

Strongly Exponential Separation Between Monotone VP and Monotone VNP

Srikanth Srinivasan*
Department of Mathematics
IIT Bombay

July 31, 2020

Abstract

We show that there is a sequence of explicit multilinear polynomials $P_n(x_1, \dots, x_n) \in \mathbb{R}[x_1, \dots, x_n]$ with non-negative coefficients that lies in monotone VNP such that any monotone algebraic circuit for P_n must have size $\exp(\Omega(n))$. This builds on (and strengthens) a result of Yehudayoff (2018) who showed a lower bound of $\exp(\tilde{\Omega}(\sqrt{n}))$.

1 Introduction

This paper deals with a problem in *Algebraic Complexity*, which is the study of the complexity of computing multivariate polynomials over some underlying field \mathbb{F} . The model of computation is the *Algebraic circuit* model, which computes polynomials from $\mathbb{F}[x_1, \dots, x_n]$ using the basic sum and product operations in this ring. This model and its variants have been studied by a large body of work (see, e.g. the surveys [18, 15]).

The central question in the area is Valiant's [19] VP vs. VNP question. The set VP contains sequences $(P_n(x_1, \dots, x_n))_{n \geq 1}$ of polynomials of polynomially bounded degree¹ that can be computed by polynomial-sized algebraic circuits. The class VNP contains sequences $(Q_n(x_1, \dots, x_n))_{n \geq 1}$ where

$$Q_n(x_1, \dots, x_n) = \sum_{b_1, \dots, b_m \in \{0,1\}} P_{n+m}(x_1, \dots, x_n, b_1, \dots, b_m)$$

where m is polynomially bounded in n and $(P_r(x_1, \dots, x_r))_{r \geq 1}$ is in VP.

Like its Boolean analogue, the VP vs. VNP question has proved stubbornly hard to resolve, the principal bottleneck being our inability to prove explicit algebraic circuit lower bounds. Given this, it is natural to look at variants of this question.

*Email: srikanth@math.iitb.ac.in

¹i.e. $\deg(P_n) \leq n^{O(1)}$

In a recent paper [21], Yehudayoff considered the *monotone* version of the VP vs. VNP question, which is defined as follows. The underlying field is \mathbb{R} and the polynomials being computed have non-negative coefficients. A *monotone* algebraic circuit is one where all the constants appearing in the circuit are non-negative. The monotone versions of VP and VNP, denoted MVP and MVNP respectively, are defined analogously: MVP contains (sequences of) polynomials that have small *monotone* algebraic circuits; MVNP contains (sequences of) polynomials that can be written as exponential Boolean sums over polynomials in MVP.

Monotone algebraic circuits have been studied since the 80s, and explicit exponential lower bounds are known for this model via the work of Schnorr [16] and Jerrum and Snir [9] (see also [20, 17, 6, 13]). However, as Yehudayoff [21] pointed out, these results do not imply a separation between MVP and MVNP. In fact, most² of the monotone circuit lower bounds proved in earlier work also imply that the same polynomials do not belong to MVNP, and hence do not imply a separation between these two classes.

The main result of [21] was the resolution of the MVP vs. MVNP question. More precisely, Yehudayoff showed that there is an explicit sequence of multilinear polynomials $(P_n(x_1, \dots, x_n))_{n \geq 1}$ in MVNP such that any monotone algebraic circuit for P_n must have size $\exp(\tilde{\Omega}(\sqrt{n}))$.

In this paper, we strengthen this result to a strongly exponential lower bound.

Theorem 1. *There is an explicit sequence of multilinear polynomials $(P_n(x_1, \dots, x_n))_{n \geq 1}$ in MVNP such that any monotone algebraic circuit for P_n must have size $2^{\Omega(n)}$.*

This theorem bears a similar relation to Yehudayoff's result as some later works [6, 13] bears to the result of Schnorr [16]. Schnorr [16] proved a lower bound of $\exp(\Omega(\sqrt{n}))$ for an explicit family of polynomials; a similar lower bound was also proved for an explicit family of polynomials by Jerrum and Snir [9].³ These bounds were strengthened to strongly exponential lower bounds by a series of works of Kuznetsov, Kasim-Zade, and Gashkov in the USSR in the 80s [11, 5, 6]⁴, and independently by a more recent result of Raz and Yehudayoff [13].

1.1 Proof Outline

High level idea. We rely on a connection between monotone algebraic circuit lower bounds and communication complexity that was made explicit by Raz and Yehudayoff [13]. As shown in [13], if a multilinear polynomial $P \in \mathbb{R}[x_1, \dots, x_n]$ has a monotone algebraic circuit of size s , then we get a decomposition

$$P = \sum_{i=1}^s g_i h_i \tag{1}$$

²The one exception to this seems to be a lower bound of Raz and Yehudayoff [13]. Here, it is unclear whether the hard polynomials lie in MVNP but we are unable to rule it out.

³These explicit polynomials were based on the Clique and the Permanent respectively.

⁴Unfortunately, journal versions of these papers are not easily available, but we refer to a survey of Gashkov and Sergeev [6] for a very interesting account of this line of work, along with details of some of these results.

where each summand $g_i h_i$ satisfies the property that g_i and h_i are *non-negative* multilinear polynomials that depend on disjoint sets of at least $n/3$ variables each. We call each such term a *non-negative product polynomial*. Thus, to prove a lower bound on the circuit complexity of P , it suffices to lower bound the number of terms in any decomposition as in (1).

As noted by Jerrum and Snir [9], one way to do this is via the *support of the polynomial* P , by which we mean the set of monomials that have non-zero coefficients in P . We think of this set, denoted $\text{Supp}(P)$, as a subset of $2^{[n]}$ by identifying each multilinear monomial on x_1, \dots, x_n with a subset of $[n]$ in the natural way. Given a decomposition of P into non-negative product polynomials as in (1), we immediately get $\text{Supp}(P) = \bigcup_{i \in [s]} \text{Supp}(g_i \cdot h_i)$. And so it suffices to obtain a P such that any such decomposition of $\text{Supp}(P)$ must have large size.

Such decompositions are closely related to a model of communication complexity known as *Multipartition Communication Complexity*, introduced by Āuris, Hromkoviĉ, Jukna, Sauerhoff and Schnitger [4] (see also the earlier result of Borodin, Razborov and Smolensky [2]). The multipartition communication complexity of a subset $\mathcal{S} \subseteq 2^{[n]}$ (or equivalently a Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$) is defined as follows. We define a *rectangle* $\mathcal{R} \subseteq 2^{[n]}$ to be any set of the form $\{A \cup B \mid A \in \mathcal{A}, B \in \mathcal{B}\}$, where $\mathcal{A} \subseteq 2^Y$ and $\mathcal{B} \subseteq 2^Z$ and (Y, Z) is a partition of $[n]$. Further, we say that both the partition and the rectangle \mathcal{R} are *balanced* if $|Y|, |Z| \geq n/3$. Finally, the multipartition communication complexity of \mathcal{S} is defined to be $\lceil \log_2 k \rceil$ where k is the smallest integer such that \mathcal{S} can be decomposed as the union of k many balanced rectangles.

To see the connection to algebraic complexity, note that if $P \in \mathbb{R}[x_1, \dots, x_n]$ has monotone algebraic circuits of size s , then (1) implies that $\text{Supp}(P)$ has multipartition communication complexity at most $\lceil \log_2 s \rceil$. In particular, linear lower bounds in this model for some explicit \mathcal{S} implies that any non-negative polynomial P with support exactly \mathcal{S} cannot be computed by monotone algebraic circuits of subexponential size.

Polynomial (but sublinear) lower bounds for multipartition communication complexity were implicit in the work of Borodin et al. [2] and were extended to linear (but somewhat non-explicit) lower bounds in the work of Āuris et al. [4]. An explicit linear lower bound for this model is implicit in a result of Bova, Capelli, Mengel and Slivovsky [3]. (See also the related work of Hayes [7]. Similar constructions are attributed to Wigderson in [13] and carried out by Jukna [10].) The hard problem of [3] is quite easy to describe. Fix a regular expander graph⁵ G on vertex set $[n]$ with constant degree d . The associated hard problem is given by taking \mathcal{S} to be the set of all vertex covers in G . Said differently, we consider the Boolean function $f_G(x_1, \dots, x_n) = \bigwedge_{\{i,j\} \in E(G)} (x_i \vee x_j)$.

As mentioned above, the communication complexity lower bound on \mathcal{S} immediately yields a strongly exponential lower bound on the monotone algebraic complexity of some explicitly defined polynomial. Unfortunately, as observed by Yehudayoff [21], this does not yield a separation between MVNP and MVP. This is because the above argument implies that *any*

⁵Recall that we call a family of d -regular graphs $(G_n)_{n \geq 1}$ (with G_n a graph on n vertices) an expander sequence if the second largest (in absolute value) eigenvalue of its adjacency matrix A is at most $d(1 - \Omega(1))$. For the problem defined above, take $G = G_n$ in such a sequence.

polynomial P_0 that has support \mathcal{S} requires monotone algebraic circuits of exponential size. Yehudayoff showed that for any polynomial P in MVNP, there is a polynomial-sized monotone algebraic circuit that computes a polynomial Q with the same support. In particular, the polynomial P_0 cannot be in MVNP as that would contradict our lower bound above. Thus, to obtain a separation between MVNP and MVP along these lines, some new idea is necessary.

We take our cue from the multipartition communication complexity lower bound above, but modify it suitably to obtain a somewhat different lower bound candidate polynomial P . Our proof method for the lower bound, as in [21], is not just based on the support of P , but rather on the sizes of the coefficients of P . We define a probability distribution μ on the monomials of P and show that for any non-negative product polynomial $g_i h_i$ in a decomposition as in (1), a random monomial (chosen according to μ) has much smaller coefficient in the product polynomial than in P . As the product polynomials sum to P , there must be many of them. This yields the lower bound.

We explain this in some more detail below.

Detailed outline. The heart of the multipartition communication complexity lower bound for the function f_G is a more standard lower bound for the *non-deterministic communication complexity* of the *Disjointness* problem. Here, the non-deterministic communication complexity of a function f (or equivalently, the set system $\mathcal{S} \subseteq 2^{[n]}$ given by $f^{-1}(1)$) is defined in a similar way to multipartition communication complexity, except that each balanced rectangle \mathcal{R} is defined over the *same equipartition* (Y, Z) of $[n]$, which we can take to be the sets $[n/2]$ and $[n] \setminus [n/2]$ respectively; and the Disjointness function $D(x)$ is defined by the Boolean predicate $\bigwedge_{i \in [n/2]} (x_i \vee x_{i+n/2})$.⁶

The lower bound for the Disjointness function is proved by a standard *Fooling set* argument (see, e.g., [12]). We consider the $2^{n/2} \times 2^{n/2}$ *communication matrix* M , where the rows and columns are labelled by Boolean settings to variables indexed by Y and Z respectively and the (i, j) th entry of M is the disjointness predicate evaluated on the corresponding input. Further, assume that the rows are ordered using the lexicographical ordering of $\{0, 1\}^Y$, and the columns are ordered according to the *reverse* lexicographic ordering of $\{0, 1\}^Z$. This ensures that for any $i \in [2^{n/2}]$, the diagonal entry $M(i, i)$ corresponds to an input of the form (a, \bar{a}) where $a \in \{0, 1\}^Y$ and \bar{a} is the bitwise complement of a . From the definition of the Disjointness function, one can check that each diagonal entry of M is 1; further, given $i \neq j$, either $M(i, j)$ or $M(j, i)$ is 0. This implies that any rectangle over (Y, Z) that contains the i th diagonal entry cannot contain the j th diagonal entry for any $j \neq i$. In particular, the number of rectangles required to cover all the diagonal entries is $2^{n/2}$, implying a linear lower bound on the non-deterministic communication complexity of the Disjointness function.

For the multipartition setting, we can follow the above strategy to prove a lower bound for the function f_G defined above. The intuition is that for any graph G , the function f_G contains many copies of the Disjointness function above. In particular, taking any induced

⁶Strictly speaking, the Disjointness function is $\bigwedge_{i \in [m]} (\neg x_i \vee \neg x_{i+n/2})$ but we keep this definition for simplicity.

matching M of size m in G and setting variables corresponding to vertices $i \notin V(M)$ to 1, we get a copy f_M of the Disjointness function on $2m$ bits. Given any rectangle \mathcal{R} over the partition (Y, Z) , one can similarly prove that \mathcal{R} cannot contain many (suitably defined) “diagonal entries” of the communication matrix of f_M , *as long as* M contains many (say $\Omega(m)$) edges from the cut defined by (Y, Z) in G .

But there is a subtle question of how to choose M as above. In the multipartition setting, the partition (Y, Z) is not known ahead of time and furthermore, each rectangle comes with its own underlying partition. This is where the expanding nature of the graph G comes in. Standard facts about expander graphs imply that given any balanced partition (Y, Z) (i.e. $|Y|, |Z| \geq n/3$), a constant fraction of the edges of G lie in the cut defined by (Y, Z) . In particular, choosing M *randomly* guarantees that many edges of M lie in the cut with high probability. This leads to a proof of the multipartition communication complexity lower bound.

We now describe how this connects to the lower bounds of this paper for monotone algebraic circuits. We will follow a similar strategy, but instead of the 0s and 1s of the Boolean predicate, we will analyze the coefficients of the multilinear monomials in P and in the terms of the decomposition in (1). The polynomial P is defined using an expander graph G on vertex set $[n]$ (let us skip over what the definition of P is for the moment) and the hard distribution μ over the monomials of P is again just the process of choosing a random induced matching⁷ M of size m in G and considering the monomial $\prod_{i \in V(M)} x_i$.

The proof of the lower bound then proceeds as follows. Assume that P has a circuit of size s and consider the decomposition given in (1). Given a term $g_i h_i$ of the decomposition, we get a balanced partition (Y_i, Z_i) of the underlying variable set x_1, \dots, x_n . We argue that for a random monomial \mathbf{m} chosen according to the distribution μ , the expected value of the coefficient of \mathbf{m} in $g_i h_i$ is much smaller than its coefficient in P . To do this, we use a numerical analogue of the fooling set technique outlined above. Again, we consider the “communication matrix” M , which now is a $2^{|Y_i|} \times 2^{|Z_i|}$ matrix whose rows and columns are labelled by multilinear monomials in Y_i and Z_i respectively, and such that the entry corresponding to monomials $(\mathbf{m}_1, \mathbf{m}_2)$ is the coefficient of the product monomial $\mathbf{m}_1 \cdot \mathbf{m}_2$ in P . The main technical part of the proof shows the following: for independently sampled monomials \mathbf{m}' and \mathbf{m}'' (chosen from distribution μ) that factor as $\mathbf{m}'_1 \cdot \mathbf{m}'_2$ and $\mathbf{m}''_1 \cdot \mathbf{m}''_2$ respectively, where $\mathbf{m}'_1, \mathbf{m}''_1$ are monomials over Y_i and $\mathbf{m}'_2, \mathbf{m}''_2$ are monomials over Z_i , the coefficients of the “cross monomials” $\hat{\mathbf{m}} := \mathbf{m}'_1 \cdot \mathbf{m}''_2$ and $\tilde{\mathbf{m}} := \mathbf{m}''_1 \cdot \mathbf{m}'_2$ in P are much smaller than the coefficients of \mathbf{m}' and \mathbf{m}'' in P . This immediately implies that the coefficients of \mathbf{m}' and \mathbf{m}'' in $g_i h_i$ are smaller than they are in P by the following simple argument. If we let $\text{Coeff}(\mathbf{m}, Q)$ denote the coefficient of monomial \mathbf{m} in a polynomial Q , then we see that

$$\begin{aligned} \text{Coeff}(\mathbf{m}', g_i h_i) \cdot \text{Coeff}(\mathbf{m}'', g_i h_i) &= \text{Coeff}(\mathbf{m}'_1, g_i) \text{Coeff}(\mathbf{m}'_2, h_i) \text{Coeff}(\mathbf{m}''_1, g_i) \text{Coeff}(\mathbf{m}''_2, h_i) \\ &= \text{Coeff}(\hat{\mathbf{m}}, g_i h_i) \cdot \text{Coeff}(\tilde{\mathbf{m}}, g_i h_i). \end{aligned}$$

The latter term is upper bounded by the product of the coefficients of the monomials $\hat{\mathbf{m}}$ and

⁷For some technical reasons, we will actually choose M so that the non-adjacent vertices of M are at distance at least 3 from each other. But this can be ignored for now.

$\tilde{\mathbf{m}}$ in P (because of the decomposition (1)), which we already argued are much smaller than the coefficients of \mathbf{m}' and \mathbf{m}'' in P . This implies that a randomly chosen monomial \mathbf{m} has much smaller coefficient in any product term $g_i h_i$ than in P . Therefore, there must be many such terms in the decomposition (1). This implies the lower bound.

The above outline also indicates the property of P that allows the lower bound proof to work: we would like that the coefficients of \mathbf{m}' and \mathbf{m}'' are much larger than those of the monomials $\hat{\mathbf{m}}$ and $\tilde{\mathbf{m}}$. We do this by designing a polynomial P in MVNP where the coefficients of any monomial $\prod_{i \in S} x_i$ grows with the number of edges in the subgraph of G induced by the set S . Recall that for a random monomial \mathbf{m} chosen according to μ , S is the vertex set of a matching of size m and hence this induced subgraph has m edges. However, if m is sufficiently smaller than n (say $m \leq \alpha n$ for a small enough $\alpha > 0$), we do not expect the vertex sets of two independently chosen matchings of size m to have too many edges between them. This is what allows us to bound the coefficients of $\hat{\mathbf{m}}$ and $\tilde{\mathbf{m}}$, and prove the lower bound as above.

2 Defining the hard polynomial

Notation. Throughout, let $n \geq 1$ be a growing integer parameter. Let $X = \{x_1, \dots, x_n\}$ be a set of indeterminates. We use x^S to denote the monomial $\prod_{i \in S} x_i$. Given a polynomial $P \in \mathbb{R}[x_1, \dots, x_n]$ and $S \subseteq [n]$, we use $\text{Coeff}(x^S, P)$ to denote the coefficient of the monomial x^S in the polynomial P .

Let $(G_n)_{n>d}$ be an explicit sequence of d -regular expander graphs on n vertices with second largest eigenvalue at most $d^{0.75}$. Here, d is a large enough constant as specified below. Such an explicit sequence of expander graphs can be constructed using, say, [14]. The only fact we will use about expanders is the following, which is an easy consequence of the Expander Mixing Lemma [1] (see also [8, Lemma 2.5]).

For any pair of *disjoint* sets $U, V \subseteq V(G_n)$, we use $E(U, V)$ to denote the set of edges $\{u, v\} \in E(G_n)$ such that $u \in U$ and $v \in V$. Also, let $E(U)$ denote the set of edges $e = \{u, v\} \in E(G_n)$ such that $u, v \in U$.

Lemma 2 (Corollary to Expander Mixing Lemma). *Let G_n be as above. Then, for any disjoint sets $U, V \subseteq [n]$ such that $|U|, |V| \in [n/3, 2n/3]$, we have*

$$|E(U, V)| \geq \frac{|E(G_n)|}{10}.$$

as long as d is a large enough constant.

From now on, d will be fixed to be a large enough constant so that the inequality in Lemma 2 holds.

We define the polynomial $P_n(x_1, \dots, x_n)$ as follows. We assume that $V(G_n) = [n]$. For each edge $e \in E(G_n)$ introduce a variable x'_e and let $X' = \{x'_e \mid e \in E(G_n)\}$. Notice that for each Boolean assignment to the variables in X' , we obtain a subgraph H of G_n . In particular, if the variables in X' are set *randomly* to Boolean values, we get a random subgraph H of

G_n with the same vertex set $[n]$. We use $\deg_H(i)$ to denote the degree of the vertex i in the graph H .

We now define

$$P_n(x_1, \dots, x_n) = \mathbf{E}_{\substack{x'_e \in \{0,1\} \\ \forall e \in E(G_n)}} \left[\prod_{i \in [n]} (1 + x_i \cdot 2^{\deg_H(i)}) \right] \quad (2)$$

$$= \mathbf{E}_{\substack{x'_e \in \{0,1\} \\ \forall e \in E(G_n)}} \left[\sum_{S \subseteq [n]} x^S \cdot 2^{\sum_{i \in S} \deg_H(i)} \right] \quad (3)$$

where the variables x'_e are set to one of $\{0, 1\}$ independently and uniformly at random.

Lemma 3. *The sequence of polynomials P_n as defined above is in MVNP.*

Proof. Using (2), we see that

$$P_n(x_1, \dots, x_n) = \frac{1}{2^{|E(G_n)|}} \sum_{x'_e \in \{0,1\}: e \in E(G)} \prod_{i \in [n]} (1 + x_i \cdot 2^{\sum_{e \ni i} x'_e}).$$

Since G_n is d -regular, it suffices to show that each function $f : \{0, 1\}^d \rightarrow \mathbb{R}$ defined by $f(x'_1, \dots, x'_d) = 2^{\sum_{j \in [d]} x'_j}$ can be represented by a constant-sized polynomial over x'_1, \dots, x'_d with *non-negative* coefficients.

But this is clear since $f(x'_1, \dots, x'_d) = \sum_{S \subseteq [d]} \prod_{i \in S} x'_i$. □

3 The lower bound

The main theorem of this section is the following.

Theorem 4. *Any monotone circuit computing P_n has size $2^{\Omega(n)}$.*

We need the following lemma from [13]. We say that a pair of multilinear polynomials $(g, h) \in \mathbb{R}[X]$ form a *non-negative product pair* if g, h are polynomials with non-negative coefficients, and there is a partition of $X = Y \cup Z$ where $n/3 \leq |Y|, |Z| \leq 2n/3$ and $g \in \mathbb{R}[Y], h \in \mathbb{R}[Z]$.

Lemma 5 ([13], Lemma 3.3). *Assume that P_n has a monotone circuit of size s . Then*

$$P_n(X) = \sum_{i=1}^{s+1} g_i h_i$$

where for each $i \in [s]$, (g_i, h_i) forms a non-negative product pair.

Corollary 6. *Assume that P_n has a monotone circuit of size s . Let μ be any probability distribution on subsets $S \subseteq [n]$. Then, there is a non-negative product pair (g, h) such that*

- $gh \leq P$, i.e., $\text{Coeff}(x^S, gh) \leq \text{Coeff}(x^S, P_n)$ for each $S \subseteq [n]$,
- $\mathbf{E}_{S \sim \mu} [\text{Coeff}(x^S, gh)/\text{Coeff}(x^S, P_n)] \geq 1/(s+1)$. (The quantity $\text{Coeff}(x^S, gh)/\text{Coeff}(x^S, P_n)$ is well defined since by (3), the denominator is non-zero for all $S \subseteq [n]$.)

Proof. Write $P_n = \sum_{i \leq s+1} g_i h_i$ as in Lemma 5. For any fixed $S \subseteq [n]$ and a uniformly random $i \in [s+1]$, we have

$$\mathbf{E}_{i \in [s+1]} \left[\frac{\text{Coeff}(x^S, g_i h_i)}{\text{Coeff}(x^S, P_n)} \right] = \frac{1}{s+1} \sum_{i \in [s]} \frac{\text{Coeff}(x^S, g_i h_i)}{\text{Coeff}(x^S, P_n)} = \frac{1}{s+1}.$$

In particular, the above also holds when S is chosen according to μ . The result now follows by averaging over $i \in [s+1]$. \square

Given Corollary 6, to prove Theorem 4, it suffices to show the following.

Lemma 7. *There is a probability distribution μ on subsets $S \subseteq [n]$ such that for any non-negative product pair (g, h) with $gh \leq P_n$, we have*

$$\mathbf{E}_{S \sim \mu} [\text{Coeff}(x^S, gh)/\text{Coeff}(x^S, P_n)] \leq \exp(-\Omega(n)). \quad (4)$$

We need some preparatory work before proving Lemma 7.

Lemma 8. *There exist constants $A, B > 1$ such that*

$$P_n(X) = \sum_{S \subseteq [n]} x^S B^{|S|} A^{|E(S)|}.$$

Proof. Using (3), we obtain

$$P_n(x_1, \dots, x_n) = \mathbf{E}_{x'_e: e \in E(G)} \left[\sum_{S \subseteq [n]} x^S \cdot 2^{\sum_{i \in S} \deg_H(i)} \right] = \sum_{S \subseteq [n]} x^S \cdot \mathbf{E}_{x'_e: e \in E(G)} [2^{\sum_{i \in S} \deg_H(i)}]$$

where H is the random subgraph of G defined by a uniformly random Boolean assignment to the variables in X' . Note that

$$\sum_{i \in S} \deg_H(i) = \sum_{i \in S} \sum_{e \ni i} x'_e = \sum_{e \in E(S, \bar{S})} x'_e + \sum_{e \in E(S)} 2x'_e.$$

Hence, we get for any $S \subseteq [n]$,

$$\begin{aligned} \mathbf{E}_{x'_e: e \in E(G)} [2^{\sum_{i \in S} \deg_H(i)}] &= \prod_{e \in E(S, \bar{S})} \mathbf{E}_{x'_e} [2^{x'_e}] \cdot \prod_{e \in E(S)} \mathbf{E}_{x'_e} [4^{x'_e}] = (3/2)^{|E(S, \bar{S})|} \cdot (5/2)^{|E(S)|} \\ &= (3/2)^{|S|d} \cdot \frac{(5/2)^{|E(S)|}}{(3/2)^{2|E(S)|}} \end{aligned}$$

where for the last equality, we have used the fact that $2|E(S)| + |E(S, \bar{S})| = |S|d$. Note that this proves the lemma with $B = (3/2)^d$ and $A = (10/9)$. \square

We now define the probability distribution μ that will be shown to have the property in (4). The distribution is defined by the following sampling process. Let $m = \alpha n$ where $\alpha \in (0, 1)$ is a small constant specified below.

Sampling Algorithm \mathcal{S} :

1. Set $M = \emptyset$. (Eventually, M will be a matching of size $m/2$ in G .)
2. For $i = 1$ to $(m/2)$, do the following.
 - (a) Remove all vertices from G_n that are at distance at most 2 from any vertex in the matching M . Let $G_n^{(i)}$ be the resulting graph.
 - (b) Choose a uniformly random edge e_i from $E(G_n^{(i)})$ and add it to M .
3. Output M .

The above algorithm defines a distribution ν over matchings M in G_n of size $m/2$. We define $S = V(M)$ to be the set of vertices sampled by the algorithm. This defines a probability distribution μ over subsets of $[n]$.

We will need the following properties of the above algorithm.

Lemma 9 (Properties of \mathcal{S}). *Let M be sampled as in \mathcal{S} above and let $S = V(M)$. Then we have*

1. *Assuming that $\alpha \leq 1/(100 \cdot d^2)$, we have $|M| = (m/2)$, $|S| = m$ and $E(S) = M$ with probability 1.*
2. *Let (U, V) be any partition of $V(G_n)$ such that $n/3 \leq |U|, |V| \leq 2n/3$. Then, as long as $\alpha \leq 1/(100 \cdot d^2)$, for some absolute constant $\gamma > 0$, we have*

$$\Pr_M[|M \cap E(U, V)| \leq \gamma m] \leq \exp(-\gamma m).$$

3. *Let M_1 and M_2 be two independent samples obtained by running \mathcal{S} twice, and let $S_i = V(M_i)$ ($i \in [2]$). Let (U, V) be a partition of $V(G_n)$ as above. Define $R_i = S_i \cap U$ and $T_i = S_i \cap V$. Then, for $\alpha \leq \gamma \ln A / (100 \cdot A^4 d^2)$ we have*

$$\mathbf{E}_{M_1, M_2} [A^{|E(R_1, T_2)| + |E(R_2, T_1)|}] \leq A^{\gamma m/4}.$$

Here, γ is as in the previous item and A is as in the statement of Lemma 8.

Proof. Item 1 easily follows from the definition of the Sampling algorithm \mathcal{S} . Note that in each iteration of Step 2, we remove at most $2 \cdot (1 + d + d^2)$ vertices and hence at most $2(d + d^2 + d^3)$ edges from the graph G_n . Hence, the upper bound on α guarantees that after $i < (m/2)$ iterations of the for loop, the number of edges removed from the graph is at most

$$2i \cdot (d^3 + d^2 + d) < 4md^3 = 4\alpha nd^3 < \frac{nd}{2},$$

which allows the algorithm to choose an edge from the graph $G_n^{(i)}$ to add to the matching M .

For Item 2, we proceed as follows. For $i \in \{1, \dots, m/2\}$, let e_i be the edge chosen by the sampling algorithm \mathcal{S} in the i th iteration of Step 2. Fix any choices of all the e_j with $j < i$ and consider the i th iteration of Step 2. The probability that e_i lies in $E(U, V)$ is $|E_i(U, V)|/|E(G_n^{(i)})|$ where $E_i(U, V)$ is the set of edges in $G_n^{(i)}$ with one endpoint each in U and V . Note that

$$|E_i(U, V)| \geq |E(U, V)| - |E(G_n) \setminus E(G_n^{(i)})| \geq \frac{nd}{10} - 2(i-1) \cdot (d^3 + d^2 + d) \geq \frac{nd}{10} - \alpha n(3d^3) \geq \frac{nd}{20}$$

where the second inequality follows from Lemma 2 and an analysis similar to Item 1 above; the last two inequalities follow from the fact that $2(i-1) < m = \alpha n \leq n/(100 \cdot d^2)$. Hence, we have shown that for each i ,

$$\Pr[e_i \in E(U, V) \mid e_1, \dots, e_{i-1}] = \frac{|E_i(U, V)|}{|E(G_n^{(i)})|} \geq \frac{(nd)/20}{(nd)/2} = \frac{1}{10}.$$

In particular, for any $T \subseteq [m]$, the probability that for every $i \in T$, $e_i \notin E(U, V)$ can be upper bounded by $(9/10)^{|T|}$.

Thus, the probability that $|M \cap E(U, V)| \leq \ell = \gamma m$ can be bounded by

$$\begin{aligned} \Pr_M[\exists T \in \binom{[m]}{m-\ell} \text{ s.t. } \forall i \in T, e_i \notin E(U, V)] &\leq \sum_T \Pr_M[\forall i \in T, e_i \notin E(U, V)] \\ &\leq \binom{m}{\ell} \left(\frac{9}{10}\right)^{m-\ell} \leq \left(\frac{em}{\ell}\right)^\ell \cdot \left(\frac{9}{10}\right)^{m-\ell} \\ &= \left(\frac{e}{\gamma} \cdot \left(\frac{9}{10}\right)^{(1/\gamma)-1}\right)^{\gamma m} \leq \exp(-\gamma m) \end{aligned}$$

as long as γ is bounded by a small enough absolute constant. This finishes the proof of Item 2.

We now prove Item 3. Fix any possible matching M_1 as sampled by the algorithm \mathcal{S} . It suffices to bound $\mathbf{E}_{M_2} [A^{|E(R_2, T_1)| + |E(R_1, T_2)|}]$ for each such M_1 . Let \tilde{S}_1 denote the set of vertices that are at distance at most 1 from S_1 and let E_1 denote the set of edges that have at least one endpoint in \tilde{S}_1 . Note that $|E_1| \leq |\tilde{S}_1|d \leq |S_1|d^2 = md^2 = \alpha nd^2$.

We claim that $|E(R_2, T_1)| + |E(R_1, T_2)| \leq 4|E_1 \cap M_2|$. The reason for this is that if a vertex $i \in S_2$ is incident to an edge e in $E(R_1, T_2) \cup E(R_2, T_1)$ then $i \in \tilde{S}_1$ and hence the edge $e' \in M_2$ involving i is an edge in $E_1 \cap M_2$. In particular, the number of such vertices $i \in S_2$ is at most $2|E_1 \cap M_2|$. Further, each such vertex i is adjacent to at most 2 vertices in S_1 since vertices in S_1 that are not adjacent via an edge in M_1 are at distance at least 3 from each other. Thus each such vertex i contributes at most 2 to $|E(R_2, T_1)| + |E(R_1, T_2)|$. This yields the claimed inequality. Thus it suffices to bound $\mathbf{E}_{M_2} [A^{4|E_1 \cap M_2|}]$.

We start with a tail bound for $|E_1 \cap M_2|$. Let $M_2 = \{e'_1, \dots, e'_{m/2}\}$ where e'_j is the j th edge added to M_2 by the algorithm \mathcal{S} . Conditioned on e'_1, \dots, e'_{j-1} , the probability that $e'_j \in E_1$ is at most

$$\frac{|E_1|}{|E(G_n^{(j)})|} = \frac{|E_1|}{|E(G_n)| - |E(G_n) \setminus E(G_n^{(j)})|} \leq \frac{\alpha nd^2}{(nd/2) - 3\alpha nd^3} \leq \frac{\alpha nd^2}{(nd)/4} = 4\alpha d$$

where for the first inequality we have bounded $|E(G_n) \setminus E(G_n^{(j)})|$ and $|E_1|$ as above and for the second inequality we have used the bound on α . Hence, we have

$$\Pr_{M_2}[|E_1 \cap M_2| \geq i] \leq \sum_{T \in \binom{[m/2]}{i}} \Pr_{M_2}[\forall j \in T, e'_j \in E_1] \leq \binom{m/2}{i} (4\alpha d)^i.$$

This allows us to bound $\mathbf{E}_{M_2}[A^{4|E_1 \cap M_2|}]$ for any fixed M_1 output by \mathcal{S} .

$$\begin{aligned} \mathbf{E}_{M_2}[A^{4|E_1 \cap M_2|}] &\leq \sum_{i=0}^{m/2} A^{4i} \Pr_{M_2}[|E_1 \cap M_2| \geq i] \\ &\leq \sum_{i=0}^{m/2} A^{4i} \cdot \binom{m/2}{i} (4\alpha d)^i = (1 + 4\alpha d A^4)^{m/2} \\ &\leq (1 + (\gamma \ln A)/2)^{m/2} \leq \exp((m\gamma \ln A)/4) = A^{\gamma m/4} \end{aligned}$$

where the third inequality follows from the bound $\alpha \leq \gamma \ln A / (100 \cdot A^4 d^2)$ assumed in the statement of the lemma. \square

We are now ready to prove Lemma 7, which will complete the proof of Theorem 4.

Proof of Lemma 7. We set $m = \alpha n$ so that α is a positive constant upper bounded by $\gamma \ln A / (100 \cdot A^4 \cdot d^2)$ and m is even. Assume that M is as sampled above by sampling algorithm \mathcal{S} and $S = V(M)$. This defines the distribution μ on subsets of $[n]$.

Let (g, h) be any non-negative product pair such $gh \leq P_n$. Consequently, there exists a partition (U, V) of $V(G_n) = [n]$ such that $n/3 \leq |U|, |V| \leq 2n/3$ and $g \in \mathbb{R}[x_i : i \in U], h \in \mathbb{R}[x_j : j \in V]$.

Let $\mathcal{E} = \mathcal{E}(M)$ denote the event that $|M \cap E(U, V)| \leq \gamma m$. By Lemma 9 item 2, we know that $\Pr_M[\mathcal{E}] \leq \exp(-\Omega(n))$ and hence we have

$$\begin{aligned} \mathbf{E}_M \left[\frac{\text{Coeff}(x^S, gh)}{\text{Coeff}(x^S, P_n)} \right] &\leq \mathbf{E}_M \left[\frac{\text{Coeff}(x^S, gh)}{\text{Coeff}(x^S, P_n)} \mid \mathcal{E} \right] \Pr_M[\mathcal{E}] + \mathbf{E}_M \left[\frac{\text{Coeff}(x^S, gh)}{\text{Coeff}(x^S, P_n)} \mid \bar{\mathcal{E}} \right] \Pr_M[\bar{\mathcal{E}}] \\ &\leq \Pr_M[\mathcal{E}] + \mathbf{E}_M \left[\frac{\text{Coeff}(x^S, gh)}{\text{Coeff}(x^S, P_n)} \mid \bar{\mathcal{E}} \right] \\ &\leq \exp(-\Omega(n)) + \frac{1}{B^m \cdot A^{m/2}} \mathbf{E}_M [\text{Coeff}(x^S, gh) \mid \bar{\mathcal{E}}] \end{aligned} \quad (5)$$

where for the second inequality we have used that $gh \leq P_n$, and for the final inequality we have used our bound on $\Pr_M[\mathcal{E}]$ along with Lemma 8 and Lemma 9 item 1.

We now bound the latter term in (5). For any i, j, k , let $\mathcal{E}_{i,j,k} = \mathcal{E}_{i,j,k}(M)$ denote the event that $|M \cap E(U, V)| = i, |M \cap E(U)| = j$, and $|M \cap E(V)| = k$. The event $\bar{\mathcal{E}}$ is partitioned into $\mathcal{E}_{i,j,k}$ where $i + j + k = (m/2)$ and $i \geq \gamma m$. Let \mathcal{T} denote the set of such triples (i, j, k) . We have

$$\mathbf{E}_M [\text{Coeff}(x^S, gh) \mid \bar{\mathcal{E}}] = \sum_{(i,j,k) \in \mathcal{T}} \mathbf{E}_M [\text{Coeff}(x^S, gh) \mid \mathcal{E}_{i,j,k}] \cdot \Pr_M[\mathcal{E}_{i,j,k} \mid \bar{\mathcal{E}}]$$

Call a triple $(i, j, k) \in \mathcal{T}$ *heavy* if $\Pr_M[\mathcal{E}_{i,j,k}] \geq A^{-\gamma m/4}$ and *light* otherwise. Note that as $\Pr_M[\bar{\mathcal{E}}] = 1 - \exp(-\Omega(n)) \geq 1/2$, we have $\Pr_M[\mathcal{E}_{i,j,k} \mid \bar{\mathcal{E}}] \leq 2 \Pr_M[\mathcal{E}_{i,j,k}]$. In particular, if (i, j, k) is light, we have $\Pr_M[\mathcal{E}_{i,j,k} \mid \bar{\mathcal{E}}] \leq 2A^{-\gamma m/4} = \exp(-\Omega(n))$. Plugging this into the expression above, we get

$$\begin{aligned} \mathbf{E}_M [\text{Coeff}(x^S, gh) \mid \bar{\mathcal{E}}] &= \sum_{(i,j,k) \in \mathcal{T}} \mathbf{E}_M [\text{Coeff}(x^S, gh) \mid \mathcal{E}_{i,j,k}] \cdot \Pr_M[\mathcal{E}_{i,j,k} \mid \bar{\mathcal{E}}] \\ &\leq |\{(i, j, k) \mid (i, j, k) \text{ light}\}| \cdot B^m A^{m/2} \cdot \exp(-\Omega(n)) + \max_{(i,j,k) \text{ heavy}} \mathbf{E}_M [\text{Coeff}(x^S, gh) \mid \mathcal{E}_{i,j,k}] \\ &\leq \exp(-\Omega(n)) B^m A^{m/2} + \max_{(i,j,k) \text{ heavy}} \mathbf{E}_M [\text{Coeff}(x^S, gh) \mid \mathcal{E}_{i,j,k}]. \end{aligned} \quad (6)$$

It suffices therefore to bound $\mathbf{E}_M [\text{Coeff}(x^S, gh) \mid \mathcal{E}_{i,j,k}]$ for any heavy (i, j, k) . This is the main part of the proof.

Fix some $(i, j, k) \in \mathcal{T}$ that is heavy. Let $C = \mathbf{E}_M [\text{Coeff}(x^S, gh) \mid \mathcal{E}_{i,j,k}]$. Thus, we get

$$C^2 = \mathbf{E}_{M_1, M_2} [\text{Coeff}(x^{S_1}, gh) \text{Coeff}(x^{S_2}, gh)]$$

where M_1 and M_2 are independent samples of M conditioned on the event $\mathcal{E}_{i,j,k}(M)$, and $S_\ell = V(M_\ell)$ for $\ell \in \{1, 2\}$. Define $R_\ell = S_\ell \cap U$ and $T_\ell = S_\ell \cap V$. We make some simple observations. For each $\ell \in [2]$

1. $|R_\ell| = i + 2j$ and $|T_\ell| = i + 2k$,
2. $|E(R_\ell)| = j, |E(T_\ell)| = k$,
3. $\text{Coeff}(x^{S_\ell}, gh) = \text{Coeff}(x^{R_\ell}, g) \cdot \text{Coeff}(x^{T_\ell}, h)$.

Thus, we have

$$\begin{aligned}
C^2 &= \mathbf{E}_{M_1, M_2} [\text{Coeff}(x^{R_1}, g) \text{Coeff}(x^{T_1}, h) \text{Coeff}(x^{R_2}, g) \text{Coeff}(x^{T_2}, h)] \\
&= \mathbf{E}_{M_1, M_2} [\text{Coeff}(x^{R_1 \cup T_2}, gh) \text{Coeff}(x^{R_2 \cup T_1}, gh)] \\
&\leq \mathbf{E}_{M_1, M_2} [\text{Coeff}(x^{R_1 \cup T_2}, P_n) \text{Coeff}(x^{R_2 \cup T_1}, P_n)] \\
&= \mathbf{E}_{M_1, M_2} [B^{|R_1|+|T_2|} \cdot A^{|E(R_1 \cup T_2)|} \cdot B^{|R_2|+|T_1|} \cdot A^{|E(R_2 \cup T_1)|}] \\
&= \mathbf{E}_{M_1, M_2} [B^{4(i+j+k)} \cdot A^{|E(R_1)|+|E(T_1)|+|E(R_2)|+|E(T_2)|+|E(R_1, T_2)|+|E(R_2, T_1)|}] \\
&= B^{2m} A^{m-2i} \cdot \mathbf{E}_{M_1, M_2} [A^{|E(R_1, T_2)|+|E(R_2, T_1)|}] \\
&\leq B^{2m} A^{m(1-2\gamma)} \cdot \mathbf{E}_{M_1, M_2} [A^{|E(R_1, T_2)|+|E(R_2, T_1)|}] \tag{7}
\end{aligned}$$

where we used the observations above for the equalities and for the final inequality, we used the fact that $i \geq \gamma m$ for all $(i, j, k) \in \mathcal{T}$.

To bound the latter term in (7), we consider a similar expression where M_1 and M_2 are replaced by M'_1 and M'_2 which are independent random outputs of the algorithm \mathcal{S} (without any conditioning). In this case, by Lemma 9 item 3, we have

$$\mathbf{E}_{M'_1, M'_2} [A^{|E(R'_1, T'_2)|+|E(R'_2, T'_1)|}] \leq A^{\gamma m/4},$$

where R'_ℓ, T'_ℓ are defined analogously for $\ell \in [2]$. Thus, using Bayes's rule we have

$$\mathbf{E}_{M_1, M_2} [A^{|E(R_1, T_2)|+|E(R_2, T_1)|}] \leq \frac{\mathbf{E}_{M'_1, M'_2} [A^{|E(R'_1, T'_2)|+|E(R'_2, T'_1)|}]}{\Pr_{M'_1, M'_2} [\mathcal{E}_{i,j,k}(M'_1) \wedge \mathcal{E}_{i,j,k}(M'_2)]} \leq A^{3\gamma m/4}$$

where the last inequality uses the fact that (i, j, k) is heavy. Plugging the above into (7), we have

$$C \leq B^m A^{m/2-5\gamma m/8} = B^m A^{m/2} \exp(-\Omega(n)).$$

As this holds for any heavy (i, j, k) , using (6) and (5), we obtain the statement of the lemma. \square

Acknowledgements. The author is grateful to Mrinal Kumar and Amir Yehudayoff for very helpful discussions and encouragement. This work was done during a visit to the ‘‘Lower Bounds Program in Computational Complexity’’ program at the Simons Institute for the Theory of Computing; the author is grateful to the organizers of this program and the Simons Institute for their hospitality. The author is also grateful to Igor Sergeev and an anonymous reviewer for pointing out the large body of work on monotone circuit lower bounds carried out by the Russian mathematical community (see [6]). Finally, the author would like to thank the anonymous reviewers (who reviewed this paper for the ACM Transactions on Computation Theory) for their comments and suggestions.

References

- [1] N. Alon and F. Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 72(1):15 – 19, 1988.
- [2] A. Borodin, A. A. Razborov, and R. Smolensky. On lower bounds for read-k-times branching programs. *Computational Complexity*, 3:1–18, 1993.
- [3] S. Bova, F. Capelli, S. Mengel, and F. Slivovsky. Expander CNFs have exponential DNNF size. *CoRR*, abs/1411.1995, 2014.
- [4] P. Duris, J. Hromkovic, S. Jukna, M. Sauerhoff, and G. Schnitger. On multi-partition communication complexity. *Inf. Comput.*, 194(1):49–75, 2004.
- [5] S. B. Gashkov. On the complexity of monotone computations of polynomials. *Vestn. Mosk. Univ., Ser. I*, 1987(5):7–13, 1987.
- [6] S. B. Gashkov and I. S. Sergeev. A method for deriving lower bounds for the complexity of monotone arithmetic circuits computing real polynomials. *Sbornik. Mathematics*, 203(10), 10 2012.
- [7] T. P. Hayes. Separating the k-party communication complexity hierarchy: an application of the Zarankiewicz problem. *Discrete Mathematics & Theoretical Computer Science*, 13(4):15–22, 2011.
- [8] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.
- [9] M. Jerrum and M. Snir. Some exact complexity results for straight-line computations over semirings. *J. ACM*, 29(3):874–897, 1982.
- [10] S. Jukna. Lower bounds for tropical circuits and dynamic programs. *Theory Comput. Syst.*, 57(1):160–194, 2015.
- [11] O. M. Kasim-Zade. The complexity of monotone polynomials. In *Proceedings of the All-Union seminar on discrete mathematics and its applications (Russian) (Moscow, 1984)*, pages 136–138. Moskov. Gos. Univ., Mekh.-Mat. Fak., Moscow, 1986.
- [12] A. Rao and A. Yehudayoff. *Communication Complexity: and Applications*. Cambridge University Press, 2020.
- [13] R. Raz and A. Yehudayoff. Multilinear formulas, maximal-partition discrepancy and mixed-sources extractors. *J. Comput. Syst. Sci.*, 77(1):167–190, 2011.
- [14] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of mathematics*, pages 157–187, 2002.

- [15] R. Saptharishi. A survey of lower bounds in arithmetic circuit complexity. *Github survey*, 2015.
- [16] C.-P. Schnorr. A lower bound on the number of additions in monotone computations. *Theoret. Comput. Sci.*, 2(3):305–315, 1976.
- [17] E. Shamir and M. Snir. *Lower bounds on the number of multiplications and the number of additions in monotone computations*. IBM Thomas J. Watson Research Division, 1977.
- [18] 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.
- [19] L. G. Valiant. Completeness classes in algebra. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1979, Atlanta, Georgia, USA*, pages 249–261, 1979.
- [20] L. G. Valiant. Negation can be exponentially powerful. *Theor. Comput. Sci.*, 12:303–314, 1980.
- [21] A. Yehudayoff. Separating monotone VP and VNP. In *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 425–429. ACM, New York, 2019.