A Quadratic Size-Hierarchy Theorem for Small-Depth Multilinear Formulas

Suryajith Chillara
Department of CSE, IIT Bombay, Mumbai, India
suryajith@cse.iitb.ac.in

Nutan Limaye
Department of CSE, IIT Bombay, Mumbai, India
nutan@cse.iitb.ac.in

Srikanth Srinivasan
Department of Mathematics, IIT Bombay, Mumbai, India
srikanth@math.iitb.ac.in

Abstract

We show explicit separations between the expressive powers of multilinear formulas of small-depth and all polynomial sizes.

Formally, for any $s = s(n) = n^{O(1)}$ and any $\delta > 0$, we construct explicit families of multilinear polynomials $P_n \in \mathbb{F}[x_1, \ldots, x_n]$ that have multilinear formulas of size s and depth three but no multilinear formulas of size $s^{1/2-\delta}$ and depth $o(\log n/\log\log n)$.

As far as we know, this is the first such result for an algebraic model of computation.

Our proof can be viewed as a derandomization of a lower bound technique of Raz (JACM 2009) using ε-biased spaces.

2012 ACM Subject Classification Theory of computation → Algebraic complexity theory

Keywords and phrases Algebraic circuit complexity, Multilinear formulas, Lower Bounds

Digital Object Identifier 10.4230/LIPIcs.ICALP.2018.36

1 Introduction

The main aim of Computational Complexity is to understand as precisely as possible the amounts of computational resources required to perform interesting computational tasks. These resources could be of various kinds depending on the computational model under consideration, e.g., time and space for traditional algorithms, size and depth for Boolean and Algebraic circuits, the number of random bits for randomized algorithms, total communication for communication protocols and so on.

A fundamental question regarding any given resource is if access to more of that resource strictly increases the power of the underlying computational model. Classical theorems in Computational Complexity theory such as the *Time Hierarchy theorem* [13] and *Space Hierarchy Theorem* [18] answer this question (in the affirmative) for the resources of time and space on multitape Turing Machines.

In this paper, we consider an analogous question for *Algebraic formulas*. Algebraic formulas (and their variants such as Algebraic circuits, Algebraic Branching Programs etc.) are the natural computational model for computing multivariate polynomials over some underlying domain, usually a field $\mathbb{F}$. Many natural problems, such as the Determinant, Permanent, Matrix Multiplication, the Fast Fourier Transform etc. fit into this general framework. Algebraic formulas compute multivariate polynomials from the ring $\mathbb{F}[x_1, \ldots, x_n]$ using the natural algebraic operations of sum and product.

45th International Colloquium on Automata, Languages, and Programming (ICALP 2018).
Editors: Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella;
Article No. 36; pp. 36:1–36:13
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The *size* of an algebraic formula is the number of algebraic operations it uses and is a measure of the efficiency of the formula (it roughly corresponds to the time in the case of traditional algorithms). One can also consider the *depth* of the formula, which measures how nested the algebraic operations in the formula are, and corresponds to how parallelizable the underlying procedure is. In this paper, we consider the question of proving a *Size-hierarchy theorem* for Algebraic formulas. Somewhat informally, we ask the following.

▶ **Question 1** (The Size-Hierarchy Question). *For any $\delta > 0$, are there explicit families of polynomials $P_n \in \mathbb{F}[x_1, \ldots, x_n]$ that can be computed by formulas of size $s(n)$ but not by formulas of size less than $s(n)^{1-\delta}$?*

(See Section 2 for the definition of "explicit". We note that requiring explicit polynomial families is necessary since counting arguments easily yield the existence of polynomials that have formulas of size $s(n)$ but not size $s(n)^{1-\delta}$ for most reasonable functions $s(n)$. However, as is standard in Circuit Complexity, the interesting question is finding an explicit function that witnesses this separation.)

As of now, the size-hierarchy question is far beyond the range of our techniques for most non-trivial parameters. Indeed, we do not have techniques to prove *any* explicit strong lower bounds for general algebraic formulas, let alone lower bounds for explicit polynomials that further have algebraic formulas of some prescribed size $s(n)$.

So, we restrict ourselves to the setting of *multilinear formulas*, which are algebraic formulas that are required to compute a multilinear polynomial[1] at each intermediate stage of computation. Note that the most efficient formula for computing some multilinear polynomial need not be multilinear (this is known to be true for small-depth multilinear formulas [8]) and so this is indeed a restriction. Nevertheless, it is a reasonable restriction for formulas that compute multilinear polynomials and has been the focus of a large body of work [20, 22, 21, 24, 23, 25, 15, 11, 17, 8, 3, 7] with interesting upper as well as lower bound results.

Therefore, it is natural to consider the size-hierarchy question in the setting of multilinear formulas. It follows from the work of Raz [21] and Raz and Yehudayoff [24] that for $s(n) \leq n^{O(1)}$, there are explicit polynomial families that can be computed by multilinear formulas of size s but not by multilinear formulas of size less than s^{δ_0} for some positive, but small, δ_0. (One needs to mine the proofs for the exact value of δ_0. The best value that we could obtain for δ_0 was less than 1/30.)

In this paper, we prove a near-tight multilinear size-hierarchy theorem for *small-depth* multilinear formulas. It is known [6, 5] that any multilinear formula of polynomial size s can be converted to another of size at most $s^{1+\delta}$ and depth $O(\log n)$ (for any fixed $\delta > 0$). Below, we consider multilinear formulas of smaller depth $O(\log n/\log\log n)$. The main result is the following.

▶ **Theorem 2.** *For any fixed $c \in \mathbb{N}$ and $\delta \in (0, 1/2)$, there exists an explicit polynomial family $P_n \in \mathbb{F}[x_1, \ldots, x_n]$ that has a multilinear formula of depth 3 and size at most $s = O(n^c)$ but no multilinear formulas of size less than $s^{(1/2)-\delta}$ and depth $\Delta < \log n/100 \log\log n$.*

As such, our result is incomparable with the separation implied by [21, 24] since we further assume that our formulas have small depth. However, in the setting of small-depth multilinear formulas, our result improves on the separation of [21, 24] in two ways. The first is that we obtain a separation of s versus $s^{(1/2)-\delta}$ as opposed to the s versus s^{δ_0} separation

[1] Recall that a *multilinear* polynomial $P \in \mathbb{F}[x_1, \ldots, x_n]$ is one in which each variable has degree at most 1.

obtained by [21, 24]. The second is that the polynomials P_n that we consider have *depth-three* formulas of size s. This is in contrast to the polynomials constructed in [21, 24], that have formulas of size s but also depth $\Omega(\sqrt{\log s})$, which is considerably larger.[2] Finally, our proof technique is based on a derandomization of a special case of a lower bound technique of Raz [22]. This derandomization leaves some scope for improvement: an optimal result along these lines would resolve the size-hierarchy question optimally yielding a separation between sizes s and $s^{1-\delta}$ for any $\delta > 0$. An optimal derandomization of the more general lower bound technique of Raz would yield the same result for general multilinear formulas (without the depth restriction).

1.1 Related Work

Our work is partially motivated by hierarchy theorems for *Boolean circuits*, which compute functions $f : \{0,1\}^n \to \{0,1\}$ using simple Boolean operations such as AND and OR. Superpolynomial lower bounds have been known for constant-depth Boolean circuits since the early 1980s [12, 1, 14]. However, a size hierarchy theorem in this setting was obtained relatively recently by Rossman [26], who proved that for any constant $k \in \mathbb{N}$, there are explicit functions that have depth-two Boolean circuits of size $O(n^k)$ but not of size less than $n^{k/4}$. This was then improved by Amano [4] who showed that for any fixed k and δ there are explicit functions that have depth two Boolean circuits of size $O(n^k)$ but no circuits of size less than $n^{k-\delta}$.

Our proofs build on standard techniques for proving lower bounds for multilinear formulas. While these ideas are essentially due to Raz [22], we use a high-level reformulation of this argument that appears in the survey of Shpilka and Yehudayoff [28].

1.2 Proof Outline

Our proof can be seen as a derandomization of (a special case of a) technique of Raz [22] for proving lower bounds for multilinear formulas. Here, we follow a well-known reformulation of this proof that appears in [15, 28, 11].

Say we want to show that a multilinear polynomial $P \in \mathbb{F}[X]$ does not have a small-depth multilinear formula of size s'. The proof strategy consists of two steps. The first step is a decomposition lemma that says that any multilinear polynomial P that is computed by a small-depth multilinear formula of size s' can be written as a sum of s' polynomials, each of which is of the form $f = f_1 \cdot f_2 \cdot \dots \cdot f_t$ where the f_i's are multilinear polynomials over pairwise disjoint *non-empty* sets of variables $X_1, \dots, X_t$ that partition the variable set X. Following [28], we call such a polynomial a *t-product polynomial.* Here, t is some growing function of the number of variables n and the depth of the formula.

Thus, to show that P does not have a small-depth multilinear formula of size s', it suffices to show that it cannot be written as a sum of s' many t-product polynomials for a large t. This is the second step. To argue this, Raz used a rank-based argument. Specifically, we partition the variables X into any two sets[3] Y and Z and consider any polynomial $P(X)$ as a polynomial in the variables in Y, with coefficients from $\mathbb{F}[Z]$. The dimension of the space of coefficients (as vectors over the base field $\mathbb{F}$) is considered to be a measure of the complexity of P. The idea is that polynomials with small formulas will have low complexity and hence, by choosing a P of high complexity we obtain a lower bound.

[2] In fact, the polynomial families from [21, 24] do not have formulas of constant-depth and comparable size. This follows from a later lower bound result of Raz and Yehudayoff [25].

[3] Actually, the sets Y and Z also need to be of equal size. We ignore this for now for the sake of exposition.

Unfortunately, this idea by itself is not enough to prove a strong lower bound. This is because of the fact that given any partition (Y, Z), there is a small depth-2 multilinear formula $F_{(Y,Z)}$ (which is also a t-product polynomial for large t) that has maximum dimension w.r.t. this partition. To overcome this, we consider a *random* partition (Y, Z) and show that any t-product polynomial will have low-rank w.r.t. this random partition with high probability. Using a union bound, we then show that any sum of s' many t-product polynomials must be of low-rank w.r.t. some partition. If, on the other hand, our choice of polynomial P has high rank w.r.t. *every* partition, we obtain a lower bound.

The crux of the matter therefore is to argue that given any t-product polynomial $f = f_1 \cdots f_t$ as above, and a random partition (Y, Z), the polynomial f is low-rank w.h.p., w.r.t. this partition. Formally, for the union bound over s' many such polynomials to go through, we need the following criterion to hold.

$$\Pr_{(Y,Z)}[f \text{ does not have small rank w.r.t. } (Y, Z)] < \frac{1}{s'}.$$

Raz [22] showed that this reduces to a combinatorial discrepancy question. Note that any choice of partition (Y, Z) of X induces a partition (Y_i, Z_i) of each X_i $(i \in [t])$. To prove the above bound, it actually suffices to show[4] that

$$\Pr_{(Y,Z)}[|\{i \in [t] \mid |Y_i| - \lfloor |X_i|/2 \rfloor \text{ odd}\}| \text{ is small}] < \frac{1}{s'}.$$

But this is quite easy to argue. Since (Y, Z) is a random partition, each $|Y_i| - \lfloor |X_i|/2 \rfloor$ is odd with probability $1/2$. Since the X_i $(i \in [t])$ are pairwise disjoint, these events are mutually independent and hence by a Chernoff bound, it is easy to show that the probability of the above event is $1/2^{\Omega(t)} < 1/s'$ for the specific t that we obtain in the decomposition lemma (this is where the small-depth assumption comes in). This completes the proof.

The derandomization. Our idea is to simulate the purely random partition argument of Raz, but using instead a random partition from a small predefined set $\mathcal{S} = \{(Y^{(1)}, Z^{(1)}), \ldots, (Y^{(s)}, Z^{(s)})\}$ of partitions. We would like to show that for a random $j \in [s]$ and any product polynomial $f = f_1 \cdots f_t$ as above, we similarly obtain

$$\Pr_j[|\{i \in [t] \mid |Y_i^{(j)}| - \lfloor |X_i|/2 \rfloor \text{ is odd}\}| \text{ is small}] < \frac{1}{s'} \tag{1}$$

where $Y_i^{(j)}$ denotes the set $X_i \cap Y^{(j)}$.

If we have a set $\mathcal{S}$ as above, we obtain a size-hierarchy theorem as follows. First, we construct a multilinear formula $F_{\mathcal{S}}$ of size roughly s that is full-rank w.r.t. each of the partitions in $\mathcal{S}$: this is done by simply taking a suitable linear combination of the formulas $F_{(Y^{(j)},Z^{(j)})}$ $(j \in [s])$ mentioned above. On the other hand, we know that given any small-depth multilinear formula F' of size s', (1) implies that F' cannot compute a polynomial that is full-rank w.r.t. all the partitions in $\mathcal{S}$, and in particular cannot compute the same polynomial as F. This proves a separation between small-depth multilinear formulas of size s and size s'. The question now is – how do we construct such a set $\mathcal{S}$ as described above while keeping s as close to s' as possible?

[4] Raz in fact shows that it suffices to bound the probability that $\sum_{i\in[t]} ||Y_i| - |X_i|/2|$ is small. Here, we only use the fact that for each i such that $|Y_i| - \lfloor |X_i|/2 \rfloor$ is odd, we must have $||Y_i| - |X_i|/2| \geq 1/2$. This harks back to an earlier result of Nisan and Wigderson [20] who use a simpler parity argument to prove a lower bound for *set-multilinear* formulas.

Our construction of the set $\mathcal{S}$ follows two steps. We first show that it suffices to construct a set that satisfies the following somewhat weaker condition.

$$\Pr_j[|\{i \in [t] \mid |Y_i^{(j)}| - \lfloor |X_i|/2 \rfloor \text{ is odd}\}| \ = 0] < \frac{1}{2s'} \tag{2}$$

We deduce (1) from (2) by adapting a combinatorial proof of the Chernoff bound that appears in a result of Impagliazzo and Kabanets [16].

Finally, we show that for (2) it suffices to use *Small-Bias* spaces, which are a standard tool in the derandomization literature [19]. Known explicit constructions of small-bias spaces [2] yield sets $\mathcal{S}$ satisfying (2) of size roughly $s = (s')^2$. This yields a separation between size s and size roughly $\sqrt{s}$ as stated in Theorem 2.

We remark that non-constructively, we can show that there exist sets $\mathcal{S}$ satisfying (2) of size roughly s'. Constructing such sets explicitly would improve our result to a near-tight size-hierarchy theorem.

2 Preliminaries

Recall that a polynomial $P \in \mathbb{F}[x_1, \ldots, x_n]$ is *multilinear* if each variable has degree at most 1 in P.

A family $\{P_n \in \mathbb{F}[x_1, \ldots, x_n] \mid n \geq 1\}$ of multilinear polynomials is said to be *explicit* if there is a deterministic algorithm that given as input n and a monomial m over the variables $x_1, \ldots, x_n$, computes in time $\mathrm{poly}(n)$ the coefficient of the monomial m in P_n.

2.1 Multilinear formulas

For the detailed introduction to algebraic formulas, we refer the reader to standard resources such as [28, 27]. Having said that, we do make a few remarks.

- All the gates in our formulas may have *unbounded* fan-in.
- The size of a formula refers to the number of gates (including input gates) in it, and depth of the formula refers to the number of gates on the longest path from an input gate to output gate.

An algebraic formula F computing a polynomial from $\mathbb{F}[X]$ is said to be *multilinear* if each gate in the formula computes a multilinear polynomial.

We state below a decomposition lemma for small-depth multilinear formulas.

Define a polynomial $f \in \mathbb{F}[X]$ to be a *t-product polynomial* if we can write f as $f_1 \cdots f_t$, where we can find a partition of X into pairwise disjoint non-empty sets $X_1^f, \ldots, X_t^f$ such that f_i is a multilinear polynomial from $\mathbb{F}[X_i^f]$.[5] We say that X_i^f is the set *ascribed* to f_i in the t-product polynomial f. We use $\mathrm{Vars}(f_i)$ (with a slight abuse of notation)[6] to denote X_i^f.

The following is easily implied by Lemma 3.8 in [28].

▶ **Lemma 3.** *Assume that $f \in \mathbb{F}[X]$ can be computed by a multilinear circuit over n variables of size at most s and depth at most Δ. Then, f is the sum of at most $s \cdot n$ many t-product polynomials for $t = \Omega(n^{1/2\Delta})$.*

[5] Note that we do not need f_i ($i \in [t]$) to depend non-trivially on all (or any) of the variables in X_i^f.

[6] $\mathrm{Vars}(\cdot)$ is used to describe variables ascribed to gates in a circuit as well as to denote variables ascribed to polynomials.

2.2 Partial derivative matrices and relative rank

From here on, for the sake of simplicity, we will assume that n is even[7].

Let X, W be disjoint sets of variables with $X = \{x_1, \dots, x_n\}$. Let $\mathbb{F}$ be any field. Let $\mathbb{G} = \mathbb{F}(W)$ be the field of rational functions over $\mathbb{F}$ generated by set of variables W. Let Y and Z be disjoint variable sets $\{y_1, \dots, y_n\}$ and $\{z_1, \dots, z_n\}$. We consider *injective* maps $\rho : X \to Y \cup Z$ which we call *partitioning functions.*

We can index partitioning functions by elements of $\{0,1\}^n$ as follows. Fix $\mathbf{a} \in \{0,1\}^n$ be any vector. Let $1_{\mathbf{a}} = \{i \mid \mathbf{a}(i) = 1\}$ and $0_{\mathbf{a}} = \{i \mid \mathbf{a}(i) = 0\}$, and thus, $|1_{\mathbf{a}}| + |0_{\mathbf{a}}| = n$. For a vector $\mathbf{a} \in \{0,1\}^n$, define the partitioning function $\rho_{\mathbf{a}}$ by $\rho_{\mathbf{a}}(x_i) = y_i$ if $i \in 1_{\mathbf{a}}$ and $\rho_{\mathbf{a}}(x_i) = z_i$ otherwise, i.e. if $i \in 0_{\mathbf{a}}$. Let $\mathrm{Img}(\rho_{\mathbf{a}})$ denote the subset of $Y \cup Z$ that $\rho_{\mathbf{a}}$ maps the set X to. Let $Y_{\mathbf{a}} = Y \cap \mathrm{Img}(\rho_{\mathbf{a}})$ and let $Z_{\mathbf{a}} = Z \cap \mathrm{Img}(\rho_{\mathbf{a}})$. If the vector $\mathbf{a}$ is balanced[8] then we also get that $|Y_{\mathbf{a}}| = |Z_{\mathbf{a}}| = n/2$. For a balanced vector $\mathbf{a}$, we call $\rho_{\mathbf{a}}$ a *balanced partition.*

Note that given any $\mathbf{a} \in \{0,1\}^n$ and any multilinear polynomial $f \in \mathbb{F}[X, W]$, the partitioning function $\rho_{\mathbf{a}}$ defines by substitution a *multilinear*[9] polynomial in $\mathbb{F}[Y_{\mathbf{a}} \cup Z_{\mathbf{a}} \cup W]$, which we denote $f|_{\rho_{\mathbf{a}}}$. We will consider $f|_{\rho_{\mathbf{a}}}$ as a polynomial in $\mathbb{G}[Y_{\mathbf{a}} \cup Z_{\mathbf{a}}]$.

For any disjoint sets of variables Y' and Z', let $g \in \mathbb{G}[Y' \cup Z']$ be a multilinear polynomial. Define the $2^{|Y'|} \times 2^{|Z'|}$ matrix $M_{(Y',Z')}(g)$ whose rows and columns are labelled by distinct multilinear monomials in Y' and Z' respectively and the (m_1, m_2)th entry of $M_{(Y',Z')}(g)$ is the coefficient of the monomial $m_1 \cdot m_2$ in g. We will use the rank of this matrix as a measure of the complexity of g.

We define the *relative-rank* of g w.r.t. (Y', Z'), denoted $\mathrm{relrk}_{(Y',Z')}(g)$, by

$$\mathrm{relrk}_{(Y',Z')}(g) = \frac{\mathrm{rank}(M_{(Y',Z')}(g))}{2^{(|Y'|+|Z'|)/2}}.$$

The above notion is implicit in the work of Nisan and Wigderson [20] and Raz [22].

We note the following properties of relative rank.

▶ **Proposition 4.** *Let $g, g_1, g_2 \in \mathbb{G}[Y' \cup Z']$ be multilinear polynomials.*

1. $\mathrm{relrk}_{(Y',Z')}(g) \leq 1$. *Further if $|Y'| \neq |Z'|$, then* $\mathrm{relrk}_{(Y',Z')}(g) \leq 1/\sqrt{2}$.
2. $\mathrm{relrk}_{(Y',Z')}(g_1 + g_2) \leq \mathrm{relrk}_{(Y',Z')}(g_1) + \mathrm{relrk}_{(Y',Z')}(g_2)$.
3. *If Y' is partitioned into Y'_1, Y'_2 and Z' into Z'_1, Z'_2 with $g_i \in \mathbb{G}[Y'_i \cup Z'_i]$ $(i \in [2])$, then* $\mathrm{rank}(M_{(Y',Z')}(g)) = \mathrm{rank}(M_{(Y'_1,Z'_1)}(g_1)) \cdot \mathrm{rank}(M_{(Y'_2,Z'_2)}(g_2))$. *In particular,* $\mathrm{relrk}_{(Y',Z')}(g_1 \cdot g_2) = \mathrm{relrk}_{(Y'_1,Z'_1)}(g_1) \cdot \mathrm{relrk}_{(Y'_2,Z'_2)}(g_2)$.

2.3 Explicit ε-biased spaces

The following notions are borrowed from [2]. For any $\mathbf{a}, \mathbf{b} \in \{0,1\}^n$, let $(\mathbf{a}, \mathbf{b})_2$ denote the inner product of the binary vectors $\mathbf{a}$ and $\mathbf{b}$ modulo 2, that is, $(\mathbf{a}, \mathbf{b})_2 = \sum_{i=1}^{n} \mathbf{a}(i) \cdot \mathbf{b}(i) \pmod 2$.

▶ **Definition 5.** Let $\mathcal{S}$ be a multiset in $\{0,1\}^n$. Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ be chosen uniformly from $\mathcal{S}$. The multiset $\mathcal{S}$ is said to be an ε-biased space if for every $\mathbf{b} = (b_1, b_2, \dots, b_n) \in$

[7] If n is odd, everything will work as is with n replaced by $n - 1$.

[8] A vector $\mathbf{a} \in \{0,1\}^n$ is said to be balanced if $|1_{\mathbf{a}}| = |0_{\mathbf{a}}|$.

[9] The polynomial is multilinear by the injectivity of $\rho_{\mathbf{a}}$.

$\{0,1\}^n \setminus \{0\}^n$, the random variable $(\mathbf{x},\mathbf{b})_2$ is ε-biased. That is, for all $\mathbf{b} \in \{0,1\}^n \setminus \{0\}^n$,

$$\left| \mathop{\mathbf{E}}_{\mathbf{x}\in\mathcal{S}}[(-1)^{(\mathbf{x},\mathbf{b})_2}] \right| \leq \varepsilon.$$

A standard probabilistic argument implies the existence of ε-biased spaces of size $O(n/\varepsilon^2)$. Explicit constructions of size $\mathrm{poly}(n/\varepsilon)$ were first presented by Naor and Naor in [19]. We use the following construction of Alon, Goldreich, Håstad and Peralta [2].

▶ **Theorem 6** ([2], Proposition 3). *There is a deterministic algorithm, which given as input $n \in \mathbb{N}$ and $\varepsilon > 0$, produces an ε-biased set $\mathcal{S}$ of size $O(n^2/\varepsilon^2)$ in time* $\mathrm{poly}(|\mathcal{S}|)$.

3 The hard polynomial and restrictions

3.1 Subspace-avoiding sets

▶ **Definition 7.** We say that a multiset $\mathcal{S} \subseteq \{0,1\}^n$ is (ε,k)-*subspace avoiding* if for any affine subspace V of $\{0,1\}^n$ (here identified with $\mathbb{F}_2^n$) with co-dimension k,

$$\Pr_{\mathbf{x}\in\mathcal{S}}[\mathbf{x} \in V] \leq \frac{1}{2^k} + \varepsilon.$$

The above definition is quite similar to the notion of *subspace evasive sets* that have been studied in the literature (see, e.g. [10]). However, there also seems to be a crucial difference between the two settings, since in [10] the interest is in evading subspaces of small *dimension* whereas we are trying to avoid subspaces of somewhat large but still relative small *co-dimension.* In particular, it is not clear to us if [10] can be used to give better constructions of subspace avoiding sets than the ones we obtain here.

The following fact is immediate from the definition above.

▶ **Fact 8.** *If $k = 10\log\frac{1}{\varepsilon}$ and $\mathcal{S}$ is an (ε,k)-subspace avoiding set, then $\Pr_{\mathbf{x}\in\mathcal{S}}[\mathbf{x} \in V] \leq 2\varepsilon$.*

It is a standard fact [19, 9] that ε-biased spaces are in particular (ε,k)-subspace avoiding. We state this claim below. The proof is omitted for lack of space.

▶ **Claim 9.** *Any ε-biased space $\mathcal{S}$ is also an (ε,k)-subspace avoiding set.*

We will use the vectors from an (ε,k)-subspace avoiding set to define our hard polynomial. For reasons that will become apparent, it is helpful to have the vectors in the subspace avoiding set to be *balanced*, i.e. have an equal number of 0s and 1s. However, a priori, there is no reason to assume that the vectors we obtain via some construction of such a set will be balanced. In order to make them balanced, we will use the following trick.

For $i \in [n]$, let $\boldsymbol{\alpha}_i \in \{0,1\}^n$ denote the vector defined by $\boldsymbol{\alpha}_i(j) = 0$ if $j > i$ and $\boldsymbol{\alpha}_i(j) = 1$ otherwise. Let $\mathbf{0}$ denote the all zero vector, i.e. $\mathbf{0} = 0^n$. Let $\mathcal{V} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n, \mathbf{0}\}$. Note that, for any $\mathbf{a} \in \{0,1\}^n$, there exists an $\mathbf{x} \in \mathcal{V}$ such that $\mathbf{a} \oplus \mathbf{x}$ is balanced. In particular for $\mathbf{x} \in \mathcal{V}$ chosen *at random*, the probability that $\mathbf{a} \oplus \mathbf{x}$ is balanced is at least $1/(n+1)$.

Now, let $\mathcal{S}$ be an (ε,k)-subspace avoiding set. Let $\mathcal{S} \oplus \mathbf{x}$ denote the set of all vectors obtained by shifting all the vectors in $\mathcal{S}$ by $\mathbf{x}$, i.e. $\mathcal{S} \oplus \mathbf{x} = \{\mathbf{a} \oplus \mathbf{x} \mid \mathbf{a} \in \mathcal{S}\}$. We have the following easily verifiable fact.

▶ **Fact 10.** *Let $\mathcal{S}$ be an (ε,k)-subspace avoiding set. Then for each $\mathbf{x} \in \{0,1\}^n$, $\mathcal{S} \oplus \mathbf{x}$ is also an (ε,k)-subspace avoiding set.*

For any $\mathbf{x}$, let $\mathcal{B}_{\mathbf{x}}$ be the balanced vectors in $\mathcal{S} \oplus \mathbf{x}$. From our reasoning above, we get that $\mathbf{E}_{\mathbf{x} \in \mathcal{V}}[|\mathcal{B}_{\mathbf{x}}|] \geq |\mathcal{S}|/(n+1)$. By averaging, we see that there exists an $\mathbf{x} \in \mathcal{V}$ such that $|\mathcal{B}_{\mathbf{x}}| \geq |\mathcal{S}|/(n+1)$. We will now fix such an $\mathbf{x}$. We will denote it by $\mathbf{x}_0$ and work with some $\mathcal{B} \subseteq \mathcal{B}_{\mathbf{x}_0}$ of size exactly $\lceil |\mathcal{S}|/(n+1) \rceil$. Clearly, given $\mathcal{S}$, such an $\mathbf{x}_0$ and $\mathcal{B}$ can be found in time $\mathrm{poly}(|\mathcal{S}|, n)$ by simply computing all the sets $\mathcal{S} \oplus \mathbf{x}$ ($\mathbf{x} \in \mathcal{V}$) and counting the number of balanced vectors in them.

We have shown the following claim.

▶ **Claim 11.** *Let $\mathcal{S} \subseteq \{0,1\}^n$ be any (ε, k)-subspace avoiding set. Then, there is an $\mathbf{x}_0 \in \{0,1\}^n$ and a set $\mathcal{B}$ of balanced vectors from $\mathcal{S} \oplus \mathbf{x}_0$ such that $|\mathcal{B}| = \Theta(|\mathcal{S}|/n)$. Further, given $\mathcal{S}$, such an $\mathbf{x}_0$ and $\mathcal{B}$ can be found deterministically in time* $\mathrm{poly}(|\mathcal{S}|, n)$.

3.2 The hard polynomial

We now define the explicit polynomial family that we will use to prove our size hierarchy theorem.

Let $n \in \mathbb{N}$ be any positive even integer and let $\varepsilon > 0$ be any positive real parameter. Let $\mathcal{S}_{n,\varepsilon}$ be the explicit ε-biased space from Theorem 6. We further fix an $\mathbf{x}_0 \in \{0,1\}^n$ and a set $\mathcal{B}$ of balanced vectors from $\mathcal{S}_{n,\varepsilon} \oplus \mathbf{x}_0$ such that $|\mathcal{B}| = \Theta(|\mathcal{S}_{n,\varepsilon}|/n)$ as guaranteed to exist by Claim 11. Note that by Theorem 6 and Claim 11, $|\mathcal{B}|$ can be computed in time $\mathrm{poly}(n/\varepsilon)$. We denote $|\mathcal{B}|$ by τ.

Fix any $\mathbf{a} \in \mathcal{B}$. Let $1_{\mathbf{a}} = \{i_1, \ldots, i_{n/2}\}$ and $0_{\mathbf{a}} = \{j_1, \ldots, j_{n/2}\}$, where $i_1 < i_2 < \ldots < i_{n/2}$ and $j_1 < j_2 \ldots < j_{n/2}$. Define $\Gamma_{\mathbf{a}}(X) = \prod_{t=1}^{n/2}(x_{i_t} + x_{j_t})$.

We will use $\Gamma_{\mathbf{a}}(X)$ for $\mathbf{a} \in \mathcal{B}$ to define our hard polynomial. As in [22, 24] we define such a polynomial using a set W of auxiliary variables. Intuitively, the variables W help us in *tagging* a certain polynomial $\Gamma_{\mathbf{a}}(X)$ with the appropriate vector $\mathbf{a}$ from the set $\mathcal{B}$. We will now formally describe this idea.

Since $|\mathcal{B}| = \tau$, we can fix a one-one map $\mathcal{C} : \mathcal{B} \to \{0,1\}^{\log \tau}$. Let the vectors in $\mathcal{B}$ be enumerated in some arbitrary order, say $\mathbf{a}_1, \ldots, \mathbf{a}_\tau$. For an index $i \in [\tau]$, let $\mathcal{C}(\mathbf{a}_i) = (u_{i,1} \ldots, u_{i,T})$. We will denote this vector by $\mathbf{u}_i$, we will call $\mathbf{u}_i$ the *encoding* of $\mathbf{a}_i$.

Let $T = \log \tau$. Let $W = \{w_1, \ldots, w_T\}$ be a new set of auxiliary variables. For a vector $\mathbf{u} \in \{0,1\}^T$, let $\phi_{\mathbf{u}}(j) = w_j$ if $u_j = 1$ and $\phi_{\mathbf{u}}(j) = 1 - w_j$ otherwise. Then let $W_{\mathbf{u}} = \prod_{j \in [T]} \phi_{\mathbf{u}}(j)$. We will say that a polynomial $W_{\mathbf{u}}$ is the *label* of the vector $\mathbf{u}$. We will say that the tagging of the polynomial $\Gamma_{\mathbf{a}}(X)$ is obtained by multiplying $\Gamma_{\mathbf{a}}(X)$ with the label of the encoding of $\mathbf{a}$, i.e. with $W_{\mathcal{C}(\mathbf{a})}$. Note that, given $\mathbf{a} \in \mathcal{B}$, the polynomial $W_{\mathcal{C}(a)} \cdot \Gamma_{\mathbf{a}}(X)$ can be computed by a depth-2 multilinear formula, which itself can be constructed in time $\mathrm{poly}(n, \tau) = \mathrm{poly}(n/\varepsilon)$.

We are now ready to define our hard polynomial.

$$P_{n,\varepsilon}(X, W) = \sum_{\mathbf{a} \in \mathcal{B}} W_{\mathcal{C}(\mathbf{a})} \cdot \Gamma_{\mathbf{a}}(X).$$

We have the following.

▶ **Lemma 12.** *The polynomial $P_{n,\varepsilon}(X, W)$ can be computed by a depth 3 multilinear formula $F_{n,\varepsilon}$ of size $s = O(\tau(n + 2\log \tau)) = O(n^2/\varepsilon^2 + (n/\varepsilon^2)\log(n/\varepsilon))$ that can be constructed in time* $\mathrm{poly}(s)$. *Further, there is a deterministic* $\mathrm{poly}(s)$*-time algorithm that, given a multilinear monomial m over the variables $X \cup W$, computes the coefficient of m in $P_{n,\varepsilon}(X, W)$.*

Proof. Everything but the last statement is immediate from the preceding discussion. To prove the final statement, it suffices to note that the coefficient of m can be found in each constituent depth-2 formula in time $\mathrm{poly}(n) = \mathrm{poly}(s)$. Summing these coefficients yields the coefficient of m in $P_{n,\varepsilon}(X, W)$. ◀

Another property of our hard polynomial that is true by construction is the following.

▶ **Lemma 13.** *For any $n \in \mathbb{N}$ and $\varepsilon > 0$, let $P_{n,\varepsilon}(X, W) \in \mathbb{F}[X, W]$ be the polynomial defined above, which we will consider as a polynomial from $\mathbb{G}[X]$ where $\mathbb{G} = \mathbb{F}(W)$. For any $\mathbf{a} \in \mathcal{B}$, let restriction $\rho_{\mathbf{a}} : X \to Y \cup Z$ be the restriction as defined in Section 2.2. Then* $\mathrm{relrk}_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(P_{n,\varepsilon}(X, W)|_{\rho_{\mathbf{a}}}) = 1$.

Proof. Fix any $\mathbf{a} \in \mathcal{B}$ and consider the balanced partition $\rho_{\mathbf{a}} : X \to Y_{\mathbf{a}} \cup Z_{\mathbf{a}}$. We analyze the partial derivative matrix $M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(P_{n,\varepsilon}(X, W)|_{\rho_{\mathbf{a}}})$ whose entries are polynomials over the variables in W. To show that $\mathrm{relrk}_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(P_{n,\varepsilon}(X, W)|_{\rho_{\mathbf{a}}}) = 1$, we need to show that $M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(P_{n,\varepsilon}(X, W)|_{\rho_{\mathbf{a}}})$ is a full rank matrix over $\mathbb{G}$. Towards that, it is sufficient to show that $\det(M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(P_{n,\varepsilon}(X, W)|_{\rho_{\mathbf{a}}}))$ is a non-zero polynomial over the variables in W. Further, it is enough to show that there is an assignment $A : W \to \{0, 1\}$ to the W-variables such that $\det(M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(P_{n,\varepsilon}(X, A(W))|_{\rho_{\mathbf{a}}}))$ evaluates to a non-zero value. This is what we will do. (A similar proof strategy is used in the proof of Claim 4.6 in [3].)

Let the vector $\mathbf{u} = \mathcal{C}(\mathbf{a})$. For all $i \in [\log \tau]$, A sets the variable w_i to 1 if $u_i = 1$ and 0 otherwise. Now, it is easy to see that $P_{n,\varepsilon}(X, A(W)) = \Gamma_{\mathbf{a}}(X)$. This also implies that

$$\det(M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(P_{n,\varepsilon}(X, A(W))|_{\rho_{\mathbf{a}}})) = \det(M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(\Gamma_{\mathbf{a}}(X)|_{\rho_{\mathbf{a}}}))$$

Now it is easy to check that $M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(\Gamma_{\mathbf{a}}(X)|_{\rho_{\mathbf{a}}})$ is a permutation matrix and hence $\det(M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(\Gamma_{\mathbf{a}}(X)|_{\rho_{\mathbf{a}}}))$ is non-zero. This implies that $\det(M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(P_{n,\varepsilon}(X, A(W))|_{\rho_{\mathbf{a}}}))$ is non-zero as well. Thus, $\det(M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(P_{n,\varepsilon}(X, W)|_{\rho_{\mathbf{a}}}))$ is a non-zero polynomial over the variables in W and we get that $M_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(P_{n,\varepsilon}(X, W)|_{\rho_{\mathbf{a}}})$ is a full rank matrix. ◀

4 The lower bound

In this section, we show that for f, a t-product polynomial, and $\mathbf{a}$ chosen randomly from an $(\varepsilon, 10\log(1/\varepsilon))$-subspace avoiding set, the polynomial $f|_{\rho_{\mathbf{a}}}$ has low relative-rank with high probability. We then use this to prove the main theorem.

▶ **Lemma 14.** *Let $s \in \mathbb{N}$ and $\varepsilon > 0$ be parameters such that $s \geq 1/\varepsilon$. Let $f \in \mathbb{G}[X]$ be a t-product polynomial with $t \geq (\log s)^3$. Let $\mathcal{S}_0$ be any $(\varepsilon, 10\log\frac{1}{\varepsilon})$-subspace avoiding set defined in Section 3.2. For any $\mathbf{a} \in \{0, 1\}^n$, let $\rho_{\mathbf{a}}$ denote the partitioning function defined in Section 2.2. Then,*

$$\Pr_{\mathbf{a}\in\mathcal{S}_0}[\mathrm{relrk}_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(f|_{\rho_{\mathbf{a}}}) \geq \frac{1}{s}] \leq 5 \cdot \varepsilon.$$

Proof of Lemma 14. For all $k \in [t]$, let $\mathbf{r}^{(k)} = (r_1^{(k)}, r_2^{(k)}, \ldots, r_n^{(k)})$ be a vector in $\{0, 1\}^n$ such that

$$r_i^{(k)} = \begin{cases} 1 & \text{if } x_i \in \mathrm{Vars}(f_k), \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathcal{E}_k(\mathbf{a})$ be a 0-1 random variable defined as follows. If $|\mathrm{Vars}(f_k)|$ is odd, $\mathcal{E}_k(\mathbf{a})$ is always 0. Otherwise, define $\beta_k = |\mathrm{Vars}(f_k)|/2 \pmod 2$ and

$$\mathcal{E}_k(\mathbf{a}) = \begin{cases} 1 & \text{if } (\mathbf{r}^{(k)}, \mathbf{a})_2 = \beta_k, \\ 0 & \text{otherwise.} \end{cases}$$

The main step in the proof of Lemma 14, is the following claim.

▶ **Claim 15.** *If $\mathcal{S}_0$ is a $(\varepsilon, 10\log\frac{1}{\varepsilon})$-subspace avoiding set, then* $\Pr_{\mathbf{a}\in\mathcal{S}_0}[\sum_{k\in[t]}\mathcal{E}_k(\mathbf{a}) \geq t - 2\log s] \leq 5\cdot\varepsilon$.

First, we will prove Lemma 14 using this claim. In order to do that, we will show that for all $\mathbf{a}\in\mathcal{S}_0$ the condition that $(\sum_{k\in[t]}\mathcal{E}_k(\mathbf{a}) < t - 2\log s$ implies $\mathrm{relrk}_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(f|_{\rho_{\mathbf{a}}}) \leq \frac{1}{s}$.

For any vector $\mathbf{a}\in\mathcal{S}_0$ and for $k\in[t]$, let $Y_{k,\mathbf{a}} = \rho_{\mathbf{a}}(\mathrm{Vars}(f_k))\cap Y_{\mathbf{a}}$ and let $Z_{k,\mathbf{a}} = \rho_{\mathbf{a}}(\mathrm{Vars}(f_k))\cap Z_{\mathbf{a}}$. By Item 3 in Proposition 4 we know that

$$\mathrm{relrk}_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(f|_{\rho_{\mathbf{a}}}) = \prod_{k\in[t]}\mathrm{relrk}_{(Y_{k,\mathbf{a}},Z_{k,\mathbf{a}})}(f_k|_{\rho_{\mathbf{a}}}).$$

We claim that for any $k\in[t]$, if $\mathcal{E}_k(\mathbf{a}) = 0$, then $\mathrm{relrk}_{(Y_{k,\mathbf{a}},Z_{k,\mathbf{a}})}(f_k|_{\rho_{\mathbf{a}}}) \leq 1/\sqrt{2}$. To see this, note that, for any $k\in[t]$, $\mathcal{E}_k(\mathbf{a}) = 1$ if and only if $|\mathrm{Vars}(f_k)|$ is even and $(\mathbf{r}^{(k)},\mathbf{a})_2 = \beta_k$. Now, if $|\mathrm{Vars}(f_k)|$ is odd then we can never have $|Y_{k,\mathbf{a}}| = |Z_{k,\mathbf{a}}|$ and hence by Item 1 of Proposition 4, $\mathrm{relrk}_{(Y_{k,\mathbf{a}},Z_{k,\mathbf{a}})}(f_k|_{\rho_{\mathbf{a}}}) \leq 1/\sqrt{2}$. Further, if $|\mathrm{Vars}(f_k)|$ is even and $|Y_{k,\mathbf{a}}| - |\mathrm{Vars}(f_k)|/2 \neq 0 \pmod 2$, then again we must have $|Y_{k,\mathbf{a}}| \neq |Z_{k,\mathbf{a}}|$ and hence $\mathrm{relrk}_{(Y_{k,\mathbf{a}},Z_{k,\mathbf{a}})}(f_k|_{\rho_{\mathbf{a}}}) \leq 1/\sqrt{2}$.

Now if $\sum_{k\in[t]}\mathcal{E}_k(\mathbf{a}) \leq t - 2\log s$ then there exist at least $(2\log s)$ elements $k\in[t]$ such that $\mathcal{E}_k(\mathbf{a}) = 0$. Hence, $\mathrm{relrk}_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(f|_{\rho_{\mathbf{a}}}) \leq (1/\sqrt{2})^{2\log s} \leq 1/s$.

Thus, in order to upper bound the probability of the event that $\mathrm{relrk}_{(Y_{\mathbf{a}},Z_{\mathbf{a}})}(f|_{\rho_{\mathbf{a}}}) \geq 1/s$, it suffices to upper bound the probability of $\sum_{k\in[t]}\mathcal{E}_k(\mathbf{a}) \geq t - 2\log s$, which by Claim 15 is at most 5ε. This concludes the proof of Lemma 14.

It remains to prove Claim 15, which we do now. The proof follows a combinatorial proof of the Chernoff bound due to Impagliazzo and Kabanets [16].

Proof of Claim 15. Let $\ell = t - 2\log s$ and $\mathcal{E}(\mathbf{a}) = \sum_{k\in[t]}\mathcal{E}_k(\mathbf{a})$. Let $R(\mathbf{a})$ be a Boolean random variable such that

$$R(\mathbf{a}) = \begin{cases} 1 & \text{if } \sum_{k\in[t]}\mathcal{E}_k(\mathbf{a}) \geq \ell, \\ 0 & \text{otherwise.}\end{cases}$$

Thus, $\Pr_{\mathbf{a}\in\mathcal{S}}[\mathcal{E}(\mathbf{a}) \geq \ell] = \mathbf{E}_{\mathbf{a}\in\mathcal{S}}[R(\mathbf{a})]$. Fix a vector $\mathbf{a}$. Let $\tilde{R}(\mathbf{a}) \in [0,1]$ be the random variable defined by $\tilde{R}(\mathbf{a}) = \mathbf{E}_A[\prod_{i\in A}\mathcal{E}_i(\mathbf{a})]$, where $A\subseteq[t]$ is an independently and uniformly randomly chosen subset of size $2\log s$.

We claim that for every $\mathbf{a}\in\mathcal{S}_0$, $R(\mathbf{a}) \leq 2\cdot\tilde{R}(\mathbf{a})$. Assuming this, we get the following.

$$\begin{aligned}\Pr_{\mathbf{a}}[\sum_{k\in[t]}\mathcal{E}_k(\mathbf{a}) \geq \ell] &= \mathop{\mathbf{E}}_{\mathbf{a}}[R(\mathbf{a})] \leq 2\mathop{\mathbf{E}}_{\mathbf{a}}[\mathop{\mathbf{E}}_{A}[\prod_{i\in A}\mathcal{E}_i(\mathbf{a})]] \\ &= 2\mathop{\mathbf{E}}_{A}[\mathop{\mathbf{E}}_{\mathbf{a}}[\prod_{i\in A}\mathcal{E}_i(\mathbf{a})]] = 2\mathop{\mathbf{E}}_{A}[\Pr_{\mathbf{a}}[\prod_{i\in A}\mathcal{E}_i(\mathbf{a}) = 1]]. \end{aligned} \tag{3}$$

Consider an individual term $\mathcal{E}_A(\mathbf{a}) := \prod_{i\in A}\mathcal{E}_i(\mathbf{a})$ in the above expression. We claim that $\mathbf{E}_{\mathbf{a}}[\mathcal{E}_A(\mathbf{a})] \leq 2\varepsilon$. To see this, note that we have one of the following two scenarios. Either A contains an i such that $|\mathrm{Vars}(f_i)|$ is odd, in which case $\mathcal{E}_A(\mathbf{a}) = \mathcal{E}_i(\mathbf{a}) = 0$ with probability 1. Otherwise, $|\mathrm{Vars}(f_i)|$ is even for each $i\in A$ and then $\mathbf{a}$ satisfies $\mathcal{E}_A(\mathbf{a}) = 1$ if and only if $\mathbf{a}$ satisfies the system of linear equations $\{(\mathbf{a},\mathbf{r}^{(i)})_2 = \beta_i \mid i\in A\}$. Since the $\mathbf{r}^{(i)}$'s are non-zero and linearly independent, this system of equations defines an affine subspace of codimension $|A| = 2\log s$. Now, by invoking Claim 9, we get that $\Pr_{\mathbf{a}}[\prod_{i\in A}\mathcal{E}_i(\mathbf{a}) = 1] \leq 1/s^2 + \varepsilon \leq 2\varepsilon$, where the final inequality uses $s \geq 1/\varepsilon$. Substituting this back in (3), we get that $\Pr_{\mathbf{a}}[\mathcal{E}(\mathbf{a}) \geq \ell] \leq 4\varepsilon$.

To complete the proof, we need to show that for every $\mathbf{a} \in \mathcal{S}_0$, $R(\mathbf{a}) \leq 2\tilde{R}(\mathbf{a})$. If $R(\mathbf{a}) = 0$, then this statement is trivial. If not, $R(\mathbf{a}) = 1$. That is, there exist at least ℓ many $k \in [t]$ such that $\mathcal{E}_k(\mathbf{a}) = 1$. Then,

$$\tilde{R}(\mathbf{a}) = \mathop{\mathbf{E}}_{A}[\prod_{i \in A} \mathcal{E}_i(\mathbf{a})] = \Pr_{A}[\text{for all } i \in A,\ \mathcal{E}_i(\mathbf{a}) = 1] \geq \frac{\binom{\ell}{2\log s}}{\binom{t}{2\log s}} = \frac{\binom{t-2\log s}{2\log s}}{\binom{t}{2\log s}} \geq 1/2,$$

where the final inequality follows from the fact that $t \geq (\log s)^3$. This completes the proof of Lemma 14. ◀

The main theorem

We now prove our main theorem. It is restated here for the sake of convenience.

▶ **Theorem 16.** *For any fixed positive $c \in \mathbb{N}$ and $\delta \in (0, 1/2)$, there exists an explicit polynomial family $P_n \in \mathbb{F}[x_1, \ldots, x_n]$ that has multilinear formulas of depth 3 and size at most $O(s)$ where $s = n^c$, but no multilinear formulas of size less than $s^{(1/2)-\delta}$ and depth $\Delta < \log n/100 \log\log n$.*

Proof. We first show that we can assume without loss of generality that $c \geq 10/\delta$. Say this is not the case: we then have $s = n^c < n^{10/\delta}$. Now, let $m = s^{\delta/10} \leq n$. We will then define a polynomial over only the variables $\{x_1, \ldots, x_m\}$. So the number of variables reduces to m and $s = m^{10/\delta}$ (in particular, the new value of c is now $10/\delta$). Thus, we can always reduce the problem to the case when $c \geq 10/\delta$.

So we assume without loss of generality that $s \geq n^{10/\delta}$. We will fix our polynomial P_n to be $P_{n/2,\varepsilon}(X, W)$ as defined in Section 3.2 for $\varepsilon = n/\sqrt{s}$, where the variable set $X = \{x_1, \cdots, x_{n/2}\}$ and $W \subseteq \{x_{n/2+1}, \ldots, x_n\}$ is some fixed set of size $\log \tau \leq \log s$ (since $s \leq n^c$, it is clear that $\log s \leq n/2$). From Lemma 12, we know that $\{P_n \mid n \in \mathbb{N}\}$ is an explicit family of multilinear polynomials such that each P_n is computed by a depth three multilinear formula of size $O(s)$. Further, from Lemma 13, we get that $\mathrm{relrk}_{(Y_{\mathbf{a}}, Z_{\mathbf{a}})}(P_n(X, W)|_{\rho_{\mathbf{a}}}) = 1$ for every choice of $\mathbf{a} \in \mathcal{B}$, where $\mathcal{B}$ is as defined in Section 3.2.

Let us assume that P_n can be computed by a depth Δ multilinear formula Φ of size $s' < s^{1/2-\delta}$. We consider $P_n(X, W)$ as a polynomial from the ring $\mathbb{G}[X]$ where $\mathbb{G} = \mathbb{F}(W)$. We also consider Φ as a multilinear formula computing a polynomial from $\mathbb{G}[X]$ (i.e. we consider the variables from W in Φ as constants from the underlying field $\mathbb{G}$.)

From Lemma 3, we know that P_n can be written as a sum of at most $s'' = s' \cdot n$ many t-product polynomials where $t = \Omega(n^{1/2\Delta})$. Note that as $\Delta < \log n/100 \log\log n$, we have $t = \Omega((\log n)^{50}) \geq (\log s)^{40}$, the latter inequality following from the fact that $\log s = O(\log n)$.

Assume that the t-product polynomials in the above decomposition are $f_1, \ldots, f_{s''} \in \mathbb{G}[X]$. So, we have $P_n = \sum_{i \in [s'']} f_i$. We would like to show that

$$\Pr_{\mathbf{a} \in \mathcal{B}}[\mathrm{relrk}_{(Y_{\mathbf{a}}, Z_{\mathbf{a}})}(P_n|_{\rho_{\mathbf{a}}}) < 1] > 0 \tag{4}$$

which would contradict Lemma 13 and hence prove the theorem.

In order to prove inequality (4), by the sub-additivity property (Proposition 4 Item 2) of the relative-rank, it suffices to show the following.

$$\Pr_{\mathbf{a} \in \mathcal{B}}[\forall i \in [s'], \mathrm{relrk}_{(Y_{\mathbf{a}}, Z_{\mathbf{a}})}(f_i|_{\rho_{\mathbf{a}}}) < 1/s''] > 0.$$

Equivalently, it suffices to prove that

$$\Pr_{\mathbf{a} \in \mathcal{B}}[\exists i \in [s''], \mathrm{relrk}_{(Y_{\mathbf{a}}, Z_{\mathbf{a}})}(f_i|_{\rho_{\mathbf{a}}}) \geq 1/s''] < 1. \tag{5}$$

Recall from Section 3.2 that the set $\mathcal{B}$ is defined to be a subset of the set $\mathcal{S}_0 = \mathcal{S}_{n,\varepsilon} \oplus \mathbf{x}_0$ of size $\tau = \Theta(|\mathcal{S}_0|/n)$. Also, by Claim 9 and Fact 10, it follows that $\mathcal{S}_0$ is an $(\varepsilon, 10\log(1/\varepsilon))$-biased set. By using Lemma 14, we get that for each t-product polynomial f_i, we have that

$$\Pr_{\mathbf{a}\in\mathcal{S}_0}[\mathrm{relrk}_{(Y_\mathbf{a},Z_\mathbf{a})}(f_i|_{\rho_\mathbf{a}}) \geq \frac{1}{s}] \leq 5\varepsilon.$$

Therefore, by a simple union bound, we get that

$$\Pr_{\mathbf{a}\in\mathcal{S}_0}[\exists i : \mathrm{relrk}_{(Y_\mathbf{a},Z_\mathbf{a})}(f_i|_{\rho_\mathbf{a}}) \geq \frac{1}{s}] \leq 5s'' \cdot \varepsilon. \tag{6}$$

As $\mathcal{B}$ is a subset of $\mathcal{S}_0$, we get that

$$\frac{|\mathcal{B}|}{|\mathcal{S}_0|} \cdot \Pr_{\mathbf{a}\in\mathcal{B}}[\exists i : \mathrm{relrk}_{(Y_\mathbf{a},Z_\mathbf{a})}(f_i|_{\rho_\mathbf{a}}) \geq \frac{1}{s}] \leq \Pr_{\mathbf{a}\in\mathcal{S}_0}[\exists i : \mathrm{relrk}_{(Y_\mathbf{a},Z_\mathbf{a})}(f_i|_{\rho_\mathbf{a}}) \geq \frac{1}{s}]. \tag{7}$$

Hence, using the inequalities (6), (7) and the fact that $|\mathcal{B}|/|\mathcal{S}_0| = \Theta(1/n)$, we get

$$\Pr_{\mathbf{a}\in\mathcal{B}}[\exists i : \mathrm{relrk}_{(Y_\mathbf{a},Z_\mathbf{a})}(f_i|_{\rho_\mathbf{a}}) \geq \frac{1}{s}] \leq O\left(\frac{s''n^2}{\sqrt{s}}\right) \leq \frac{s'n^4}{\sqrt{s}}. \tag{8}$$

As $s \geq n^{10/\delta}$ we get $n^4 \leq s^{\delta/2}$. Therefore, as long as $s' \leq O(s^{1/2-\delta})$, inequality (5) is satisfied, which then implies that inequality (4) is also satisfied.

If inequality (4) is satisfied, then there exists a partitioning function $\rho_\mathbf{a}$ for $\mathbf{a} \in \mathcal{B}$ such that $\mathrm{relrk}_{(Y_\mathbf{a},Z_\mathbf{a})}(P_n|_{\rho_\mathbf{a}}) < 1$. This contradicts Lemma 13 which tells us $\mathrm{relrk}_{(Y_\mathbf{a},Z_\mathbf{a})}(P_n|_{\rho_\mathbf{a}}) = 1$ with respect to every partitioning function $\rho_\mathbf{a}$ ($\mathbf{a} \in \mathcal{B}$). ◀

References

1 M. Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1–48, 1983. doi:10.1016/0168-0072(83)90038-6.

2 Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple construction of almost k-wise independent random variables. *Random Struct. Algorithms*, 3(3):289–304, 1992. doi:10.1002/rsa.3240030308.

3 Noga Alon, Mrinal Kumar, and Ben Lee Volk. Unbalancing sets and an almost quadratic lower bound for syntactically multilinear arithmetic circuits, 2017. URL: https://mrinalkr.bitbucket.io/papers/synt-multilinear.pdf.

4 Kazuyuki Amano. *k*-subgraph isomorphism on AC^0 circuits. *Computational Complexity*, 19(2):183–210, 2010. doi:10.1007/s00037-010-0288-y.

5 Maria Luisa Bonet and Samuel R. Buss. Size-depth tradeoffs for boolean formulae. *Information Processing Letters*, 49(3):151–155, 1994. doi:10.1016/0020-0190(94)90093-0.

6 Richard P. Brent. The parallel evaluation of general arithmetic expressions. *Journal of the ACM*, 21(2):201–206, 1974. doi:10.1145/321812.321815.

7 Suryajith Chillara, Christian Engels, Nutan Limaye, and Srikanth Srinivasan. A Near-Optimal Depth-Hierarchy Theorem for Small-Depth Multilinear Circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:066, 2018. URL: https://eccc.weizmann.ac.il/report/2018/066/s.

8 Suryajith Chillara, Nutan Limaye, and Srikanth Srinivasan. Small-depth multilinear formula lower bounds for iterated matrix multiplication, with applications. In *STACS*, volume 96 of *LIPIcs*, pages 21:1–21:15. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018.

9 Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. In *APPROX-RANDOM*, volume 6302 of *Lecture Notes in Computer Science*, pages 504–517. Springer, 2010.

10 Zeev Dvir and Shachar Lovett. Subspace evasive sets. In *STOC*, pages 351–358. ACM, 2012.

11 Zeev Dvir, Guillaume Malod, Sylvain Perifel, and Amir Yehudayoff. Separating multilinear branching programs and formulas. In *proceedings of Symposium on Theory of Computing (STOC)*, pages 615–624, 2012. `doi:10.1145/2213977.2214034`.

12 Merrick Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical systems theory*, 17(1):13–27, Dec 1984. `doi:10.1007/BF01744431`.

13 J. Hartmanis and R. E. Stearns. On the computational complexity of algorithms. *Transactions of the American Mathematical Society*, 117:285–306, 1965. URL: `http://www.jstor.org/stable/1994208`.

14 J. Håstad. *Computational limitations of small-depth circuits*. ACM doctoral dissertation award. MIT Press, 1987. URL: `https://books.google.co.in/books?id=_h0ZAQAAIAAJ`.

15 Pavel Hrubeš and Amir Yehudayoff. Homogeneous formulas and symmetric polynomials. *Computational Complexity*, 20(3):559–578, 2011.

16 Russell Impagliazzo and Valentine Kabanets. Constructive proofs of concentration bounds. In *APPROX-RANDOM*, volume 6302 of *Lecture Notes in Computer Science*, pages 617–631. Springer, 2010.

17 Neeraj Kayal, Vineet Nair, and Chandan Saha. Separation between read-once oblivious algebraic branching programs (ROABPs) and multilinear depth three circuits. In *proceedings of Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 46:1–46:15, 2016. `doi:10.4230/LIPIcs.STACS.2016.46`.

18 P. M. Lewis, R. E. Stearns, and J. Hartmanis. Hierarchies of memory limited computations. In *6th Annual Symposium on Switching Circuit Theory and Logical Design (SWCT 1965)(FOCS)*, volume 00, pages 179–190, 10 1965. `doi:10.1109/FOCS.1965.11`.

19 Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993. `doi:10.1137/0222053`.

20 Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997. `doi:10.1007/BF01294256`.

21 Ran Raz. Multilinear-$\mathsf{NC}^2 \neq$ multilinear-NC^1. In *proceedings of Foundations of Computer Science (FOCS)*, pages 344–351, 2004. `doi:10.1109/FOCS.2004.42`.

22 Ran Raz. Separation of multilinear circuit and formula size. *Theory of Computing*, 2(1):121–135, 2006. `doi:10.4086/toc.2006.v002a006`.

23 Ran Raz, Amir Shpilka, and Amir Yehudayoff. A lower bound for the size of syntactically multilinear arithmetic circuits. *SIAM Journal of Computing*, 38(4):1624–1647, 2008. `doi:10.1137/070707932`.

24 Ran Raz and Amir Yehudayoff. Balancing syntactically multilinear arithmetic circuits. *Computational Complexity*, 17(4):515–535, 2008. `doi:10.1007/s00037-008-0254-0`.

25 Ran Raz and Amir Yehudayoff. Lower bounds and separations for constant depth multilinear circuits. *Computational Complexity*, 18(2):171–207, 2009. `doi:10.1007/s00037-009-0270-8`.

26 Benjamin Rossman. *Average-case Complexity of Detecting Cliques*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2010. AAI0823246.

27 Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity. Github survey, 2015. URL: `https://github.com/dasarpmar/lowerbounds-survey/releases/`.

28 Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5:207–388, March 2010. `doi:10.1561/0400000039`.