results

Remark. *Throughout this paper, we assume that for formulas of size 2^{n^δ}, the underlying field $\mathbb{F}$ is of size at least $|\mathbb{F}| \geq 2^{n^{2\delta}\mathsf{poly}\log(n)}$, and that if this is not the case then we are allowed to query the formula on inputs from an extension field of the appropriate size. In particular, all our results hold over fields of characteristic zero or over fields of size* $\exp(n)$.

We give subexponential size hitting sets for depth-3, depth-4 and regular depth-d multilinear formulas, of subexponential size. In particular we obtain the following results.

Theorem 1.1. *There exists an explicit hitting set $\mathcal{H}$ of size $2^{\tilde{O}(n^{2/3+2\delta/3})}$ for the class of $\Sigma\Pi\Sigma$ multilinear formulas of size 2^{n^δ}.*

This gives a significant improvement to the recent result, mentioned above, of [AGKS15] who studied sums of set-multilinear formulas. From the connection between hitting sets and

circuit lower bounds [HS80, Agr05] we obtain the following corollary.

Corollary 1.2. *There is an explicit multilinear polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$, whose coefficients can be computed in exponential time, such that any depth-3 multilinear formula for f has size $\exp(\tilde{\Omega}(\sqrt{n}))$.*

This lower bound is weaker than the exponential lower bound of [NW96] for this model .[3] Yet, it is interesting to note that we can get such a strong lower bound from a PIT algorithm. Next, we present our result for depth-4 multilinear formulas.

Theorem 1.3. *There exists an expicit hitting set $\mathcal{H}$ of size $2^{\tilde{O}(n^{2/3+4\delta/3})}$ for the class of $\Sigma\Pi\Sigma\Pi$ multilinear formulas of size 2^{n^δ}.*

Again, from the connection between hitting sets and circuit lower bounds we obtain the following corollary.

Corollary 1.4. *There is an explicit multilinear polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$, whose coefficients can be computed in exponential time, such that any depth-4 multilinear formula for f has size $\exp(\tilde{\Omega}(n^{1/4}))$.*

The best known lower bound for depth-4 multilinear formulas is $\exp(n^{1/2})$ due to [RY09], thus, as in the previous case, the term in the exponent of our lower bound is the square root of the corresponding term in the best known lower bound. For regular depth-d multilinear formulas we obtain the following result.

Theorem 1.5. *There exists an explicit hitting set $\mathcal{H}$ of size $2^{\tilde{O}(n^{1-\delta/3})}$ for the class of regular depth-d multilinear formulas of size 2^{n^δ}, where $\delta \leq \frac{1}{5^{\lfloor d/2 \rfloor + 1}} = O\left(\frac{1}{\sqrt{5}^d}\right)$.*

As before we obtain a lower bound for such formulas.

Corollary 1.6. *There is an explicit multilinear polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$, whose coefficients can be computed in exponential time, such that any regular depth-d multilinear formula for f has size $\exp(\tilde{\Omega}(n^{\frac{1}{5^{\lfloor d/2 \rfloor + 1}}}))$.*

We note that Raz and Yehudayoff gave an $\exp(n^{\Omega(\frac{1}{d})})$ lower bound for depth-d multilinear formulas, which is much stronger than what Corollary 1.6 gives. Yet, our result also gives a PIT algorithm, which does not follow from the results of [RY09]. As we later explain, we lose a square root in the term at the exponent for every increase of the depth and this is the reason that we get only $\exp(n^{1/\exp(d)})$ instead of $\exp(n^{1/d})$.

In addition to lower bounds, our work also implies deterministic factorization of multilinear polynomials. [SV10] proved that if one can perform PIT deterministically for certain classes of multilinear polynomials then a deterministic factoring algorithm for those classes follows. Specifically, for a class of polynomials $\mathcal{C}$ they defined the class $\mathcal{C}_V$, consisting of all polynomials that can be computed by circuits of the form $C = C_1 + C_2 \times C_3$, where the circuits C_i belong

[3] It is also weaker in the sense that the lower bound of Nisan and Wigderson applies to a polynomial in VP, the arithmetic analog of P, whereas our lower bound is for a polynomial in $\mathsf{DTIME}(2^{O(n)})$.

to the class $\mathcal{C}$ and the circuits C_2 and C_3 are defined over disjoint sets of variables. They proved that if the class $\mathcal{C}_V$ has a deterministic PIT that runs in time $T(n,s)$ for circuits on n variables of size s then there is a deterministic factoring algorithm for the class $\mathcal{C}$ that runs in time $O(n^3 \cdot T(n,s))$ (Theorem 1.1 in [SV10]). Furthermore, if $\mathcal{C}$ is a multilinear circuit class, f is a polynomial computed by a size s circuit from $\mathcal{C}$ and g is a factor of f, then g can also be computed by a circuit from $\mathcal{C}$ of size at most s (this follows from the fact that the irreducible factors of f are variable disjoint; we again refer to [SV10] for the details).

In our case, since the product of two variable disjoint multilinear $\Sigma\Pi\Sigma$ ($\Sigma\Pi\Sigma\Pi$) formulas of size 2^{n^δ} is a multilinear $\Sigma\Pi\Sigma$ ($\Sigma\Pi\Sigma\Pi$) formula of size 2^{2n^δ}, which is still inside of the class $\Sigma\Pi\Sigma$ ($\Sigma\Pi\Sigma\Pi$), the result of [SV10], when combined with our PIT results, implies that we can deterministically factor such formulas. Therefore, we obtain the following corollary:

Corollary 1.7 (Deterministic Factorization)**.** *Given a multilinear polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ that can be computed by a $\Sigma\Pi\Sigma$ ($\Sigma\Pi\Sigma\Pi$) formula of size 2^{n^δ}, there exists an efficient deterministic algorithm that outputs the factors of f. The algorithm outputs $\Sigma\Pi\Sigma$ ($\Sigma\Pi\Sigma\Pi$) formulas for the factors if the formula for f is given to it explicitly, and black-boxes if it only has black-box access to f. The running time of this algorithm is $2^{\tilde{O}(n^{2/3+2\delta/3})}$ when f is computed by a $\Sigma\Pi\Sigma$ formula and $2^{\tilde{O}(n^{2/3+4\delta/3})}$ when it is computed by a $\Sigma\Pi\Sigma\Pi$ formula.*

1.4 Proof Overview

We first discuss our proof technique for the case of depth-3 multilinear formulas. Our (idealized) aim is to reduce such a formula Φ to a depth-3 multilinear formula in which each linear function is of the form $\alpha x + \beta$. That is, each linear function contains at most one variable. If we manage to do that then we can use the quasi-polynomial sized hitting set of [FS13, AGKS15] for this model.

Of course, the problem with the above argument is that in general, depth-3 formulas have more than one variable per linear function. To overcome this difficulty, we will partition the variables to several sets $T_1, \ldots, T_m$ and hope that each linear function in the formula contains at most one variable from each T_i. If we can do that then we would use the hitting set for each set of variables T_i and combine those sets together to get our hitting set. That is, the combined hitting set is constructed by "concatenating" the smaller hitting sets, so that it contains all vectors $v \in \mathbb{F}^n$ such that the restriction of v to the indices of T_i belongs to the hitting set constructed for the variables of T_i. Thus, if we can carry out this procedure then we will get a hitting set of size roughly $n^{m \log n}$. This step indeed yields a hitting set, since when we restrict our attention to each T_i and think of the other variables as constants in some huge extension field, then we do get a small ROABP (in the variables of T_i) and hence plugging in the hitting set of [FS13] and [AGKS15] gives a non-zero polynomial. Thus, we can first do this for T_1 and obtain some good assignment v_1 that makes the polynomial non-zero after substituting the variables in T_1 to v_1. Then we can find v_2, etc.

There are two problems with the above argument. One problem is how to find such a good partition. The second is that this idea simply cannot work as is. For example, if we have the linear function $x_1 + \cdots + x_n$, then it will have a large intersection with each T_i.

We first deal with the second question. To overcome the difficulty posed by the example, we would like to somehow "get rid" of all linear functions of large support and then carry out the idea above. To remove linear functions with large support from the formula we use another trick. Consider a variable x_k that appears in a linear function ℓ_0 that has a large support. Assume that $\frac{\partial f}{\partial x_k} \not\equiv 0$ as otherwise we can ignore x_k. Now, because of multilinearity, we can transform our original formula Φ to a formula computing $\frac{\partial f}{\partial x_k}$. This is done by replacing each linear function $\ell(X) = \sum_{i=1}^{n} \alpha_i x_i + \alpha_0$ with the constant α_k. In particular, the function ℓ_0 that used to have a high support does not appear in the new formula. Furthermore, this process does not increase the support size of any other linear function. A possible issue is that if we have to repeat this process for every function of large support then it seems that we need to take a fresh derivative for every such linear function. The point is that because we only care about linear functions that have a large support to begin with, we can find a variable that simultaneously appears in many of those functions and thus one derivative will eliminate many of the "bad" linear functions. Working out the parameters, we see that we need to take roughly $n^{\epsilon} \cdot \log |\Phi|$ many derivatives to reduce to the case where all linear functions have support size at most $n^{1-\epsilon}$. We can then "lift" the hitting set obtained for the derivative to a hitting set for the original polynomial, with a cost which is exponential in the number of the derivatives we took.

Now we go back to our first problem. We can assume that we have a depth-3 formula in which each linear function has support size at most $n^{1-\epsilon}$ and we wish to find a partition of the variables to sets $T_1, \ldots, T_m$ so that each T_i contains at most one variable from each linear function. This cannot be achieved for our choice of parameters, as shown by a simple probabilistic argument, so we relax our requirement and only demand that in each multiplication gate (of the formula) only a few linear functions have a large intersection. If at most k linear functions in each gate have a large intersection, we can expand each multiplication gate to at most n^k new gates (by simply taking the product of all linear functions that have large intersection) and then apply our argument. As we will be able to handle subexponential size formulas, this blow up is tolerable for us.

Note that if we were to pick the partition at random, when $m = n^{1-\epsilon+\gamma}$, for some small γ, then we will get that with a very high probability at most n^{δ} linear functions will have intersection at least n^{δ} with each T_i, where δ is such that $|\Phi| < \exp(n^{\delta})$. To get a deterministic version of this partitioning, we simply use an n^{δ}-wise independent family of hash functions $\{h : [n] \to [m]\}$. Each hash function h induces a partition of the variables to $T_i = \{x_k \mid h(k) = i\}$. Because of the high independence, we are guaranteed that there is at least one hash function that induces a good partition.

Now we have all the ingredients in place. To get our hitting set we basically do the following (we describe the construction as a process, but it should be clear that every step can be performed using some evaluation vectors).

1. Pick $n^{\epsilon} \cdot \log |\Phi|$ many variables and compute a black-box for the polynomial that is obtained by taking the derivative of f with respect to those variables. The cost of this step is roughly $\binom{n}{n^{\epsilon} \cdot \log |\Phi|} \cdot 2^{n^{\epsilon} \cdot \log |\Phi|}$, where the first term accounts for trying all possible choices of $n^{\epsilon} \cdot \log |\Phi|$ variables and the second is what we have to pay to get access to the

derived polynomial.

2. Partition the remaining variables to (roughly) $n^{1-\epsilon/2}$ many sets using a (roughly) $\log|\Phi|$-wise independent family of hash functions. The cost of this step is roughly $n^{\log|\Phi|}$ as this is the size of the hash function family.

3. Plug in a fresh copy of the hitting set of [FS13] and [AGKS15] to each of the sets of variables T_i. The cost is roughly $n^{\log n \cdot n^{1-\epsilon/2}}$.

Combining everything we get a hitting set of size roughly

$$\left(\binom{n}{n^\epsilon \cdot \log|\Phi|} \cdot 2^{n^\epsilon \cdot \log|\Phi|}\right) \cdot \left(n^{\log|\Phi|}\right) \cdot \left(n^{\log n \cdot n^{1-\epsilon/2}}\right) \approx 2^{\tilde{O}\left(n^{1-\epsilon/2}+n^\epsilon \log|\Phi|\right)}.$$

Optimizing the parameters we get our hitting set.

We would like to use the same approach also for the case of depth-4 formulas. Here the problem is that in the two bottom layers the formula computes a polynomial and not a linear function. In particular, when taking a derivative we are no longer removing functions that have a large support. Nevertheless, we can still use a similar idea. We show that there is a variable x_i such that by either setting $f|_{x_i=0}$ or considering $\frac{\partial f}{\partial x_i}$, we are guaranteed that the total sparsity of all polynomials that have large support goes down by some non-negligible factor. Thus, repeating this process (of either setting a variable to 0 or taking a derivative) $n^\epsilon \cdot \log|\Phi|$ many times we reach a depth-4 formula where all polynomials computed at the bottom addition gate have small support. Next, we partition the variables to sets and consider a single set T_i. Now, another issue is that even if the intersection of a low-support polynomial with some T_i is rather small, the sparsity of the resulting polynomial (which is considered as a polynomial in the variable in the intersection) can still be exponential in the size of the intersection. This is why we lose a bit in the upper bound compared to the depth-3 case. Combining all steps again we get the result for depth-4 formulas.

The proof for regular formulas works by first reducing to the depth-4 case and then applying our hitting set. The reduction is obtained in a similar spirit to the reduction of [KSS14]. We break the formula at an appropriate layer and then express the top layers as a $\Sigma\Pi$ circuit and the bottom layers as products of polynomials of not too high degree. We then use the trivial observation that if the degree of a polynomial is at most $n^{1-\epsilon}$ then its sparsity is at most $n^{n^{1-\epsilon}}$ and proceed as before. Due to the different requirements of the reduction and of the hashing part, we roughly lose a constant factor in the exponent of n, in the size of the hitting set, whenever the depth grows, resulting in a hitting set of size roughly $\exp(n^{1-1/\exp(d)})$.

To obtain the lower bounds we simply use the idea of [HS80] and [Agr05]. That is, given a hitting set we find a non-zero multilinear polynomial that vanishes on all points of the hitting set by solving a homogeneous system of linear equations.

1.5 Related Work

The work of [AGKS15]: The closest work to ours is the one by [AGKS15]. In addition to other results, they gave a white-box PIT algorithm that runs in time $n^{O(2^c n^{1-2/2^c} \log s)}$ for depth-3 formulas that can be represented as a sum of c set-multilinear formulas, each of size s (potentially with respect to different partitions of the variables).

Theorem 1.1 improves upon this results in several ways. First, the theorem gives a hitting set, i.e., a black-box PIT. Secondly, for $c = O(\log \log n)$ the running time of the algorithm of [AGKS15] is $\exp(n)$, whereas our construction can handle a sum of $\exp(n^\beta)$ set-multilinear formulas and still maintain a subexponential complexity (recall that also the improvement in [GKST15] requires exponential running time for $c = O(\log n)$).

However, there are some similarities behind the basic approach of this work and the work of Agrawal et al. Recall that a set-multilinear depth-3 formula is based on a partition of the variables, where each linear function in the formula contains variables from a single partition. Agrawal et al. start with a sum of c set-multilinear circuits, each with respect to a different partitioning of the variables, and their first goal is to reduce the formula to a set-multilinear formula, i.e., to have only one partition of the variables. For this they define a distance between different partitions and show, using an involved combinatorial argument, that one can find some partition $T_1, \ldots, T_m$ of the variables so that when restricting our attention to T_i, all the c set-multilinear formulas will be somewhat "close to each other". If the distance is Δ (according to their definition) then they prove that they can express the sum as a ROABP of size roughly $s \cdot n^\Delta$, where s is the total size of the depth-3 formula. Unlike our work, they find the partition in a white-box manner by gradually refining the given c partitions of the set-multilinear circuits composing the formula. The final verification step is done, in a similar manner to ours, by substituting the hitting set of [ASS13] (or that of [FS13]) to each of the sets T_i. The step of finding the partition $T_1, \ldots, T_m$ is technically involved and is the only step where white-box access is required.

Lower bounds for multilinear circuits and formulas: Lower bounds for the multilinear model were first proved by [NW96], who gave exponential lower bounds for depth-3 formulas. Raz first proved quasi-polynomial lower bounds for multilinear formulas computing the Determinant and Permanent polynomials [Raz09] and later gave a separation between multilinear NC_1 and multilinear NC_2 [Raz06]. [RY09] proved a lower bound of $\exp(n^{\Omega(\frac{1}{d})})$ for depth-d multilinear formulas. As in the general case, the depth reduction techniques of [VSBR83, AV08, Koi10, Tav13] also work for multilinear formulas. Thus, proving a lower bound of the form $\exp(n^{1/2+\epsilon})$ for $\Sigma\Pi\Sigma\Pi$ multilinear formulas, would imply a super-polynomial lower bound for multilinear circuits. Currently, the best lower bound for syntactic multilinear circuits is $n^{4/3}$ by Raz, Shpilka and Yehudayoff [RSY08].

Kayal, Saha and Saptharishi [KSS14] proved a quasi-polynomial lower bound for regular formulas that have the additional condition that the syntactic degree of the formula is at most twice the degree of the output polynomial.

1.6 Organization

In Section 2 we provide basic definitions and notations, and also prove some general lemmas which will be helpful in the next sections. In Section 3, we explain how to reduce general depth-3 and depth-4 formulas to formulas such that every polynomial at the bottom has small support. Then, in Section 4, we construct a hitting set for those types of formulas. In Section 5, we explain how to combine the ideas of the previous two sections and construct our hitting set for depth-3 and depth-4 multilinear formulas.

We then move on in Section 6 to depth-d regular formulas, and show how to reduce them to depth-4 formulas and obtain a hitting set for this class. In the short Section 7 we spell out briefly how, using known observations about the relation between PIT and lower bounds, we obtain our lower bounds for multilinear formulas. Finally, in Section 8 we discuss some open problems and future directions for research.

2 Preliminaries

In this section, we establish notation, some definitions and useful lemmas that will be used throughout the paper.

2.1 Notations and Basic Definitions

For any positive integer n, we denote by $[n]$ the set of integers from 1 to n, and by $\binom{[n]}{\leq r}$ the family of subsets $A \subseteq [n]$ such that $|A| \leq r$. We often associate a subset $A \subseteq [n]$ with a subset of variables $\mathsf{var}(A) \subseteq \{x_1, \ldots, x_n\}$ in a natural way (i.e., $\mathsf{var}(A) = \{x_i \mid i \in A\}$). In those cases we make no distinction between the two and use A to refer to $\mathsf{var}(A)$. Additionally, if A and B are disjoint subsets of $[n]$, we denote their disjoint union by $A \sqcup B$. For a vector $v \in \mathbb{F}^n$ we denote with $v|_A$ the restriction of v to the coordinates A.

In order to improve the readability of the text, we omit floor and ceiling notations.

Let $f(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$ be a polynomial. We will denote by $\partial_x f$ the formal derivative of f with respect to the variable x, and by $f|_{x=0}$ the polynomial obtained from f by setting $x = 0$. Moreover, if $A \subseteq [n]$, we will denote by $\partial_A f$ the polynomial obtained when taking the formal derivative of f with respect to all variables in A. In a similar fashion, we denote by $f|_{A=0}$ the polynomial obtained when we set all the variables in A to zero, and more generally, if $|A| = r$ and $\overline{\alpha} = (\alpha_1, \ldots, \alpha_r) \in \mathbb{F}^r$, $f|_{A=\overline{\alpha}}$ will denote the restriction of f obtained when setting the i'th variable in A to α_i, for $1 \leq i \leq r$.

In addition to the conventions above, the following definitions will be very useful in the next sections.

Definition 2.1 (Variable Set and Non-trivial Variable Set)**.** *Let $f(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$ be a polynomial. Define the variable set (var) and the non-trivial variable set (var^*) as follows:*

$$\mathsf{var}(f) = \{x \in \{x_1, \ldots, x_n\} \mid \partial_x f \neq 0\}$$
$$\mathsf{var}^*(f) = \{x \in \{x_1, \ldots, x_n\} \mid \partial_x f \neq 0 \text{ and } f|_{x=0} \neq 0\}.$$

That is, the variable set of a polynomial f is the set of variables $x \in \{x_1, \ldots, x_n\}$ that appear in the representation of f as a sum of monomials, whereas the non-trivial variable set is the set of variables of f that do not divide it.

We shall say that f has a small support if $\mathsf{var}(f)$ (or $\mathsf{var}^*(f)$) is not too large.

Definition 2.2 (Monomial Support and Sparsity of a polynomial)**.** *Let $f(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$ be a polynomial. We define the* monomial support *of f, written $\mathsf{mon}(f)$, as the set of monomials that have a non-zero coefficient in f. In addition, we define the sparsity of f, written $\|f\|$, as the size of the set $\mathsf{mon}(f)$, that is,*

$$\|f\| = |\mathsf{mon}(f)|.$$

In other words, the sparsity of f is the number of monomials that appear with a non-zero coefficient in f.

In the constructions of our hitting sets we will need to combine assignments to different subsets of variables. The following notation will be useful. For a partition of $[n]$, $T_1 \sqcup T_2 \sqcup \cdots \sqcup T_m = [n]$, and sets $\mathcal{H}_i \subseteq \mathbb{F}^{|T_i|}$, we denote with $\mathcal{H}_1^{T_1} \times \cdots \times \mathcal{H}_m^{T_m}$ the set of all vectors of length n whose restriction to T_i is an element of $\mathcal{H}_i$:

$$\mathcal{H}_1^{T_1} \times \cdots \times \mathcal{H}_m^{T_m} = \{v \in \mathbb{F}^n \mid \forall i \in [m],\ v|_{T_i} \in \mathcal{H}_i\}.$$

2.2 Depth-3 and Depth-4 Formulas

We define some special classes of depth-3 and depth-4 formulas that will be used throughout this paper.

Definition 2.3 (Restricted Top Fan-in)**.** *Let Φ be a multilinear depth-4 formula. We say that Φ is a multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula if it is of the form $\sum_{i=1}^{m}\prod_{j=1}^{t_i} f_{i,j}$, where $m \leq M$. If, in addition to the conditions above, we have that each $f_{i,j}$ is a linear function, that is, Φ is actually a depth-3 formula, we will say that Φ is a multilinear $\Sigma^{[M]}\Pi\Sigma$ formula.*

Our next definition considers the case where polynomials computed at the bottom layers do not contain too many variables, that is, they have small support.

Definition 2.4 (Restricted Top Fan-in and Variable Set)**.** *Let Φ be a multilinear depth-4 formula. We say that Φ is a multilinear $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula if it is of the form $\sum_{i=1}^{m}\prod_{j=1}^{t_i} f_{i,j}$, where $m \leq M$ and for each $1 \leq i \leq m$ we have that*

(i) $|\mathsf{var}(f_{i,j})| \leq \tau$ for all $1 \leq j \leq t_i$

(ii) $\mathsf{var}(f_{i,j_1}) \cap \mathsf{var}(f_{i,j_2}) = \emptyset$, for any $j_1 \neq j_2$.

If, in addition to the conditions above, we have that each $f_{i,j}$ is a linear function, that is, Φ is actually a depth-3 formula, we will say that Φ is a multilinear $\Sigma^{[M]}\Pi\Sigma^{\{\tau\}}$ formula.

Since the formula will be given to us as a black-box, we can make some assumptions about it, which will help us to preserve non-zeroness when taking derivatives or setting variables to zero. To this end, we define a notion of simplicity of depth-4 formulas,[4] and prove that we can assume without loss of generality that any input formula is simple.

Definition 2.5. *Let $f(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$ be a multilinear polynomial and let*

$$\Phi = \sum_{i=1}^{M} \prod_{j=1}^{t_i} f_{i,j}$$

be a multilinear depth-4 formula computing f. We say that Φ is a simple *multilinear depth-4 formula if for each variable $x \in \mathsf{var}(f)$ that divides f, it must be the case that for every $1 \leq i \leq M$, there exists $j \in [t_i]$ such that $f_{i,j} = x$.*

In words, Φ is simple if whenever a variable x divides f, it also divides every product gate. The following proposition tells us that we can indeed assume, without loss of generality, that any multilinear depth-4 formula given to us is a simple formula.

Proposition 2.6. *If Φ is a depth-4 multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula computing $f(x_1, \ldots, x_n)$, then f can be computed by a simple depth-4 multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula Ψ where $|\Psi| \leq |\Phi|$.*

Proof. Since Φ is a $\Sigma^{[M]}\Pi\Sigma\Pi$ formula, we have that

$$f = \sum_{i=1}^{M} \prod_{j=1}^{t_i} f_{i,j}.$$

Let $x \in \mathsf{var}(f)$ be such that $x \mid f$. Notice that we can write each $f_{i,j}$ in the following form:

$$f_{i,j} = x \cdot g_{i,j} + h_{i,j}, \quad \text{where } x \notin \mathsf{var}(g_{i,j}) \cup \mathsf{var}(h_{i,j}).$$

Moreover, observe that if $x \notin \mathsf{var}(f_{i,j})$, then we must have that $f_{i,j} = h_{i,j}$. Since the formula is multilinear, for each $i \in [M]$ there exists at most one j such that $x \in \mathsf{var}(f_{i,j})$. If such j exists, we might as well assume without the loss of generality that $j = 1$.

Let $A = \{i : 1 \leq i \leq M, \text{ and } x \in \mathsf{var}(f_{i,1})\}$ and $B = [M] \setminus A$. Now, rewriting the formula above for f, we get:

$$\begin{aligned} f = \sum_{i=1}^{M} \prod_{j=1}^{t_i} f_{i,j} &= \sum_{i \in A} f_{i,1} \cdot \prod_{j=2}^{t_i} f_{i,j} + \sum_{i \in B} \prod_{j=1}^{t_i} f_{i,j} \\ &= \sum_{i \in A} (x g_{i,1} + h_{i,1}) \cdot \prod_{j=2}^{t_i} h_{i,j} + \sum_{i \in B} \prod_{j=1}^{t_i} h_{i,j} \\ &= \sum_{i \in A} x g_{i,1} \cdot \prod_{j=2}^{t_i} h_{i,j} + \sum_{i \in A} h_{i,1} \cdot \prod_{j=2}^{t_i} h_{i,j} + \sum_{i \in B} \prod_{j=1}^{t_i} h_{i,j}. \end{aligned}$$

[4] Note that this is not the same notion as used, e.g., in [DS06].

Since $x \mid f$, it follows that $f = xg$. Hence, we must have that (in the above equation)

$$\sum_{i \in A} h_{i,1} \cdot \prod_{j=2}^{t_i} h_{i,j} + \sum_{i \in B} \prod_{j=1}^{t_i} h_{i,j} = 0$$

and therefore

$$f = \sum_{i \in A} x g_{i,1} \cdot \prod_{j=2}^{t_i} h_{i,j}.$$

Since $|A| \leq M$ and $\|g_{i,1}\| \leq \|f_{i,1}\|$, $\|h_{i,j}\| \leq \|f_{i,j}\|$ for every $i \in [M]$ and $2 \leq j \leq k_i$, the formula

$$\Phi' = \sum_{i \in A} x \cdot g_{i,1} \cdot \prod_{j=2}^{t_i} h_{i,j} = \sum_{i \in A} \prod_{j=2}^{t_i} x \cdot g_{i,1} \cdot h_{i,j}$$

is a multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula computing f, of size $|\Phi'| \leq |\Phi|$ and such that x appears as a polynomial at each product gate.

By repeating this process for each variable $\mathsf{var}(f) \setminus \mathsf{var}^*(f)$, we get our $\Sigma^{[M]}\Pi\Sigma\Pi$ formula Ψ. Since at each step we preserve the invariant that the size of the formula does not increase, we must have that $|\Psi| \leq |\Phi|$. □

As a corollary, together with the simple observation that any derivative or restriction of a multilinear formula results in a multilinear formula of at most the same size, we obtain that partial derivatives or restrictions of a multilinear polynomial can also be computed by simple formulas.

Corollary 2.7. *If Φ is a depth-4 multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula computing $f(x_1, \ldots, x_n)$, then for any two disjoint sets $A, B \subseteq \mathsf{var}(f)$, $\partial_A f|_{B=0}$ can be computed by a simple depth-4 multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula Ψ where $|\Psi| \leq |\Phi|$. We will refer to Ψ as $\partial_A \Phi|_{B=0}$.*

Therefore, from now on we will always assume that any depth-4 multilinear formula given to us is a simple formula.

2.3 ROABPs for Products of Sparse Polynomials

Another important model that we need for our constructions is that of Algebraic Branching Programs.

Definition 2.8 ([Nis91])**.** *An Algebraic Branching Program (ABP) is a directed acyclic graph with one vertex s of in-degree zero (the* source*) and one vertex t of out-degree zero (the* sink*). The vertices of the graph are partitioned into levels labeled $0, 1, \ldots, D$. Edges in the graph can only go from level $\ell - 1$ to level ℓ, for $\ell \in [D]$. The source is the only vertex at level 0 and the sink is the only vertex at level D. Each edge is labeled with an affine function in the input variables. The* width *of an ABP is the maximum number of nodes in any layer, and the* size *of an ABP is the number of vertices in the ABP.*

Each path from s to t computes the polynomial which is the product of the labels of the path edges, and the ABP computes the sum, over all $s \to t$ paths, of such polynomials.

Definition 2.9 (Ordered Read-Once Algebraic Branching Programs)**.** *An* Ordered Read-Once Algebraic Branching Program (ROABP) *in the variable set $\{x_1, \ldots, x_D\}$ is an ABP of depth D, such that each edge between layer $\ell - 1$ and ℓ is labeled by a univariate polynomial in x_ℓ.*

In this section we show an elementary construction of ROABPs for a very specific class of polynomials. This construction however will be useful in the upcoming sections.

Lemma 2.10. *Let $\mathbb{F}$ be a field, and $f(y_1, \ldots, y_m) = \sum_{i=1}^{M} \prod_{j=1}^{t_i} f_{i,j}$ be a multivariate polynomial over $\mathbb{F}$, such that for every $1 \leq i \leq M$:*

1. *At most k different $1 \leq j \leq t_i$, satisfy $|\mathsf{var}(f_{i,j})| > 1$.*
2. *For every $1 \leq j \leq t_i$, $\|f_{i,j}\| \leq s$.*

Then f can be computed by an ROABP of width at most $M \cdot s^k$.

Proof. Assume without the loss of generality that for every i there is $k_i \leq k$ such that $f_{i,1}, \ldots, f_{i,k_i}$ are those polynomials that contain more than a single variable. Note that the sparsity of every product $g_i \stackrel{\text{def}}{=} \prod_{j=1}^{k_i} f_{i,j}$ is at most s^k. We construct an ROABP of width s^k for each g_i. The final ROABP is constructed by connecting the ROABPs for all M products in parallel.

Fix $1 \leq i \leq M$. Expanding the product $g_i = \prod_{j=1}^{k_i} f_{i,j}$ we get at most s^k monomials. Now, multiply each such monomial with the remaining functions in the i'th gate, $\prod_{j=k_i+1}^{t_i} f_{i,j}$. Notice that the multiplicands in each such term can be reordered so that first x_1 appears then x_2 etc. Thus, we can construct a ROABP of width s^k for computing each such product of a monomial with $\prod_{j=k_i+1}^{t_i} f_{i,j}$. Then, connecting all those ROABPs in parallel we get a ROABP of width s^k for the i'th multiplication gate. Connecting in a similar fashion all ROABPs for the different multiplication gates we get a ROABP of width (at most) $M \cdot s^k$ computing f. □

Remark 2.11. *Since we only need to construct ROABPs of the above kind, which are in fact set multilinear formulas, we do not really need the full power of the ROABP model. In fact, we could have used any hitting set for the model of depth 3 set-multilinear formulas, and our results would remain true (albeit with no significant improvement in the parameters).*

2.4 Hashing

In this section we present the basic hashing tools that we will use in our construction. We first recall the notion of a k-wise independent hash family.

Definition 2.12. *A family of hash functions $\mathcal{F} = \{h : [n] \to [m]\}$ is k-wise independent if for any k distinct elements $(a_1, \ldots, a_k) \in [n]^k$ and any k (not necessarily distinct) elements $(b_1, \ldots, b_k) \in [m]^k$, we have:*

$$\Pr_{h \in \mathcal{F}} [h(a_1) = b_1 \wedge \cdots \wedge h(a_k) = b_k] = m^{-k}.$$

Our next lemma studies the case where several sets are hashed simultaneously by the same hash function. We present the proof in a general form and only later, in the application, fix the parameters.

Lemma 2.13. *Let $0 < \delta < \epsilon$, and $n, M \in \mathbb{N}$ such that $M = 2^{n^\delta}$. Assume $\mathcal{A}_1, \ldots, \mathcal{A}_M$ are families of pairwise disjoint subsets of $[n]$ such that for every $1 \leq i \leq M$ and every $A \in \mathcal{A}_i$, $|A| \leq n^{1-\epsilon}$. Let $\gamma > 0$ be such that $\gamma \geq (\epsilon - \delta)/2$. Let $\mathcal{F}$ be a family of k-wise independent hash functions from $[n]$ to $[m]$ for $k = n^\delta + 2\log n$ and $m = 10n^{1-\epsilon+\gamma}$.*

Then there exists $h \in \mathcal{F}$ such that for every $1 \leq i \leq M$ and every $1 \leq j \leq m$, both of the following conditions hold:

1. *For every set $A \in \mathcal{A}_i$, $\left|h^{-1}(j) \cap A\right| \leq k$.*
2. *The number of sets $A \in \mathcal{A}_i$ such that $\left|h^{-1}(j) \cap A\right| > 1$ is at most $k \log n$.*

Proof. We show that for a random $h \in \mathcal{F}$, both items happen with probability at least $2/3$.

Fix $1 \leq i \leq M$, $1 \leq j \leq m$ and $A \in \mathcal{A}_i$. By k-wise independence and the assumption that $|A| \leq n^{1-\epsilon}$, we have that

$$\begin{aligned}
\Pr\left[\left|h^{-1}(j) \cap A\right| \geq k\right] &\leq \sum_{B \subseteq A, |B|=k} \Pr[\forall b \in B, h(b) = j] \\
&\leq \binom{n^{1-\epsilon}}{k} \cdot \frac{1}{\left(10n^{1-\epsilon+\gamma}\right)^k} \\
&\leq n^{(1-\epsilon)k} \cdot \frac{1}{\left(n^{(1-\epsilon)k+\gamma k}\right)} \cdot \frac{1}{10^k} \\
&\leq n^{-\gamma k} \cdot 10^{-k}. \qquad (1)
\end{aligned}$$

Taking a union bound over all $1 \leq i \leq M$ and $1 \leq j \leq m$, and using the estimate (1) and the fact that $m \leq n$, we get that the first item in the statement of the lemma does not happen with probability at most

$$M \cdot m \cdot n^{-\gamma k} \cdot 10^{-k} \leq \frac{1}{3}$$

for large enough n, by the choice of k.

Turning to the second item in the statement of the lemma, it is convenient to partition every family of subsets $\mathcal{A}$ into $(1-\epsilon)\log n$ disjoint buckets, according to the size of the sets in $\mathcal{A}$. Fix such $\mathcal{A}$, and, for $1 \leq i \leq (1-\epsilon)\log n$, define the bucket

$$\mathcal{B}_i = \left\{ A \in \mathcal{A} : \frac{n^{1-\epsilon}}{2^i} \leq |A| \leq \frac{n^{1-\epsilon}}{2^{i-1}} \right\}.$$

We show that with high probability over the choice of h, and for every $j \in [m]$, in every bucket there are at most k sets whose intersection with $h^{-1}(j)$ has size larger than 1.

For every set $A \in \mathcal{A}$, by k-wise independence (in particular, pairwise independence) the probability that $|A \cap h^{-1}(j)| \geq 2$ is at most

$$\binom{|A|}{2} \cdot \frac{1}{(10n^{1-\epsilon+\gamma})^2} \leq |A|^2 \cdot \frac{1}{100n^{2-2\epsilon+2\gamma}}.$$

Fix a bucket $\mathcal{B}_i$. By definition, for every $A \in \mathcal{B}_i$ it holds that $|A| \leq \frac{n^{1-\epsilon}}{2^{i-1}}$, and so for every set $A \in \mathcal{B}_i$, the probability that $|A \cap h^{-1}(j)| \geq 2$ is at most

$$\frac{n^{2-2\epsilon}}{2^{2i-2}} \cdot \frac{1}{100} \cdot n^{2\epsilon-2-2\gamma} = \frac{1}{100} \cdot n^{-2\gamma} \cdot 2^{2-2i}. \tag{2}$$

Since $\mathcal{A}$ is a partition, by pairwise disjointness, the number of sets in $\mathcal{B}_i$ is at most $n^\epsilon \cdot 2^i$. Hence, by k-wise independence and (2), the probability there exist $k/2$ sets in the bucket $\mathcal{B}_i$, with intersection sizes at least 2, is at most

$$\begin{aligned}\binom{n^\epsilon \cdot 2^i}{k/2} \cdot \left(\frac{1}{100} \cdot n^{-2\gamma} \cdot 2^{2-2i}\right)^{k/2} \leq \\ \left(\frac{en^\epsilon \cdot 2^i}{k/2}\right)^{k/2} \left(\frac{1}{100} \cdot n^{-2\gamma} \cdot 2^{2-2i}\right)^{k/2} = \\ \left(\frac{e \cdot 2^{3-i}}{100}\right)^{k/2} \cdot \left(\frac{n^{\epsilon-2\gamma}}{k}\right)^{k/2},\end{aligned} \tag{3}$$

where we have used the inequality $\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b$. Observe that $n^{\epsilon-2\gamma} \leq k$, by the choice of γ.

Taking a union bound over all $\log n$ buckets, and then over all M partitions and all m possible values of j, and using the estimation (3), we get that the probability that there more than $k/2$ sets that intersect $h^{-1}(j)$ in more than one element, for some j, is at most

$$M \cdot m \cdot \log n \cdot \left(\frac{e \cdot 2^{3-i}}{100}\right)^{k/2} \cdot \left(\frac{n^{\epsilon-2\gamma}}{k}\right)^{k/2} \leq \frac{1}{3},$$

for large enough n, by the choices of k and γ. Hence, the second item in the statement of the lemma follows as well. □

We conclude this section with the following well known fact (see, e.g., Chapter 16 in [AS08], and the references therein):

Fact 2.14. *There exists an explicitly constructible family $\mathcal{F}$ of k-wise independent hash functions from $[n]$ to $[10n^{1-\epsilon+\gamma}]$ of size $|\mathcal{F}| = n^{O(k)}$.*

3 Reducing Bottom Support of Depth-3 and Depth-4 Formulas

In this section we make the first step towards proving Theorem 1.1 and Theorem 1.3. As outlined in Section 1.4, our first step is making the functions computed at the bottom layers

(linear functions in the case of depth-3 and "sparse" polynomials in the case of depth-4) have small (variable) support. Hence, we establish reductions from any $\Sigma^{[M]}\Pi\Sigma$ or $\Sigma^{[M]}\Pi\Sigma\Pi$ formula to a $\Sigma^{[M]}\Pi\Sigma^{\{\tau\}}$ or $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula, respectively. We first describe the simple depth-3 case. We continue by elaborating on the more general case of depth-4 formulas, which is slightly more involved. Both proofs follow the outline described in Section 1.4.

3.1 Reducing Bottom Support for Depth-3

Given a depth 3 formula $\sum_{i=1}^{M}\prod_{j=1}^{t_i}\ell_{i,j}$, we would like to eliminate all linear functions that contain many variables. To this end, we observe that there must exist a variable that appears in many of these functions, and that taking a derivative according to that variable eliminates those functions from the formula.

Lemma 3.1. *Let $f(x_1,\ldots,x_n)=\sum_{i=1}^{M}\prod_{j=1}^{t_i}\ell_{i,j}$ be a non-zero, multilinear polynomial computed by a multilinear $\Sigma^{[M]}\Pi\Sigma$ formula Φ and let $\epsilon>0$. Then, there exists a set of variables A of size $|A|\leq\tilde{O}(n^{\epsilon}\cdot\log M)$ such that $\partial_A f$ is a non-zero multilinear polynomial that can be computed by a multilinear $\Sigma^{[M]}\Pi\Sigma^{\{n^{1-\epsilon}\}}$ formula.*

Proof. Define

$$\mathcal{B}=\{\ell_{i,j}\mid|\mathsf{var}(\ell_{i,j})\cap\mathsf{var}(f)|\geq n^{1-\epsilon}\}$$

to be the set of "bad" linear functions. Those are linear functions that contain more than $n^{1-\epsilon}$ variables that also appear in f. We show how to eliminate those linear functions from the formula while preserving non-zeroness.

Since for every $\ell\in\mathcal{B}$, $|\mathsf{var}(\ell)\cap\mathsf{var}(f)|\geq n^{1-\epsilon}$, there exists a variable x_i that appears in at least $|\mathcal{B}|n^{1-\epsilon}/n=|\mathcal{B}|/n^{\epsilon}$ linear functions in $\mathcal{B}$ (and also in f). Consider the polynomial $\partial_{x_i}f$. Since $x_i\in\mathsf{var}(f)$, this is a non-zero polynomial. Furthermore, using the fact that deriving with respect to a variable is a linear operation, and the fact that every multiplication gate in the formula multiplies linear functions with disjoint support, a formula for $\partial_{x_i}f$ can be obtained from Φ by replacing every linear function in which x_i appears with an appropriate constant. Therefore, every such function is removed from $\mathcal{B}$, and so the set of bad linear functions in $\partial_{x_i}f$ is of size at most $|\mathcal{B}|-|\mathcal{B}|/n^{\epsilon}=|\mathcal{B}|\cdot(1-1/n^{\epsilon})$. We continue this process for at most $O(n^{\epsilon}\cdot\log|\mathcal{B}|)$ steps, until we reach a point where $|\mathcal{B}|<1$ and so $|\mathcal{B}|=0$.

Finally, it remains to be noted that in the original formula, the number $|\mathcal{B}|$ of bad linear functions is at most Mn, since by multilinearity each multiplication gate multiplies linear functions with disjoint support, and so its fan-in is at most n. □

3.2 Reducing Bottom Support for Depth-4

The process for depth-4 formulas follows the same outline as the depth-3 case, but there are a few more details. Given a parameter $\tau\in\mathbb{N}$, we want to transform any multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula computing a non-zero polynomial $f(x_1,\ldots,x_n)$ into a $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula, while preserving non-zeroness. By Proposition 2.6, we can assume any formula that we work with is a *simple formula*. We again define the "bad" polynomials as those that contain many variables

(that also appear in f). Our progress measure for their elimination will be the *total sparsity* of all bad polynomials, which we define below.

Definition 3.2. *Let $\tau \in \mathbb{N}$ and $\Phi = \sum_{i=1}^{M} \prod_{j=1}^{t_i} f_{i,j}$ be a multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula computing a non-zero multilinear polynomial $f(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$. Let $\mathcal{B} = \{f_{i,j} \mid \mathsf{var}(f_{i,j}) > \tau\}$. We say that Φ is Δ-far from a $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula if*

$$\sum_{g \in \mathcal{B}} \|g\| = \Delta.$$

A polynomial $f(x_1, \ldots, x_n)$ is Δ-far from a $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula if it can be computed by a formula that is Δ-far from such a formula.

Observation 3.3. *Notice that a formula Φ is 0-far from being $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ iff Φ itself is a $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula.*

Now that we have a measure of how far a given $\Sigma^{[M]}\Pi\Sigma\Pi$ formula (computing a non-zero polynomial) is from the class of $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formulas, we can show the existence of a small set of variables such that when we either take derivatives or set these variables to zero, we obtain a $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula computing another non-zero polynomial. Since we are working with *simple* formulas, if a variable x appears in a bad polynomial $f_{i,j} \in \mathcal{B}$, then it must be the case that $x \in \mathsf{var}^*(f)$, and therefore we are free to either take a partial derivative with respect to x or to set x to zero, while preserving non-zeroness of the input polynomial $f(x_1, \ldots, x_n)$. Therefore, the non-zeroness condition is taken care of by simplicity.

We begin by showing that we can always make good progress in this measure. More precisely, we have the following lemma:

Lemma 3.4. *Let $\Phi = \sum_{i=1}^{M} \prod_{j=1}^{t_i} f_{i,j}$ be a multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula computing a non-zero multilinear polynomial $f(x_1, \ldots, x_n)$*
$\in \mathbb{F}[x_1, \ldots, x_n]$. If Φ is Δ-far from a $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula, then there exists $x \in \mathsf{var}^*(f)$ such that one of the polynomials $\partial_x f$ or $f|_{x=0}$ is non-zero and is at most $\Delta(1 - \frac{\tau}{2n})$-far from a $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula.

Proof. By Proposition 2.6, we can assume without loss of generality that Φ is simple. We note that making Φ simple can only reduce Δ.

Let $\mathcal{B} = \{f_{i,j} \mid |\mathsf{var}(f_{i,j})| > \tau\}$. Notice that by simplicity of Φ, we have that $|\mathsf{var}(f_{i,j})| > 1 \Rightarrow \mathsf{var}(f_{i,j}) \subseteq \mathsf{var}^*(f)$. Since Φ is Δ-far from a $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula, we have that

$$\Delta = \sum_{g \in \mathcal{B}} \|g\|.$$

For each $x \in \mathsf{var}^*(f)$, let $F_x = \{g \in \mathcal{B} \mid x \in \mathsf{var}(g)\}$. Notice that

$$\sum_{x \in \mathsf{var}^*(f)} \left(\sum_{g \in F_x} \|g\| \right) = \sum_{g \in \mathcal{B}} |\mathsf{var}(g)| \cdot \|g\| > \tau \cdot \sum_{g \in \mathcal{B}} \|g\| = \tau\Delta.$$

This implies that there exists $x \in \mathsf{var}^*(f)$ for which

$$\sum_{g \in F_x} \|g\| \geq \frac{\Delta \cdot \tau}{|\mathsf{var}^*(f)|} \geq \frac{\Delta \cdot \tau}{n}. \tag{4}$$

Since $\|g\| = \|g|_{x=0}\| + \|\partial_x g\|$ for any multilinear polynomial g, we have that

$$\sum_{g \in F_x} \|g\| = \sum_{g \in F_x} \|g|_{x=0}\| + \sum_{g \in F_x} \|\partial_x g\|.$$

Hence, by equation (4), one of $\sum_{g \in F_x} \|g|_{x=0}\|$ or $\sum_{g \in F_x} \|\partial_x g\|$ must be larger than $\frac{\Delta \cdot \tau}{2n}$. If

$$\sum_{g \in F_x} \|g|_{x=0}\| > \frac{\Delta \tau}{2n},$$

by taking the derivative of f with respect to x, we have that $\partial_x f \neq 0$ (since $x \in \mathsf{var}^*(f)$) and that the distance of $\partial_x \Phi$ to a $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula is upper bounded by

$$\begin{aligned}\sum_{g \in \mathcal{B} \setminus F_x} \|g\| + \sum_{g \in F_x} \|\partial_x g\| &= \sum_{g \in \mathcal{B}} \|g\| - \sum_{g \in F_x} \|g|_{x=0}\| \\ &< \Delta - \frac{\Delta \tau}{2n} = \Delta \left(1 - \frac{\tau}{2n}\right).\end{aligned}$$

Indeed, notice that in $\partial_x \Phi$ the support of any other polynomial $f_{i,j}$ cannot increase (so that the subset of polynomials with large support in $\partial_x \Phi$ is a subset of $\mathcal{B}$).

Analogously, if $\sum_{C \in F_x} \|\partial_x C\| > \frac{\Delta \tau}{2n}$, then we take $f|_{x=0}$. The upper bounds are the same as those obtained for the first case. This finishes the proof of the lemma. □

By repeatedly applying Lemma 3.4, we obtain the following corollary, which guarantees the existence of a small set of variables that allow us to transform our $\Sigma^{[M]}\Pi\Sigma\Pi$ formula into a $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ one.

Corollary 3.5 (Reduction to Depth-4 with Small Bottom Support)**.** *Let Φ be a multilinear simple $\Sigma^{[M]}\Pi\Sigma\Pi$ formula computing a non-zero multilinear polynomial $f(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$. There exist disjoint sets $A, B \subset [n]$ with $|A \sqcup B| \leq \frac{2n}{\tau} \cdot \log(|\Phi|)$ such that the polynomial $\partial_A f|_{B=0}$ is non-zero and can be computed by a simple multilinear $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ formula Φ.*

Proof. Let Δ be such that Φ is Δ-far from being $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$. Notice that $\Delta \leq |\Phi|$.

We show by induction that there exist disjoint sets A_k and B_k such that $|A_k \sqcup B_k| \leq k$, and the polynomial $\partial_{A_k} f|_{B_k=0}$ is non-zero and at most $\Delta(1 - \frac{\tau}{2n})^k$-far from being $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$.

For $k \geq 0$, define $A_k, B_k \subseteq [n]$, $f_k(x_1, \ldots, x_n) = \partial_{A_k} f|_{B_k=0}$ and let Δ_k be an upper bound on how far $f_k(x_1, \ldots, x_n)$ is from being $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$. Initially, set $A_0 = B_0 = \emptyset$. In this case, we have that $f_0(x_1, \ldots, x_n) = f(x_1, \ldots, x_n)$ and $\Delta_0 = \Delta = \Delta(1 - \frac{\tau}{2n})^0$. This is the base case for our induction.

Suppose our hypothesis is true for some $k \geq 0$. If $\Delta(1-\frac{\tau}{2n})^k < 1$, then we know that our formula is already $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$ and therefore we are done. Else, by applying Lemma 3.4, we have that there is a variable $x \in \mathsf{var}^*(f_k)$ such that either $\partial_x f_k$ or $f_k|_{x=0}$ is (at most) $\Delta_k(1-\frac{\tau}{2n})$-far from being $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{\tau\}}$. Thus, $\Delta_k(1-\frac{\tau}{2n}) \leq \Delta(1-\frac{\tau}{2n})^k \cdot (1-\frac{\tau}{2n}) = \Delta(1-\frac{\tau}{2n})^{k+1}$ and $x \in \mathsf{var}^*(f_k) \subseteq [n] \setminus (A_k \sqcup B_k)$. Therefore, if $\partial_x f_k$ is the close polynomial then we set $A_{k+1} = A_k \cup \{x\}, B_{k+1} = B_k$. Otherwise, we set $A_{k+1} = A_k, B_{k+1} = B_k \cup \{x\}$. It is easy to see that the inductive properties hold in this case as well. This ends the inductive proof.

Since $\Delta(1-\frac{\tau}{2n})^k < 1$ for $k \geq \frac{2n}{\tau}\log\Delta$, and since $\frac{2n}{\tau}\log(|\Phi|) \geq \frac{2n}{\tau}\log\Delta$, it is enough to choose at most $\frac{2n}{\tau}\log(|\Phi|)$ variables. This proves this corollary. □

4 Hitting Set for $\Sigma\Pi\Sigma^{\{n^{1-\epsilon}\}}$ and $\Sigma\Pi(\Sigma\Pi)^{\{n^{1-\epsilon}\}}$ Formulas

In this section we construct subexponential sized hitting set for the classes of $\Sigma^{[M]}\Pi\Sigma^{\{n^{1-\epsilon}\}}$ and $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{n^{1-\epsilon}\}}$ multilinear formulas. Recall that in Section 3 we showed how to reduce general depth-3 and depth-4 formulas to these types of formulas. In the next section, we will show how to tie all loose ends and combine the arguments of Section 3 with those of this section in order to handle the general case.

An essential ingredient in our construction is a quasi-polynomial sized hitting set for Read-Once Algebraic Branching Programs (ROABPs) [FS13, AGKS15]. We note that in our setting, we may assume that the reading order of the variables by the ABP is known.

Theorem 4.1 ([FS13, AGKS15])**.** *Let $\mathcal{C}$ be the class of n-variate polynomials computed by a ROABP of width w, such that the degree of each variable is at most d, over a field $\mathbb{F}$ so that $|\mathbb{F}| \geq \mathsf{poly}(n,w,d)$. Then $\mathcal{C}$ has a hitting set of size $\mathsf{poly}(n,w,d)^{\log n}$ that can be constructed in time $\mathsf{poly}(n,w,d)^{\log n}$.*

We begin by describing a unified construction for both $\Sigma^{[M]}\Pi\Sigma^{\{n^{1-\epsilon}\}}$ and $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{n^{1-\epsilon}\}}$ formulas. We then describe how to set the parameters of the construction for each of the cases.

Construction 4.2 (Hitting set for multilinear $\Sigma^{[M]}\Pi\Sigma^{\{n^{1-\epsilon}\}}$ and $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{n^{1-\epsilon}\}}$ formulas)**.** *Let $0 < \delta < \epsilon$ and n, k, s, M integers, such that $M = 2^{n^\delta}$ and $k = n^\delta + 2\log n$. Set $m = 10n^{1-(\epsilon+\delta)/2}$ and $t = k\log n$. Let $\mathcal{F}$ be a family of k-wise independent hash functions from $[n]$ to $[m]$, as in Lemma 2.13. For every $h \in \mathcal{F}$, define the set I_h as follows:*

1. *Partition the variables to sets[5] $T_1 \sqcup T_2 \sqcup \cdots \sqcup T_m = h^{-1}(1) \sqcup h^{-1}(2) \sqcup \cdots \sqcup h^{-1}(m)$.*
2. *For every $1 \leq i \leq m$, let $\mathcal{H}_i$ be a hitting set for ROABPs of width $M \cdot s^t$ and individual degree $d = 1$ (as promised by Theorem 4.1), on the variables of T_i (of course, $|T_i| \leq n$).*
3. *We define I_h as the set of all vectors v such that the restriction of v to the coordinates T_i, $v|_{T_i}$, is in $\mathcal{H}_i$. I.e., in the notation of Section 2.1,*

$$I_h = \mathcal{H}_1^{T_1} \times \mathcal{H}_2^{T_2} \times \cdots \times \mathcal{H}_m^{T_m}.$$

[5] Recall that we associate subsets of $[n]$ with subsets of the variables, and make no distinction in the notation.

Finally, define $\mathcal{H} = \bigcup_{h \in \mathcal{F}} I_h$.

The following lemma gives an upper bound to the size of the hitting set constructed in Construction 4.2.

Lemma 4.3. *Let* $\delta, \epsilon, k, n, s$ *and* M *be the parameters of Construction 4.2. The set* $\mathcal{H}$ *constructed in Construction 4.2 has size* $n^{O(k)} \cdot \left(M \cdot s^{k \log n}\right)^{\tilde{O}(n^{1-(\epsilon+\delta)/2})} = \left(M \cdot s^{k \log n}\right)^{\tilde{O}(n^{1-(\epsilon+\delta)/2})}$, *and it can be constructed in time* $\mathsf{poly}(|\mathcal{H}|)$.

Proof. This is a direct consequence of the construction, 2.14 and Theorem 4.1. □

4.1 Depth-3 Formulas

We begin by describing the argument for depth-3 formulas. The following lemma proves that indeed, by setting the proper parameters, the set $\mathcal{H}$ from Construction 4.2 does hit $\Sigma^{[M]}\Pi\Sigma^{\{n^{1-\epsilon}\}}$ formulas.

Lemma 4.4. *Let* $f(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$ *be a non-zero multilinear polynomial computed by a multilinear* $\Sigma^{[M]}\Pi\Sigma^{\{n^{1-\epsilon}\}}$ *formula* $\Phi = \sum_{i=1}^{M} \prod_{j=1}^{t_i} \ell_{i,j}$. *Let* $\mathcal{H}$ *be the set constructed in Construction 4.2 with* $s = k + 1$. *Then there exists a point* $\overline{\alpha} \in \mathcal{H}$ *such that* $f(\overline{\alpha}) \neq 0$.

Proof. For every multiplication gate $1 \leq i \leq M$ in Φ, define a partition of the variables

$$\mathcal{A}_i = \{\mathsf{var}(\ell_{i,j}) \cap \mathsf{var}(f) \mid 1 \leq j \leq t_i\}.$$

Let $h \in \mathcal{F}$ be the function guaranteed by Lemma 2.13 with respect to the partitions $\mathcal{A}_1, \ldots \mathcal{A}_M$, and assume the setup of Construction 4.2. We claim that there exists $\overline{\alpha} \in I_h$ such that $f(\overline{\alpha}) \neq 0$.

To that end, consider the partition of the variables induced by h:

$$T_1 \sqcup T_2 \sqcup \cdots \sqcup T_m = h^{-1}(1) \sqcup h^{-1}(2) \sqcup \cdots \sqcup h^{-1}(m).$$

We view the polynomial as a polynomial f_1 in the variables of T_1, over the extension field $\mathbb{F}(T_2 \sqcup \cdots \sqcup T_n)$. We claim that f_1 can be computed by an ROABP of width $M \cdot (k+1)^{k \log n}$. To see this note that, by Lemma 2.13, in any multiplication gate, at most $k \log n$ linear functions contain more than one variable from T_1, and each contains at most k variables. It follows that the sparsity of every linear function (with respect to the variables in T_1) among those $k \log n$ functions, is at most $k + 1$. By Lemma 2.10, f_1 can be computed by an ROABP over $\mathbb{F}(T_2 \sqcup \cdots \sqcup T_n)$ of width $M \cdot (k+1)^{k \log n}$. By the hitting set property of Theorem 4.1, there exists $\overline{\alpha_1} \in \mathcal{H}_1 \subseteq \mathbb{F}^{|T_1|}$ such that $f_2 \stackrel{\text{def}}{=} f_1|_{T_1 = \overline{\alpha_1}} \not\equiv 0$.

Similarly, the same conditions now hold for f_2, considered as a polynomial over the field $\mathbb{F}(T_3 \sqcup \cdots \sqcup T_n)$, and so there exists $\overline{\alpha_2} \in \mathcal{H}_2 \subseteq \mathbb{F}^{|T_2|}$ such that $f_3 \stackrel{\text{def}}{=} f_2|_{T_2 = \overline{\alpha_2}} \not\equiv 0$.

We continue this process for m steps, and at the last step we find $\overline{\alpha_m}$ such that $f_{m-1}(\overline{\alpha_m}) = f(\overline{\alpha_1}, \cdots, \overline{\alpha_m}) \neq 0$, with $(\overline{\alpha_1}, \cdots, \overline{\alpha_m}) \in \mathbb{F}^n$ being the length n vector obtained by filling the entires of $\overline{\alpha_i} \in \mathbb{F}^{|T_i|}$ in the positions indexed by T_i. □

4.2 Depth-4 Formulas

The argument for depth-4 formulas is very similar, apart from a small change in the setting of the parameters.

Lemma 4.5. *Let $f(x_1,\ldots,x_n) \in \mathbb{F}[x_1,\ldots,x_n]$ be a non-zero multilinear polynomial computed by a multilinear $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{n^{1-\epsilon}\}}$ formula $\Phi = \sum_{i=1}^{M}\prod_{j=1}^{t_i} f_{i,j}$. Let $\mathcal{H}$ be the set constructed in Construction 4.2 with $s = 2^k$. Then, there exists a point $\overline{\alpha} \in \mathcal{H}$ such that $f(\overline{\alpha}) \neq 0$.*

Proof. The proof is almost identical to that of Lemma 4.4. In this case, for every $1 \leq i \leq M$ we define the partition

$$\mathcal{A}_i = \{\mathsf{var}(f_{i,j}) \mid 1 \leq j \leq t_i\}.$$

Note that the assumptions of Lemma 2.13 still hold, and so we denote by $h \in \mathcal{F}$ the function guaranteed by Lemma 2.13 with respect to the partitions $\mathcal{A}_1,\ldots\mathcal{A}_M$, and again claim that there exists $\overline{\alpha} \in I_h$ such that $f(\overline{\alpha}) \neq 0$.

Consider once more the partition on the variables induced by h,

$$T_1 \sqcup T_2 \sqcup \cdots \sqcup T_m = h^{-1}(1) \sqcup h^{-1}(2) \sqcup \cdots \sqcup h^{-1}(m),$$

and view the polynomial as a polynomial f_1 in the variables of T_1, over the field $\mathbb{F}(T_2 \sqcup \cdots \sqcup T_n)$.

We now claim that f_1 can be computed by an ROABP of width $M \cdot \left(2^k\right)^{k\log n} = 2^{k^2\log n}$. The proof for this claim is exactly as in the depth-3 case, except that now the best bound we can give on the sparsity of each polynomial which intersects T_1 in more than one variable is 2^k, as it is a multilinear polynomials in at most k variables.

Similarly, we move on to handle $T_2,\ldots,T_m$ and obtain a point $\overline{\alpha}$ such that $f(\overline{\alpha}) \neq 0$. □

5 Hitting Set for Depth-3 and Depth-4 Multilinear Formulas

Recall that, in Section 3, we showed that any non-zero $\Sigma^{[M]}\Pi\Sigma$ or $\Sigma^{[M]}\Pi\Sigma\Pi$ formula has a non-zero partial derivative (and, possibly, a restriction) which is computed by a non-zero $\Sigma^{[M]}\Pi\Sigma^{\{n^{1-\epsilon}\}}$ or $\Sigma^{[M]}\Pi(\Sigma\Pi)^{\{n^{1-\epsilon}\}}$ formula, respectively. Then, in Section 4 we gave hitting sets for such formulas. In this section we provide the final ingredient, which is to show how to "lift" those hitting sets back to the general class, via a simple method, albeit one that requires some notation.

Handling restrictions is fairly easy, and causes no blow up in the hitting set size: If we have a set $\mathcal{H} \subseteq \mathbb{F}^{n-r}$ that hits $f|_{B=0}$ for some multilinear polynomial $f(x_1,\ldots,x_n)$ and $B \subseteq [n]$ with $|B| = r$, then simply extending $\mathcal{H}$ into a subset of $\mathbb{F}^n$ by filling out zeros in all the entries indexed by B will hit f itself.

In order to handle partial derivates, first note that if $f(x_1,\ldots,x_n)$ is a multilinear polynomial, then

$$\begin{aligned}\partial_{x_i} f = {} & f(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n) \\ & - f(x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_n),\end{aligned}$$

and so if $\partial_{x_i} f(\overline{\alpha}) \neq 0$ for some $\overline{\alpha} \in \mathbb{F}^n$ then at least one of the two evaluations on the right hand side must be non-zero as well.

Applying this fact repeatedly, given a set $A \subseteq [n]$ we can evaluate $\partial_A f$ at any point by making $2^{|A|}$ evaluations of f. Motivated by this, we introduce the following notation:

Definition 5.1. *Let $f(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$ be a multilinear polynomial and $A, B \subseteq [n]$ be two disjoint subsets of variables with $|A| = r, |B| = r'$. Let $\mathcal{H} \subseteq \mathbb{F}^{n-(r+r')}$.*

We define the "lift" of $\mathcal{H}$ with respect to (A, B) to be

$$\mathcal{L}_A^B(\mathcal{H}) = \left(\{0,1\}^r\right)^A \times \left(\{0\}^{r'}\right)^B \times \mathcal{H}^{[n]\setminus(A \sqcup B)}.$$

In the special case where $B = \emptyset$, we simply denote $\mathcal{L}_A^B(\mathcal{H}) = \mathcal{L}_A(\mathcal{H})$.

That is, for all $\overline{\alpha} \in \mathcal{H}$, $\mathcal{L}_A^B(\mathcal{H})$ contains all the possible 2^r ways to extend $\overline{\alpha}$ into $\overline{\beta} \in \mathbb{F}^n$ by filling out zeros and ones within the r entries that are indexed by A, and zeros in all the r' entries indexed by B.

5.1 Depth-3 Formulas

In this section we prove Theorem 1.1. For the reader's convenience, we first restate the theorem:

Theorem 5.2 (Theorem 1.1, restated)**.** *Let $\mathcal{C}$ be the class of multilinear $\Sigma^{[M]}\Pi\Sigma$ formulas for $M = 2^{n^\delta}$. There exists a hitting set $\mathcal{H}$ of size $|\mathcal{H}| = 2^{\tilde{O}(n^{2/3+2\delta/3})}$ for $\mathcal{C}$, that can be constructed in time $\mathsf{poly}(|\mathcal{H}|)$.*

The size of the hitting set is subexponential for any constant $\delta < 1/2$. Also, if $M = \mathsf{poly}(n)$ then the size of the hitting set is $2^{\tilde{O}(n^{2/3})}$.

With Definition 5.1 in hand, we now provide our construction for $\Sigma^{[M]}\Pi\Sigma$ formulas, towards the proof of Theorem 5.2.

Construction 5.3 (Hitting set for multilinear $\Sigma^{[M]}\Pi\Sigma$ formulas)**.** *Let $M = 2^{n^\delta}$ and $\epsilon = 2/3 - \delta/3$. Let $r = \tilde{O}(n^\epsilon \log M) = \tilde{O}(n^{2/3+2\delta/3})$ be an upper bound on the size of the set guaranteed to exist by Lemma 3.1. For every $A \in \binom{[n]}{\leq r}$, construct a set $\mathcal{H}_A \in \mathbb{F}^{n-|A|}$ using Construction 4.2 with parameters $\delta, \epsilon, n, k, s = k+1$ and M (recall that in Construction 4.2 we set $k = n^\delta + 2\log n$). Finally, let*

$$\mathcal{H} = \bigcup_{A \in \binom{[n]}{\leq r}} \mathcal{L}_A(\mathcal{H}_A).$$

We are now ready to prove Theorem 5.2:

Proof of Theorem 5.2. We show that the set $\mathcal{H}$ constructed in Construction 5.3 has the desired properties. First, note that by Lemma 4.3, for every $A \subseteq [n]$ with

$$|A| \leq \tilde{O}(n^\epsilon \log M) = \tilde{O}(n^{2/3-\delta/3} \log M) = \tilde{O}(n^{2/3+2\delta/3}),$$

the set $\mathcal{H}_A$ has size

$$(M \cdot (k+1)^{k \log n})^{\tilde{O}(n^{2/3-\delta/3})} = 2^{\tilde{O}(n^{2/3+2\delta/3})},$$

where we have used the fact that, in Construction 5.3, we take $k = n^\delta + 2\log n$. It therefore follows that

$$|\mathcal{L}_A(\mathcal{H}_A)| \leq 2^{|A|} \cdot |\mathcal{H}_A| = 2^{\tilde{O}(n^{2/3+2\delta/3})},$$

and that

$$|\mathcal{H}| \leq \sum_{i=0}^{\tilde{O}(n^{2/3+2\delta/3})} \sum_{A \subseteq [n], |A|=i} |\mathcal{L}_A(\mathcal{H}_A)| = 2^{\tilde{O}(n^{2/3+2\delta/3})}.$$

To show the hitting property of $\mathcal{H}$, let $f(x_1, \ldots, x_n)$ be a non-zero multilinear polynomial computed by a $\Sigma^{[M]}\Pi\Sigma$ formula, and let $A' \subseteq [n]$ be the set guaranteed by Lemma 3.1. Thus, $|A'| \leq \tilde{O}(n^\epsilon \log M) = \tilde{O}(n^{2/3+2\delta/3})$. Then by Lemma 4.4, there exists $\overline{\alpha} \in \mathcal{H}_{A'}$ such that $\partial_{A'} f(\overline{\alpha}) \neq 0$, and so there must exist

$$\overline{\beta} \in \mathcal{L}_{A'}(\mathcal{H}_{A'}) \subseteq \mathcal{H}$$

such that $f(\overline{\beta}) \neq 0$. □

5.2 Depth-4 Formulas

Moving on to depth-4, the construction and proof are both very similar, with a slight change in the parameters. We first give a slightly more general form of Theorem 1.3 that we will later use for regular formulas.

Theorem 5.4 (General theorem for multilinear ΣΠΣΠ formulas)**.** *Let $\mathcal{C}$ be the class of multilinear ΣΠΣΠ formulas of top fan-in M and size S so that $(\log M)^3 \cdot \log S = o(n)$. There exists a hitting set $\mathcal{H}$ of size $|\mathcal{H}| = 2^{\tilde{O}(n^{2/3} \cdot \log M \cdot (\log S)^{1/3})}$ for $\mathcal{C}$, that can be constructed in time $\mathsf{poly}(|\mathcal{H}|)$.*

We note that Theorem 1.3 is an immediate corollary of Theorem 5.4.

Proof of Theorem 1.3. Apply Theorem 5.4 with $M = S = 2^{n^\delta}$ for some constant $0 < \delta < 1/4$ (the bound in Theorem 1.3 is meaningless for $\delta \geq 1/4$). It is clear that the conditions of Theorem 5.4 are met. Thus, we obtain a hitting set of size $|\mathcal{H}| = 2^{\tilde{O}(n^{2/3} \cdot \log M \cdot (\log S)^{1/3})} = 2^{\tilde{O}(n^{2/3+4\delta/3})}$ for the class. □

For the proof of Theorem 5.4 we will use the following construction that is similar to Construction 5.3.

Construction 5.5 (Hitting set for multilinear ΣΠΣΠ formulas)**.** *Let M and S be such that $(\log M)^3 \cdot \log S = o(n)$. Denote $M = 2^{n^\delta}$ (hence $S = 2^{o(n^{1-3\delta})}$). Let ϵ be such that*

$$n^\epsilon = n^{2/3} \cdot \log M / (\log S)^{2/3}.$$

Set

$$r = 2n^{\epsilon} \log S = 2n^{2/3} \cdot \log M \cdot (\log S)^{1/3}.$$

For every two disjoint sets $A, B \subseteq [n]$ with $|A|, |B| \leq r$, construct a set $\mathcal{H}_{A,B} \in \mathbb{F}^{n-(|A|+|B|)}$ using Construction 4.2 with parameters $\delta, \epsilon, n, k, s = 2^k$ and M (recall that in Construction 4.2 we set $k = n^{\delta} + 2\log n$). Finally, let

$$\mathcal{H} = \bigcup_{\substack{A,B \in \binom{[n]}{\leq r} \\ A \cap B = \emptyset}} \mathcal{L}_A^B(\mathcal{H}_{A,B}).$$

Proof of Theorem 5.4. We show that the set $\mathcal{H}$ constructed in Construction 5.5 has the desired properties. First, note that by Lemma 4.3, for every $A, B \subseteq [n]$ with $|A|, |B| \leq 2n^{\epsilon} \log S$, the set $\mathcal{H}_{A,B}$ has size

$$\left(M \cdot 2^{k^2 \log n}\right)^{\tilde{O}(n^{1-(\epsilon+\delta)/2})} = \left(2^{k^2 \log n}\right)^{\tilde{O}(n^{1-(\epsilon+\delta)/2})} = 2^{\tilde{O}(n^{1-\epsilon/2+3\delta/2})},$$

for $k = n^{\delta} + 2\log n$. It therefore follows that

$$|\mathcal{L}_A^B(\mathcal{H}_{A,B})| \leq 2^{|A|} \cdot |\mathcal{H}_{A,B}| = 2^{2n^{\epsilon} \log S} \cdot 2^{\tilde{O}(n^{1-\epsilon/2+3\delta/2})},$$

and that

$$|\mathcal{H}| \leq \sum_{i,j=0}^{2n^{\epsilon} \log S} \sum_{\substack{A \subseteq [n] \\ |A|=i}} \sum_{\substack{B \subseteq [n] \\ |B|=j}} |\mathcal{L}_A^B(\mathcal{H}_{A,B})| = 2^{\tilde{O}(n^{\epsilon} \log S)} \cdot 2^{\tilde{O}(n^{1-\epsilon/2+3\delta/2})}.$$

By our setting of parameters

$$\begin{aligned} |\mathcal{H}| \leq 2^{\tilde{O}(n^{\epsilon} \log S)} \cdot 2^{\tilde{O}(n^{1-\epsilon/2+3\delta/2})} &= 2^{\tilde{O}(n^{\epsilon} \log S + n^{1-\epsilon/2+3\delta/2})} \\ &= 2^{\tilde{O}(n^{2/3} \cdot \log M \cdot (\log S)^{1/3})}, \end{aligned}$$

where the last equality follows from our choice of ϵ in Construction 5.5 and the fact that $\log M = n^{\delta}$.

To show the hitting property of $\mathcal{H}$, let $f(x_1, \ldots, x_n)$ be a non-zero multilinear polynomial computed by a multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula of size S, and let $A', B' \subseteq [n]$ be the sets guaranteed by Corollary 3.5 with $\tau = n^{1-\epsilon}$. Thus, $|A'|, |B'| \leq 2n^{\epsilon} \log S$. Then, by Lemma 4.5, there exists $\overline{\alpha} \in \mathcal{H}_{A',B'}$ such that $\partial_{A'} f|_{B'=0}(\overline{\alpha}) \neq 0$, and so there must exist

$$\overline{\beta} \in \mathcal{L}_{A'}^{B'}(\mathcal{H}_{A',B'}) \subseteq \mathcal{H}$$

such that $f(\overline{\beta}) \neq 0$. □

6 Multilinear Depth-d Regular Formulas

6.1 Definition and Background

[KSS14] defined *regular formulas*, which consist of formulas with alternating layers of sum and product gates such that the fan-in of all the gates in any fixed layer is the same. In addition, they require the formal (syntactic) degree of the formula must be at most twice the (total) degree of its output polynomial. They showed that any $n^{O(1)}$-sized arithmetic circuit can be computed by a regular formula of size $n^{O(\log^2 n)}$ and proved a lower bound of $n^{\Omega(\log n)}$ on the size of regular formulas that compute some explicit polynomial in VNP.

In this section, we consider *multilinear regular formulas*, which are regular formulas with the extra condition that each gate computes a multilinear polynomial. Our syntactic requirement is different than that of [KSS14]: we require that the formal degree of the formula, which is the product of the fan-ins of the multiplication levels, is at most n. This restriction seems more suitable in the context of multilinear computation, in which the (semantic) degree of the polynomial computed is never more than n.

As was noted by [KSS14], by adding dummy gates to the computation tree, any formula can be converted to a formula in which the fan-in of all gates in any layer is the same. This transformation causes a blow-up to the size of the formula which is at most polynomial in the original size; it may, however, increase the fan-in of multiplication gates in a way that would make the formal degree larger than n. Thus, the requirement that the formal degree is at most n is quite restrictive, and in particular, it does not allow the easy transformation from multilinear formulas to regular multilinear formulas.

We now give the formal definition.

Definition 6.1 (Multilinear Regular Formulas)**.** *We say that a formula Φ is a multilinear $(a_1, p_1, a_2, p_2, \ldots, a_d, p_d, a_{d+1})$-regular formula, computing a multilinear polynomial $f(x_1, \ldots, x_n)$, if it is a multilinear $\Sigma^{[a_1]}\Pi^{[p_1]}\Sigma^{[a_2]}\Pi^{[p_2]}\ldots\Sigma^{[a_d]}\Pi^{[p_d]}\Sigma^{[a_{d+1}]}$-formula, and $\prod_{1\le i\le d} p_i \le n$. Notice that the size of such a formula is $(\prod_{1\le i\le d+1} a_i)\cdot(\prod_{1\le i\le d} p_i)$, and the formal degree of such a formula is given by $\deg(\Phi) = \prod_{1\le i\le d} p_i \le n$.*

Comparing with the definition given in Section 1.1, an $(a_1, p_1, a_2, p_2, \ldots, a_d, p_d, a_{d+1})$-regular formula has depth $2d+1$.

6.2 Reduction to Depth-4 Formulas

In this section, we reduce a multilinear depth-d regular formula to a depth-4 formula. We first give a depth reduction lemma (Lemma 6.2) that tells us that we can reduce the depth by one with a slight blow up in the fan-ins of the regular formula. We then use this lemma to obtain a depth-4 formula. The idea is to break the regular formula into two formulas (the top part and the bottom part), and then to apply the depth reduction lemma separately to these two formulas. The delicate part is that we wish to obtain a depth-4 formula that has a subexponential size hitting set, as in Theorem 5.4. For this we need the top fan-in M and the total size S to satisfy that $(\log M)^3 \cdot \log S = o(n)$. To achieve this we should carefully select

the point in which to divide the formula. This is done in Theorem 6.3.

We start with the depth reduction lemma.

Lemma 6.2 (Depth Reduction Lemma)**.** *Let Ψ be a multilinear $(a_1, p_1, a_2, p_2, 1)$-regular formula computing a polynomial $f(x_1, \ldots, x_n)$. Then, there exists a multilinear $(a_1 a_2^{p_1}, p_1 p_2, 1)$-regular formula Φ computing $f(x_1, \ldots, x_n)$.*

Proof. Notice that a multilinear $(a_1, p_1, a_2, p_2, 1)$-regular formula is a $\Sigma^{[a_1]}\Pi^{[p_1]}\Sigma^{[a_2]}\Pi^{[p_2]}$ formula. Writing Ψ as $\sum_{i=1}^{a_1}\prod_{j_i=1}^{p_1}\sum_{k_{j_i}=1}^{a_2} m(i, j_i, k_{j_i})$, where each $m(i, j_i, k_{j_i})$ is a monomial that is a product of p_2 input gates, and by expanding the expression above by computing all the products, we obtain:

$$\sum_{i=1}^{a_1}\prod_{j_i=1}^{p_1}\sum_{k_{j_i}=1}^{a_2} m(i, j_i, k_{j_i}) =$$

$$\sum_{i=1}^{a_1}\left(\sum_{k_{i,1}=1}^{a_2}\sum_{k_{i,2}=1}^{a_2}\cdots\sum_{k_{i,p_1}=1}^{a_2}\prod_{t=1}^{p_1} m(i, t, k_{i,t})\right). \tag{5}$$

Since $m(i, j_i, k_{j_i})$ is a product of p_2 input gates, and since the right hand side of (5) computes a product of p_1 of these terms, each monomial computed by $\prod_{t=1}^{p_1} m(i, t, k_{i,t})$ is a product of $p_1 p_2$ input gates.

Since the sums on the right hand side of (5) are over all tuples of the form $(i, k_{i,1}, k_{i,2} \ldots, k_{i,p_1}) \in [a_1] \times [a_2]^{p_1}$, we have that there are exactly $a_1 \cdot a_2^{p_1}$ summands. Hence, the right hand side of (5) is the expression of a multilinear $(a_1 a_2^{p_1}, p_1 p_2, 1)$-regular formula. □

By repeatedly applying the depth reduction lemma above, we obtain the following theorem:

Theorem 6.3 (Depth Reduction of Regular Formulas)**.** *Let $d \geq 2$ be an integer, $c \in \mathbb{R}$ a constant such that $c \geq 3$, and Ψ a multilinear $(a_1, p_1, a_2, p_2, \ldots, a_d, p_d, a_{d+1})$-regular formula of size S computing a multilinear polynomial $f(x_1, \ldots, x_n)$. Then, one of the following conditions must happen:*

(i) For $M = S$, there exists a $\Sigma^{[M]}\Pi\Sigma\Pi$ formula of size $O(S \cdot n^{n^{1-(1/c)^d}})$ computing $f(x_1, \ldots, x_n)$, or

(ii) There exists $t \in [d-1]$ such that there is a multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula Φ computing $f(x_1, \ldots, x_n)$, with

$$M = S^{n^\alpha} \quad \text{and} \quad |\Phi| \leq 2Mn \cdot n^{n^{1-(c-1)\alpha}},$$

for $\alpha = \frac{1}{c-1} \cdot \left(\frac{1}{c}\right)^{d-t}$.

Proof. Recall that the size of the formula S satisfies $S = \left(\prod_{1 \leq i \leq d+1} a_i\right) \cdot \left(\prod_{1 \leq i \leq d} p_i\right)$. We have three cases to analyze:

Case 1 (small total degree): If $\prod_{i=1}^{d} p_i \leq n^{1-(1/c)^d}$, then we can simply write the polynomial $f(x_1, \ldots, x_n)$ as a sum of monomials, which would give us a multilinear $\Sigma\Pi$ formula Φ of size

$$|\Phi| \leq \sum_{i=0}^{n^{1-(1/c)^d}} \binom{n}{i} = O(n^{n^{1-(1/c)^d}}),$$

which is clearly of the form $\Sigma^{[1]}\Pi\Sigma\Pi \subseteq \Sigma^{[S]}\Pi\Sigma\Pi$ and of the required size for item i.

Case 2 (large p_1): If $p_1 > n^{(1/c)^d}$, then notice that the regular formula Ψ can be written in the form

$$\sum_{i=1}^{a_1} \prod_{j=1}^{p_1} f_{i,j},$$

where each $f_{i,j}$ is a multilinear $(a_2, p_2, \ldots, a_d, p_d, a_{d+1})$-regular formula. Hence, each $f_{i,j}$ is a polynomial of degree bounded by $\prod_{i=2}^{d} p_i \leq n/p_1 < n^{1-(1/c)^d}$. Therefore, expanding each $f_{i,j}$ into a sum of monomials, we obtain a formula Φ of the form $\Sigma^{[a_1]}\Pi\Sigma\Pi$ and of size

$$|\Phi| \leq a_1 \cdot p_1 \cdot \sum_{i=0}^{n^{1-(1/c)^d}} \binom{n}{i} = O(a_1 p_1 n^{n^{1-(1/c)^d}}) = O(S \cdot n^{n^{1-(1/c)^d}}).$$

This too satisfies item i.

Case 3 (high degree but small p_1): In this case, we can assume that $\prod_{i=1}^{d} p_i > n^{1-(1/c)^d}$ and that $p_1 \leq n^{(1/c)^d}$. It follows there exists an index $t \in [d-1]$ satisfying

$$p_t \leq n^{(1/c)^{d+1-t}} \quad \text{and} \quad p_{t+1} > n^{(1/c)^{d+1-(t+1)}} = n^{(1/c)^{d-t}},$$

since otherwise, using $c \geq 3$, we would have that

$$\prod_{i=1}^{d} p_i \leq \prod_{i=1}^{d} n^{(1/c)^{d+1-i}} = n^{\sum_{i=1}^{d}(1/c)^{d+1-i}} < n^{\frac{1}{c-1}} < n^{1-(1/c)^d},$$

which contradicts the assumption on the degree.

Notice that we can express Ψ in the form

$$\sum_{i_1=1}^{a_1} \prod_{j_1=1}^{p_1} \cdots \sum_{i_{t+1}=1}^{a_{t+1}} \prod_{j_{t+1}=1}^{p_{t+1}} f_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}}, \tag{6}$$

where each $f_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}}$ is a multilinear $(a_{t+2}, p_{t+2}, \ldots, a_d, p_d, a_{d+1})$-regular formula. We shall analyze separately each of the $f_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}}$ and the $(a_1, p_1, \ldots, a_{t+1}, p_{t+1}, 1)$-regular formula "above" them.

Notice that each $f_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}}$ is a polynomial of degree bounded by

$$\prod_{i=t+2}^{d} p_i < n/p_{t+1} < n^{1-(1/c)^{d-t}}.$$

Therefore, when expressing each $f_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}}$ as a sum of monomials, its sparsity is upper bounded by

$$\sum_{i=0}^{n^{1-(1/c)^{d-t}}} \binom{n}{i} \leq 2n^{n^{1-(1/c)^{d-t}}}. \tag{7}$$

Now, if in (6) we replace each polynomial $f_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}}$ with a new variable $y_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}}$, then we get an $(a_1, p_1, \ldots, a_{t+1}, p_{t+1}, 1)$-regular formula in the y variables

$$\Phi_1 = \sum_{i_1=1}^{a_1} \prod_{j_1=1}^{p_1} \cdots \sum_{i_{t+1}=1}^{a_{t+1}} \prod_{j_{t+1}=1}^{p_{t+1}} y_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}}.$$

It is clear that Ψ is the composition of Φ_1 with the assignment $y_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}} = f_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}}$.

By applying the Depth Reduction (Lemma 6.2) repeatedly to Φ_1, we obtain that Φ_1 becomes a multilinear $\left(\prod_{i=1}^{t+1} a_i^{\pi_{i-1}}, \pi_{t+1}, 1\right)$-regular formula Φ_2, where $\pi_k = \prod_{i=1}^{k} p_i$, for any $1 \leq k \leq d$ (and $\pi_0 = 1$).

Composing Φ_2 with the assignment $y_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}} = f_{i_1,\ldots,i_{t+1},j_1,\ldots,j_{t+1}}$, we obtain the following depth-4 regular formula Φ (after some proper relabelling):

$$\Phi = \sum_{i=1}^{M} \prod_{j=1}^{\pi_{t+1}} f_{i,j}, \quad \text{where} \quad M = \prod_{i=1}^{t+1} a_i^{\pi_{i-1}}. \tag{8}$$

Notice that, by our choice of the parameter t, we have

$$M = \prod_{i=1}^{t+1} a_i^{\pi_{i-1}} \leq S^{\pi_t} = S^{\prod_{i=1}^{t} p_i} \leq S^{\prod_{i=1}^{t} n^{(1/c)^{d+1-i}}} < S^{n^{\alpha}},$$

where $\alpha = \left(\frac{1}{c}\right)^{d-t} \frac{1}{c-1}$. Since, by equation (7), each $f_{i,j}$ has sparsity bounded by $2n^{n^{1-(1/c)^{d-t}}} = 2n^{n^{1-(c-1)\alpha}}$, we have that Φ is a $\Sigma^{[M]}\Pi\Sigma\Pi$ formula of size bounded by:

$$|\Phi| \leq M \cdot \pi_{t+1} \cdot 2n^{1-(c-1)\alpha} \leq M \cdot n \cdot 2n^{1-(c-1)\alpha}.$$

This satisfies the conditions of item ii, and so this concludes the proof of the theorem. □

6.3 PIT for Regular Formulas

In this section, we construct our hitting set for regular formulas. Since the reduction done in Theorem 6.3 reduces a multilinear depth-d regular formula to one of two types of depth-4 formulas, our hitting set will be the union of two hitting sets, $\mathcal{F}_d$ and $\mathcal{G}_d$, each one designed to hit a specific type.

Theorem 6.4 (Theorem 1.5, restated)**.** *For $d \geq 2$, let $\mathcal{C}_d$ be the class of multilinear polynomials computed by $(a_1, p_1, a_2, p_2, \ldots, a_d, p_d, a_{d+1})$-regular formulas of size $S \leq 2^{n^\delta}$ computing a multilinear polynomial $f(x_1, \ldots, x_n)$, where $\delta = \frac{1}{5^{d+1}}$. Then, there exists a hitting set $\mathcal{H}_d$ of size $|\mathcal{H}_d| = 2^{\tilde{O}(n^{1-\delta/3})}$ for $\mathcal{C}_d$, that can be constructed in time $\mathsf{poly}(|\mathcal{H}_d|)$.*

The proof follows by combining Theorem 6.3 with the hitting set guaranteed by Theorem 5.4.

Proof. Let $f(x_1, \ldots, x_n)$ be a polynomial computed by a formula $\Psi \in \mathcal{C}_d$. By Theorem 6.3, with the constant $c = 5$, there are two cases of the depth reduction to analyze. For each case we will give a hitting set (using Theorem 5.4) and thus the union of the sets will be a hitting set for $\mathcal{C}_d$.

Case 1: For $M = S \leq 2^{n^\delta}$, there exists a $\Sigma^{[M]}\Pi\Sigma\Pi$ formula Φ of size $|\Phi| = O(S \cdot n^{n^{1-(1/5)^d}}) = O(2^{n^\delta} \cdot n^{n^{1-5\delta}})$ computing $f(x_1, \ldots, x_n)$. By Theorem 5.4, there exists a hitting set $\mathcal{H}'$ that hits all such non-zero formulas with

$$|\mathcal{H}'| = 2^{\tilde{O}(n^{2/3} \cdot \log M \cdot (\log |\Phi|)^{1/3})}.$$

Observe that

$$\begin{aligned}(\log M)^3 \cdot \log |\Phi| = n^{3\delta} \cdot (n^\delta + n^{1-5\delta} \cdot \log n) &= n^{4\delta} + n^{1-2\delta} \log n \\ &= \tilde{O}(n^{1-2\delta}).\end{aligned}$$

Hence

$$|\mathcal{H}'| = 2^{\tilde{O}(n^{2/3} \cdot \log M \cdot (\log |\Phi|)^{1/3})} = 2^{\tilde{O}(n^{2/3} \cdot n^{(1-2\delta)/3})} = 2^{\tilde{O}(n^{1-2\delta/3})}.$$

Case 2: There exists $t \in [d-1]$ such that for $\alpha_t = \frac{1}{4} \cdot \left(\frac{1}{5}\right)^{d-t} \leq \frac{1}{20}$, there exists a multilinear $\Sigma^{[M]}\Pi\Sigma\Pi$ formula Φ computing $f(x_1, \ldots, x_n)$, where the top fan-in $M = S^{n^{\alpha_t}} = 2^{n^{\delta+\alpha_t}}$ and the size is bounded by $|\Phi| \leq 2Mn \cdot n^{n^{1-4\alpha_t}}$. Again, by Theorem 5.4, there exists a hitting set $\mathcal{H}''$ that hits all such non-zero formulas with

$$|\mathcal{H}''| = 2^{\tilde{O}(n^{2/3} \cdot \log M \cdot (\log |\Phi|)^{1/3})}.$$

We now have that

$$\begin{aligned}(\log M)^3 \cdot \log |\Phi| &= \tilde{O}\left(n^{3(\delta+\alpha_t)} \cdot (n^{\delta+\alpha_t} + n^{1-4\alpha_t})\right) \\ &= \tilde{O}\left(n^{4(\delta+\alpha_t)} + n^{1+3\delta-\alpha_t}\right) \\ &= \tilde{O}\left(n^{1-\delta}\right),\end{aligned}$$

where the last equality holds as, by our choice of parameters, for all $t \in [d-1]$, $5\delta + 4\alpha_t < 1$ and $4\delta < \alpha_t$. This implies that

$$|\mathcal{H}''| = 2^{\tilde{O}(n^{2/3} \cdot \log M \cdot (\log |\Phi|)^{1/3})} = 2^{\tilde{O}(n^{2/3} \cdot n^{(1-\delta)/3})} = 2^{\tilde{O}(n^{1-\delta/3})},$$

as stated. □

We note that we did not attempt to optimize the parameters in the theorem as, using our current proof, the exponent is going to be of the form $n^{1-1/\exp(d)}$ anyway.

7 Lower Bounds for Bounded Depth Multilinear Formulas

As we noted earlier, the connection between construction of hitting sets and lower bounds for explicit polynomials is well established. The following theorem was proved by [HS80] and [Agr05], albeit we cite only a special case which matches our use of it:

Theorem 7.1 (A special case of [HS80, Agr05])**.** *Suppose there is a black-box deterministic PIT algorithm for a class $\mathcal{C}$ of multilinear circuits, that outputs a hitting set $\mathcal{H}$ of size $|\mathcal{H}| = 2^{n^\alpha} < 2^n$ and runs in time $\mathsf{poly}(|\mathcal{H}|)$, such that $\mathcal{H}$ hits circuits of size at most 2^{n^δ}. Then, there exists a multilinear polynomial $f(x_1, \ldots, x_n)$ such that any circuit from the class $\mathcal{C}$ computing f must be of size at least 2^{n^δ}, and the coefficients of f can be found in time $2^{O(n)}$.*

Theorem 7.1 is proved by finding a non-zero polynomial $f(x_1, \ldots, x_n)$ which vanishes on the entire hitting set $\mathcal{H}$ of size 2^{n^α}, and then, by definition, f cannot have circuits of size 2^{n^δ}. Finding f amounts to finding a non-zero solution to a homogenous system of linear equations whose unknowns are the coefficients of the 2^n possible multilinear monomials in $x_1, \ldots, x_n$. As long as $2^n > |\mathcal{H}| = 2^{n^\alpha}$, a non-zero solution is guaranteed to exist.

Our lower bounds now follow as a direct application of our hitting set constructions and Theorem 7.1.

Proofs of Corollary 1.2, Corollary 1.4 and Corollary 1.6. In light of Theorem 7.1, we only need to pick δ so that the hitting sets we constructed have size less than 2^n. The appropriate choices, by Theorem 1.1, Theorem 1.3 and Theorem 1.5, respectively, can be seen to be $\delta = 1/2 - O(\log\log n/\log n)$ (for depth-3), $\delta = 1/4 - O(\log\log n/\log n)$ (for depth-4) and $\delta = \frac{1}{5^{\lfloor d/2 \rfloor + 1}} = O\left(\frac{1}{\sqrt{5}^d}\right)$ (for depth-d regular formulas). □

8 Conclusion and Open Questions

We conclude this paper with some obvious open problems. First, as noted in Section 1.3, the lower bounds that we get from our hitting sets are not as good as the best lower bounds for these models. Can one improve our construction to yield lower bounds matching the best known lower bounds?

Currently, the size of the hitting set that we get for depth-d regular multilinear formulas is roughly $\exp(n^{1-1/\exp(d)})$. Can the bound be improved to $\exp(n^{1-\Omega(1/d)})$? In our proof the reason for this exponential loss is that we reduce the regular formula to a $\Sigma^{[M]}\Pi\Sigma\Pi$ formula of size S and we need M and S to satisfy (because of Theorem 5.4) $(\log M)^3 \cdot \log S = o(n)$. In particular, if $M = 2^{n^\delta}$ and $S = 2^{n^\gamma}$ then we require that $3\delta + \gamma < 1$. Notice that, in the depth reduction theorem (Theorem 6.3), if we start with a regular formula of size 2^{n^δ} then, if we break the formula at layer t, we roughly get a top fan-in of $M = 2^{n^\delta \cdot p_1 \cdot p_2 \cdots p_t}$ and bottom sparsity of (roughly) $\exp(n^{1-p_{t+1}})$. This gives a size upper bound of (roughly) $S = 2^{n^\delta \cdot p_1 \cdot p_2 \cdots p_t} \cdot \exp(n^{1-p_{t+1}})$. To match the requirement $(\log M)^3 \cdot \log S = o(n)$, we get that p_{t+1} must be larger than $3p_1 \cdots p_t$. This leads to an argument in which we require the degree of the product gates to increase exponentially. This is more or less the cause of the exponential loss in our argument.

Finally, another natural question is to extend our argument from depth-d regular multilinear formulas to arbitrary depth-d multilinear formulas.

Acknowledgements

The authors would like to thank Zeev Dvir and Avi Wigderson for helpful discussions during the course of this work, and Michael Forbes and Mrinal Kumar for helpful suggestions and remarks on earlier versions of this paper. We also thank the anonymous reviewers for their thoughtful comments.

Rafael Oliveira is supported by NSF Career award (1451191) and by CCF-1523816 award. Amir Shpilka is supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 257575, and by the Israel Science Foundation (grant number 339/10). Ben Lee Volk is supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 257575.

An extended abstract of this work appeared as [OSV15].

References

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

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

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

[AS08] N. Alon and J. Spencer. *The probabilistic method*. J. Wiley, 3 edition, 2008.

[ASS13] M. Agrawal, C. Saha, and N. 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.

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

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

[AvMV11] M. Anderson, D. van Melkebeek, and I. Volkovich. Derandomizing polynomial identity testing for multilinear constant-read formulae. In *Proceedings of the 26th Annual CCC*, pages 273–282, 2011.

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

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

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

[FS12] M. A. Forbes and A. Shpilka. On identity testing of tensors, low-rank recovery and compressed sensing. In *Proceedings of the 44th annual STOC*, pages 163–172, 2012.

[FS13] M. A. Forbes and A. 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] M. A. Forbes, R. Saptharishi, and A. 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] A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi. Arithmetic circuits: A chasm at depth three. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013*, pages 578–587, 2013.

[GKST15] R. Gurjar, A. Korwar, N. Saxena, and T. Thierauf. Deterministic identity testing for sum of read-once oblivious arithmetic branching programs. In D. Zuckerman, editor, *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, volume 33 of *LIPIcs*, pages 323–346. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015.

[Gup14] A. Gupta. Algebraic geometric techniques for depth-4 PIT & sylvester-gallai conjectures for varieties. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:130, 2014.

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

[KI03] V. Kabanets and R. Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. In *Proceedings of the 35th Annual STOC*, pages 355–364, 2003.

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

[KMSV13] Z. S. Karnin, P. Mukhopadhyay, A. Shpilka, and I. Volkovich. Deterministic identity testing of depth-4 multilinear circuits with bounded top fan-in. *SIAM J. Comput.*, 42(6):2114–2131, 2013.

[Koi10] P. Koiran. Arithmetic circuits: the chasm at depth four gets wider. *CoRR*, abs/1006.4700, 2010.

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

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

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

[KSS14] N. Kayal, C. Saha, and R. Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *Symposium on Theory of Computing, STOC 2014*, pages 146–153, 2014.

[Nis91] N. Nisan. Lower bounds for non-commutative computation. In *Proceedings of the 23rd Annual STOC*, pages 410–418, 1991.

[NW96] N. Nisan and A. Wigderson. Lower bound on arithmetic circuits via partial derivatives. *Computational Complexity*, 6:217–234, 1996.

[OSV15] R. Oliveira, A. Shpilka, and B. L. Volk. Subexponential size hitting sets for bounded depth multilinear formulas. In D. Zuckerman, editor, *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, volume 33 of *LIPIcs*, pages 304–322. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015.

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

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

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

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

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

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

[SS11] N. Saxena and C. Seshadhri. Blackbox identity testing for bounded top fanin depth-3 circuits: the field doesn't matter. In *Proceedings of the 43rd Annual STOC*, pages 431–440, 2011.

[SV08] A. Shpilka and I. Volkovich. Read-once polynomial identity testing. In *Proceedings of the 40th Annual STOC*, pages 507–516, 2008.

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

[SV10] A. Shpilka and I. Volkovich. On the relation between polynomial identity testing and finding variable disjoint factors. In *ICALP (1)*, pages 408–419, 2010.

[SV11] S. Saraf and I. Volkovich. Black-box identity testing of depth-4 multilinear circuits. In *Proceedings of the 43rd annual STOC*, pages 421–430, 2011.

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

[Tav13] S. Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *Mathematical Foundations of Computer Science 2013 - 38th International Symposium, MFCS 2013*, pages 813–824, 2013.

[VSBR83] L. G. Valiant, S. Skyum, S. Berkowitz, and C. Rackoff. Fast parallel computation of polynomials using few processors. *SIAM J. on Computing*, 12(4):641–644, November 1983.

[Zip79] R. Zippel. Probabilistic algorithms for sparse polynomials. In *EUROSAM*, pages 216–226, 1979.