Electronic Colloquium on Computational Complexity, Report No. 22 (2015)

Lower bounds for non-commutative skew circuits*

Nutan Limaye† Guillaume Malod‡ Srikanth Srinivasan§

Abstract

Nisan (STOC 1991) exhibited a polynomial which is computable by linear sized non-commutative circuits but requires exponential sized non-commutative algebraic branching programs. Nisan's hard polynomial is in fact computable by linear sized *skew* circuits (skew circuits are circuits where every multiplication gate has the property that all but one of its children is an input variable or a scalar). We prove that any non-commutative skew circuit which computes the square of the polynomial defined by Nisan must have exponential size.

As a further step towards proving exponential lower bounds for general non-commutative circuits, we also extend our techniques to prove an exponential lower bound for a class of circuits which is a restriction of general non-commutative circuits and a generalization of non-commutative skew circuits. More precisely, we consider non-commutative circuits of small *non-skew depth*, which denotes the maximum number of non-skew gates on any path from the output gate to an input gate. We show that for any $k < d$, there is an explicit polynomial of degree d over n variables that has non-commutative circuits of polynomial size but such that any circuit with non-skew depth k must have size at least $n^{\Omega(d/k)}$. It is not hard to see that any polynomial of degree d that has polynomial size circuits has a polynomial-sized circuit with non-skew depth d. Hence, our results should be interpreted as proving lower bounds for the class of circuits with non-trivially small non-skew depth.

As far as we know, this is the strongest model of non-commutative computation for which we have superpolynomial lower bounds.

1 Introduction

If we want to design an efficient algorithm for a computational problem that is naturally stated as a polynomial — such as the determinant or the permanent, matrix multiplication, Fast Fourier Transform, etc. — then *arithmetic circuits* capture most natural candidate algorithms that we might consider. An arithmetic circuit is an algorithm that starts with the input variables and possibly some constants in the underlying field, and iteratively applies addition and multiplication operations until it computes the desired polynomial. There has been a large body of work proving upper and lower bounds on the arithmetic circuit complexity of various polynomials (see, e.g. the

*This research was funded by IFCPAR/CEFIPRA Project No 4702-1(A) and by the ANR project CompA (ANR-13-BS02-0001-01)

†IIT Bombay, Department of Computer Science and Engineering, Mumbai, India. nutan@cse.iitb.ac.in

‡Univ Paris Diderot, Sorbonne Paris Cité, Institut de Mathématiques de Jussieu, UMR 7586 CNRS, F-75205 Paris, France. malod@math.univ-paris-diderot.fr

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

ISSN 1433-8092

surveys [21, 6]). In particular, proving explicit superpolynomial lower bounds for general arithmetic circuits is a celebrated open question in complexity theory and one of the possible approaches to the P versus NP question (see, e.g., [4]). However, despite more than three decades of intensive study, it has seen little tangible progress (in the sense of concrete lower bounds for general circuits).

In this paper, we concentrate on *non-commutative arithmetic circuits*, which compute polynomials in the *non-commutative* polynomial ring $\mathbb{F}\langle X\rangle$: here, variables do not commute upon multiplication; that is, xy and yx (for distinct $x, y \in X$) are distinct monomials. There are two reasons for looking at such circuits. The first is that such circuits yield algorithms for polynomial functions over non-commutative *algebras*, which arise naturally and can even have applications for *commutative* computations (see [7, 2], in particular the use of non-commutative determinants to approximate the commutative permanent). The second reason is that proving explicit lower bounds for non-commutative arithmetic circuits is formally an easier problem than that of proving lower bounds for (commutative) arithmetic circuits described in the previous paragraph, and it is hoped that techniques discovered in the course proving non-commutative lower bounds will be useful in the commutative setting as well.

The works of Hyafil [10] and Nisan [14] were among the first to motivate the study of arithmetic circuits from this latter point of view. In a breakthrough result, Nisan [14] showed exponential lower bounds for non-commutative arithmetic formulas (a restriction of general non-commutative arithmetic circuits) and more generally for non-commutative algebraic branching programs (ABPs). This might have led one to think that a superpolynomial lower bound for general (non-commutative) arithmetic circuits[1] was also close at hand. However, Nisan also showed using the same techniques that general arithmetic circuits are exponentially more powerful than arithmetic formulas and ABPs, hinting that his techniques are not sufficient to prove lower bounds for general arithmetic circuits. Indeed, there is no known lower bound for general non-commutative arithmetic circuits that is stronger than those that we already have for general *commutative* arithmetic circuits.

In a more recent work, Hrubeš, Wigderson, and Yehudayoff [9] suggested a new line of attack on the general arithmetic circuit lower bound question. Their result introduces a "product lemma" for general arithmetic circuits that generalizes a decomposition of ABPs due to Nisan [14]. Using this lemma, they are able to show that superpolynomial lower bounds for general arithmetic circuits would follow from a strong enough lower bound for the classical *Sum-of-squares* problem. However, as of now, this approach has not yielded superpolynomial arithmetic circuit lower bounds. Therefore, the strongest known computational model for which we have superpolynomial lower bounds remains the ABPs from the work of Nisan [14].

In this work, we prove exponential lower bounds for *skew circuits*. Skew circuits are arithmetic circuits where every multiplication involves at least one argument[2] that is either an input variable or a field element. They are a well-studied model of computation [22, 13, 1, 12], especially in the commutative setting, where they are equivalent in power to ABPs and to the evaluation of the determinant polynomial. However, the picture seems more complicated in the non-commutative setting. Nisan [14] has shown that skew circuits are exponentially more powerful than ABPs. Thus, our lower bound for this model can be seen as one step towards the goal of superpolynomial lower

[1] From here on, all circuits, formulas, ABPs, and polynomials, unless explicitly mentioned otherwise, will be non-commutative.

[2] We assume fan-in 2 for all gates.

bounds for general non-commutative circuits.[3]

Our result also clarifies the relative power of skew circuits vis-à-vis general arithmetic circuits. In fact, our lower bound shows that skew circuits are exponentially less powerful than circuits with just *one* non-skew gate (that is, neither of its arguments is an input variable or field element). This is because the explicit polynomial for which we prove a lower bound is just the square of a polynomial considered by Nisan, and this polynomial in turn has skew circuits of linear size.

We also consider the problem of extending our lower bound to more powerful classes of circuits. A natural way to do this (and one that is analogous to many works in the *Boolean circuit* setting; see, e.g. [3, 5, 11]) is to augment a circuit for which we do have lower bounds with a few "powerful" gates and see if one can still prove a lower bound. We therefore consider the problem of proving lower bounds for skew circuits with a "few" non-skew multiplication gates.

We say that the *non-skew depth* of a non-commutative circuit is the maximum number of non-skew gates on a path from a variable to the output gate in the DAG underlying the circuit. We prove that for infinitely many $d \in \mathbb{N}$ and any $k, n \in \mathbb{N}$, there exists a polynomial of degree d on n variables which is computable by a polynomial sized non-commutative circuit of non-skew depth $O(k)$ but requires size $n^{\Omega(d/k)}$ for any non-commutative circuit of non-skew depth k.

In particular our result implies that there exists a polynomial of degree d which is computable by a polynomial sized non-commutative circuit of non-skew depth d, but requires a superpolynomial size for any non-commutative circuit of non-skew depth $k(d) = o(d)$. It is not hard to see that any polynomial of degree d that can be computed by a polynomial-sized arithmetic circuit can also be computed by a polynomial-sized arithmetic circuit of non-skew depth d: hence, strengthening our lower bound substantially would prove lower bounds for general non-commutative circuits.

We also show that the determinant polynomial can simulate our hard polynomial, thus completing the picture in the non-commutative setting by showing that skew circuits are exponentially less powerful than the determinant polynomial. Finally, we show that to prove superpolynomial lower bounds for general non-commutative circuits, our complexity measure (to be defined formally in the upcoming section) will need to be further refined. Slightly more precisely, we show that there is a polynomial that has polynomial-sized non-commutative circuit, but for which our complexity measure is as large as possible.

The rest of the paper is organized as follows. We start with a proof outline in Section 2. We then present some definitions in Section 3 and preliminaries in Section 4. The proof of the lower bound for skew circuits is presented in Section 5 and the proof for the lower bound for non-skew depth bounded circuits is presented in Section 6[4]. Finally, we extend the lower bound to the permanent and determinant polynomials in Section 7.

[3] A superpolynomial lower bound for non-commutative skew circuits was claimed by Allender et al. [1], but, unfortunately, the proof of this particular result in the paper (Theorem 7.12) seems to fail because it did not take into account possible cancellations (Meena Mahajan, personal communication).

[4] As skew circuits are a subset of bounded non-skew depth circuits, our lower bound for bounded non-skew depth circuits subsumes the lower bound for skew circuits. However, for the sake of exposition we first describe the lower bound proof for skew circuits and then prove the lower bound for bounded non-skew depth circuits.

2 Proof Outline

Our overall proof strategy is similar to that of Nisan [14] for non-commutative formulas and algebraic branching programs (ABPs). In his work, Nisan considered the *partial derivative matrix* corresponding to a homogeneous polynomial $f \in \mathbb{F}\langle X \rangle$ of degree d — originally introduced by Hyafil [10] — which is defined to be an $n^{d/2} \times n^{d/2}$ matrix $M[f]$ where the rows and columns are labelled by monomials in X of degree $d/2$. The (m_1, m_2)th entry of the matrix $M[f]$ is defined to be the coefficient of the monomial $m_1 m_2$ in f.[5]

Nisan observed that if f has a formula or ABP of small size, then f can be decomposed as a small sum of polynomials of the form $g \cdot h$ where g and h are homogeneous polynomials of degree $d/2$. Crucially, it may be seen that for any such g, h the matrix $M[g \cdot h]$ has rank 1 and hence, by subadditivity of rank, $M[f]$ has small rank. Thus, choosing an f such that $\mathrm{rank}(M[f])$ is large gives us a lower bound.

Intuitively speaking, the rank of the matrix $M[f]$ is a measure of how "correlated" the first half of a monomial appearing in f is with its second half: $M[f]$ being full rank would mean that they are perfectly correlated, whereas $M[f]$ being low rank would mean that they are not very correlated at all. Nisan's argument shows that small ABPs have "information bottlenecks" at degree $d/2$ (and indeed at any degree $d' \leq d$), and hence the amount of correlation is small.

A natural question to ask is if this argument can give a lower bound for non-commutative skew circuits as well. Unfortunately, the answer is no, as is already implicit in Nisan's work. Consider the Palindrome polynomial $\mathsf{PAL}_{d/2}(X)$, which is the sum of all monomials of degree d that are palindromes when viewed as strings of length d over the alphabet X. Nisan observed that $\mathsf{PAL}_{d/2}(X)$ has a skew circuit of linear size but at the same time $M[\mathsf{PAL}_{d/2}(X)]$ has full rank: in fact, $M[\mathsf{PAL}_{d/2}(X)]$ is a permutation matrix since the first half of a palindrome uniquely determines the second half (thus, the first and second halves of monomials appearing in f are perfectly correlated). Hence, the partial derivative matrix of polynomials with small skew circuits can have as large a rank as possible. This means that in our lower bound argument for skew circuits, we need to use a different measure of complexity.

The measure that we use is a modified version of the partial derivative matrix, defined as follows: let $f \in \mathbb{F}\langle X \rangle$ be a homogeneous polynomial of degree d over n variables, and given an ordered partition $\Pi = (Y, Z)$ of $[d]$ into two parts, we define $M[f, \Pi]$ to be the matrix whose rows and columns are indexed by monomials in X of degree $|Y|$ and $|Z|$ respectively. The (m_1, m_2)th entry of $M[f, \Pi]$ is defined to the coefficient of the unique monomial m of degree d which equals m_1 if we keep only the variables indexed by locations in Y and delete the others, and equals m_2 if we only keep the variables indexed by locations in Z. As above, the rank of $M[f, \Pi]$ measures the correlation between the restriction of a monomial to the locations in Y and the locations in Z. We are usually interested in Π where $|Y| \leq |Z|$, since in this case we know that the maximum possible rank is $\min\{n^{|Y|}, n^{|Z|}\} \leq n^{|Y|}$.

In this notation, the measure of complexity used by Nisan is $\mathrm{rank}(M[f, ([d/2], [d] \setminus [d/2])])$ and we have seen above that this measure is as large as it can be for, say, the Palindrome polynomial $\mathsf{PAL}_{d/2}(X)$, which has a small skew circuit. However, it is an easy observation that if one considers

[5] More generally, Nisan also considered the matrix where the rows and columns are labelled by monomials of degree d' and $d - d'$ respectively.

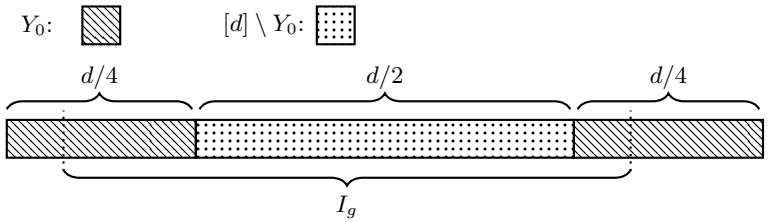
the partition $\Pi_0 = (Y_0, Z_0)$ where $Z_0 := [d/4 + 1, 3d/4]$ and $Y_0 := [d] \setminus Z_0$, then $M[\mathsf{PAL}_{d/2}(X), \Pi_0]$ has rank 1.

Thus, we might hope that for every polynomial f that has a small skew circuit, we could find a Π such that $M[f, \Pi]$ has low rank. We are in fact able to show something much stronger: we can show in general that if f has a small skew circuit, then $\mathrm{rank}(M[f, \Pi_0])$ is 'small' for the particular Π_0 defined above. (Here, 'small' means that the rank is much smaller than full rank.)

In terms of correlation, this statement could be interpreted as saying that though skew circuits can compute polynomials that are perfectly correlated w.r.t. Nisan's partition $([d/2], [d] \setminus [d/2])$, they can only do so by correlating the initial few indices in the monomial with the final few indices, as in the Palindrome polynomial. Consequently, these "extreme" indices end up uncorrelated with those in the middle. This is the weakness of skew circuits that we exploit in our lower bound.

The proof of this fact rests on a decomposition of skew circuits that is motivated by the similar ABP decomposition mentioned above. Like in the ABP decomposition, we can show that given any homogeneous polynomial f of degree d that has a small skew circuit and any degree parameter $d' \in [d]$, we can decompose f as a small sum of polynomials of the form $g \times_j h$ where g and h are polynomials of degree d' and $d - d'$ respectively (we refer the reader to Section 3 for the definition of $\times_j$, but it intuitively means that the polynomial g is multiplied on the left by the sum of the prefixes of the monomials of h of degree j and on the right by the sum of the suffixes of degree $d - d' - j$). The proof of this lemma is obtained by specializing the proof of a lemma of Hrubeš, Wigderson and Yehudayoff [9] regarding general non-commutative arithmetic circuits to the case of skew circuits, where it yields a stronger conclusion.

Given this decomposition lemma, we prove the lower bound as follows. We apply the lemma with d' being a large number close to d: for concreteness, say $d' = 3d/4$. In other words, we decompose f as a small sum of polynomials $g \times_j h$ where g and h are homogeneous polynomials of degrees $3d/4$ and $d/4$ respectively. In each such polynomial, a set $I_g \subseteq [d]$ of $3d/4$ indices corresponds to g and a set $I_h = [d] \setminus I_g$ corresponds to the polynomial h as shown below:



As we mentioned above, we will consider the rank of the matrix $M[g \times_j h, \Pi_0]$. Now, it is easy to show that

$$\mathrm{rank}(M[g \times_j h, \Pi_0]) = \mathrm{rank}(M[g, \Pi_g]) \cdot \mathrm{rank}(M[h, \Pi_h])$$

where the partitions $\Pi_g = (Y_g, Z_g)$ and $\Pi_h = (Y_h, Z_h)$ are the natural restrictions of Π_0 to I_g and I_h respectively.

Note that if $\mathrm{rank}(M[g \times_j h, \Pi_0])$ is to be close to full — i.e. $n^{|Y_0|}$ — then we need both $\mathrm{rank}(M[g, \Pi_g])$ and $\mathrm{rank}(M[h, \Pi_h])$ to be close to $n^{|Y_g|}$ and $n^{|Y_h|}$ respectively. However, it is easily seen that, irrespective of the value of j, the matrix $M[h, \Pi_h]$ is *always* a rank 1 matrix (this happens since Y_h occupies all of I_h and thus $Z_h = \emptyset$) and hence $\mathrm{rank}(M[g \times_j h, \Pi_0])$ falls *exponentially* short

of its maximum possible value. Since f is a small sum of such polynomials, the same is true of $\mathrm{rank}(M[f, \Pi_0])$ as well. More generally, the same strategy shows that $\mathrm{rank}(M[f, \Pi])$ is small as long as $\Pi = (Y, Z)$ has the "left-right monochromatic" form (LRM partitions for short) shown in Figure 1 (for d_1, d_2 large enough).

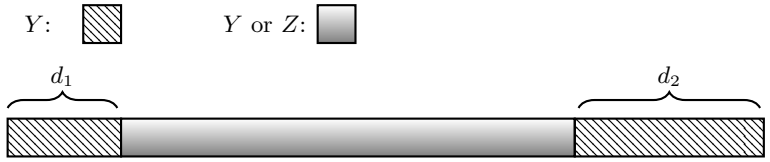


Figure 1: Left-right monochromatic (LRM) partitions, where segments on both the left and right ends are contained in Y

The above argument implies a strong exponential lower bound on the size of a skew circuit computing any homogeneous polynomial F of degree d such that $M[F, \Pi_0]$ is full rank. It is easy to find explicit examples of such polynomials: for example, we could take F to be the square of $\mathsf{PAL}_{d/4}(X)$ or the Lifted Identity polynomial of Hrubeš et al. [9]. In either of these cases, it can be checked that $M[F, \Pi_0]$ is again a permutation matrix and hence full rank. Since $(\mathsf{PAL}_{d/4}(X))^2$ can be computed by a small circuit with just a *single* non-skew gate, this also gives an exponential separation between skew circuits and circuits with one non-skew gate. However, this also implies that if we want to extend our lower bound to non-commutative circuits of small non-skew depth, then we need to modify our measure further.

We prove our lower bound for circuits of small non-skew depth by induction on the non-skew depth k of the circuit. As in the skew case, we choose a partition Π_k of $[d]$ such that no small non-skew depth k circuit can compute a polynomial that has large rank w.r.t. the partition Π_k. The inductive argument is based on showing that if a non-skew depth k circuit C computes a polynomial of large rank w.r.t. Π_k, then it must contain a depth $k-1$ circuit that computes a polynomial of large rank w.r.t. Π_{k-1} (or an even 'harder' partition). We then apply the inductive hypothesis to prove the lower bound.

Let us consider the problem of constructing such a partition in the case $k = 1$ (i.e. non-skew depth 1). Ideally, we would like to construct a partition Π_1 such that if C is a circuit of non-skew depth 1 that is high rank w.r.t. Π_1, then a sub-circuit of C is high rank w.r.t. an LRM partition as in Figure 1 (with perhaps a slightly smaller degree). However, it can be checked that we *cannot* choose such a partition even if we know beforehand that C is just a product of two skew circuits. That is, for any candidate partition Π_1, there are skew circuits of degree $d' \leq d$ and $d - d'$ computing polynomials g_1 and g_2 such that neither the partition restricted to g_1, nor the partition restricted to g_2, is LRM.

Hence, we are first led to the problem of enlarging the family of partitions that are hard for skew circuits. Building on the techniques outlined for skew circuits above, we can also show that small skew circuits cannot compute high rank polynomials w.r.t. the larger family of "extended LRM" (XLRM) partitions — illustrated in Figure 2 — which are obtained by extending an LRM partition on the left and right sides with segments of length ℓ that are contained in Y and Z respectively.[6] Intuitively, a skew circuit that computes a large rank polynomial w.r.t. such a partition would try to

[6]The actually family of partitions we consider is a little more general.

pairwise correlate indices in the segments (of length ℓ) on the two extremes. However, after having done this, it is still left with the task of computing a high rank polynomial w.r.t. an LRM partition, which we know to be a hard problem.

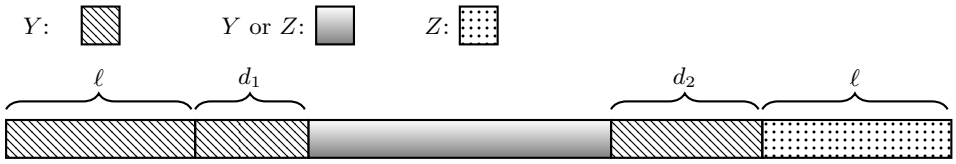


Figure 2: Extended left-right monochromatic (XLRM) partitions

We are now ready to tackle the problem of proving lower bounds for circuits of non-skew depth k. We choose our hard partition $\Pi_k = (Y_k, Z_k)$ to have the form shown in Figure 3. That is, starting from the left, our partition assigns an initial segment of length roughly $d/4$ to Y_k. The remaining indices are assigned to Y_k and Z_k in k' pairs of segments of length roughly $d/4k'$ and $d/2k'$ respectively — for $k' = O(k)$ — so that overall we have $|Y_k| = |Z_k| = d/2$. Note that Π_k is in particular an XLRM partition, and hence is clearly hard for skew circuits. We show that any small circuit C of non-skew depth at most k cannot compute a polynomial of large rank w.r.t. Π_k.

To get an idea of the proof, consider first the easier case when the output of C is a non-skew homogeneous multiplication gate and hence C is a product of two homogeneous polynomials g_1 and g_2 that have small circuits of non-skew depth at most $k-1$. In this case, the indices in $[d]$ are distributed between g_1 and g_2 as shown in Figure 3. Now, as we have argued previously, if the polynomial f computed by C is to have rank nearly $n^{|Y_k|}$ w.r.t. Π_k, then $\mathrm{rank}(M[g_i, \Pi_{k,i}])$ should be close to $n^{|Y_k^{(i)}|}$ where $\Pi_{k,i} = (Y_{k,i}, Z_{k,i})$ is the natural restriction of Π_k to the indices corresponding to g_i for $i \in [2]$. For this to occur, however, we must have $|Y_{k,i}| \approx |Z_{k,i}|$ for each i: since otherwise for some i, we will have $|Z_{k,i}|$ much smaller than $|Y_{k,i}|$, and then $\mathrm{rank}(M[g_i, \Pi_{k,i}]) \leq n^{|Z_{k,i}|} \ll n^{|Y_{k,i}|}$. However, it is easy to check that if $|Y_{k,i}| \approx |Z_{k,i}|$ for each i, then the only possibility is that one of g_1 or g_2 — say g_1 for concreteness — has very small degree and the other "occupies" almost all the indices in $[d]$ and is hence already computing a polynomial of large rank w.r.t. Π_k. Since g_2 has a small skew circuit of non-skew depth at most $k-1$, this allows us to induct on g_2.

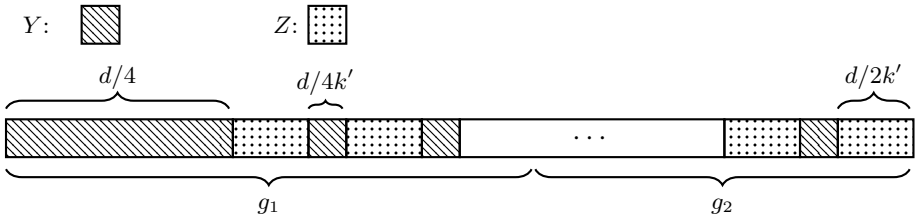


Figure 3: The partition Π_k

The general case puts together a couple of arguments we have already outlined. Using a decomposition lemma that is similar in spirit to the skew circuit decomposition lemma described above, we can show that any homogeneous polynomial f of degree d computed by a small circuit of

non-skew depth at most k can be written as a small sum of polynomials of the form

$$(g_1 \cdot g_2) \times_j h$$

where g_1 and g_2 are homogeneous polynomials computed by small circuits of non-skew depth at most $k-1$ and h has a small *skew* circuit. In the easy case above, we have already handled the case when $\deg(h) = 0$, and so now we try to see how h can help produce a polynomial of large rank w.r.t. the partition Π_k. As in the proof of the hardness of XLRM partitions, one would guess that the worst that h could do is to match up the $d/2k'$ indices in Y and Z on either extreme. In this case, we can argue as in the easier case above that one of g_1 or g_2 occupies all that is remaining, which corresponds to a partition that is hard for non-skew depth at most $k-1$, as desired.

As might be expected, the actual proof is not quite as neat, since we need to handle some other cases that we have not describe above. It turns out, however, that these cases are easy, even if somewhat tedious, to handle.

3 Definitions

Throughout, fix the set $X = \{x_1, \ldots, x_n\}$ of indeterminates. We work over the non-commutative ring of polynomials $\mathbb{F}\langle X \rangle$.

For $i, j \in \mathbb{N}$, we define $[i, j]$ to be the set $\{i, i+1, \ldots, j\}$ (the set is empty if $i > j$). We also use the standard notation $[i]$ to denote the set $[1, i]$.

For $d \in \mathbb{N}$, we use $\mathcal{M}_d(X)$ to denote the set of monomials over the variables in X of degree exactly d.

Definition 1 (j-products)**.** *Given homogeneous polynomials $g, h \in \mathbb{F}\langle X \rangle$ of degrees d_g and d_h respectively and an integer $j \in [0, d_h]$, we define the* j-product of g and h *— denoted $g \times_j h$ — as follows:*

- *When g and h are monomials, then we can factor h uniquely as a product of two monomials $h_1 h_2$ such that $\deg(h_1) = j$ and $\deg(h_2) = d_h - j$. In this case, we define $g \times_j h$ to be $h_1 \cdot g \cdot h_2$.*

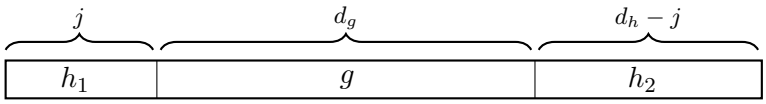


Figure 4: j product for monomials g, h

- *The map is extended bilinearly to general homogeneous polynomials g, h. Formally, let g, h be general homogeneous polynomials, where $g = \sum_\ell g_\ell$, $h = \sum_i h_i$ and g_ℓ, h_i are monomials of g, h respectively. For $j \in [0, d_h]$, each h_i can be factored uniquely into h_{i_1}, h_{i_2} such that $\deg(h_{i_1}) = j$ and $\deg(h_{i_2}) = d_h - j$. And $g \times_j h$ is defined to be $\sum_i \sum_\ell h_{i_1} g_\ell h_{i_2}$.*

Note that $g \times_0 h$ and $g \times_{d_h} h$ are just the products $g \cdot h$ and $h \cdot g$ respectively.

The following easily verifiable facts about j-products will be useful:

Fact 2. *1. The operator $\times_j$ is bilinear: i.e. $(g_1+g_2)\times_j h = g_1\times_j h + g_2\times_j h$ and $g\times_j(h_1+h_2) = g\times_j h_1 + g\times_j h_2$ provided that g, g_1, g_2, h, h_1, h_2 are such that all the above expressions are well defined.*

2. Assume g and h are such that $g\times_j h$ is defined and let f be a homogeneous polynomial of degree d. Then $(g\times_j h)\cdot f = g\times_j (h\cdot f)$ and $f\cdot(g\times_j h) = g\times_{d+j}(f\cdot h)$.

3. Assume g and h are as above and further that $g = g_1\cdot g_2$. Then $g\times_j h = g_1\times_j (g_2\times_j h) = g_2\times_{j+d_{g_1}}(g_1\times_j h)$ where $d_{g_1} = \deg(g_1)$. If instead we have $g = g_1\times_k g_2$, then $g\times_j h = g_1\times_{j+k}(g_2\times_j h)$.

Given a monomial $m = x_{i_1}x_{i_2}\cdots x_{i_d}\in\mathbb{F}\langle X\rangle$ and a subset $S\subseteq[d]$, we denote by m_S the product of all the variables in the locations indexed by S: i.e. $m_S = \prod_{j\in S} x_{i_j}$ where the product is taken in increasing order of j.

For any pair of subsets $S, I\subseteq[d]$ such that $S\subseteq I$, we denote by $\texttt{Collapse}(S, I)$ the subset of $[|I|]$ which contains the ranks of all elements in I which are contained in S. Formally,

$$\texttt{Collapse}(S, I) = \{j\in[|I|] \mid S \text{ contains the } j\text{th smallest element of } I\}.$$

Let Π denote a partition of $[d]$ given by an ordered pair (Y, Z), where $Y\subseteq[d]$ and $Z = [d]\setminus Y$. In what follows we only use partitions of sets into two parts.

Definition 3 (Partial Derivative matrix)**.** *Let $f\in\mathbb{F}\langle X\rangle$ be a homogeneous polynomial of degree d. Given a partition $\Pi = (Y, Z)$ of $[d]$, we define a $n^{|Y|}\times n^{|Z|}$ matrix $M[f,\Pi]$ with entries from $\mathbb{F}$ as follows: the rows of $M[f,\Pi]$ are labelled by monomials from $\mathcal{M}_{|Y|}(X)$ and the columns by elements of $\mathcal{M}_{|Z|}(X)$. Let $m'\in\mathcal{M}_{|Y|}(X)$ and $m''\in\mathcal{M}_{|Z|}(X)$; the (m', m'')th entry of $M[f,\Pi]$ is the coefficient in the polynomial f of the unique monomial m such that $m_Y = m'$ and $m_Z = m''$.*

We will use the rank of the matrix $M[f,\Pi]$ (for a suitably defined $\Pi = (Y, Z)$) as a measure of the complexity of f. Note that since the rank of the matrix is at most the number of rows, we have for any $f\in\mathbb{F}\langle X\rangle$

$$\text{rank}(M[f,\Pi])\leq n^{|Y|}$$

As in many works on multilinear formulas and circuits [15, 16, 17, 18, 19, 8], we will be interested in how close the rank of $M[f,\Pi]$ can be to this trivial upper bound.

Definition 4 (Relative Rank)**.** *Let $f\in\mathbb{F}\langle X\rangle$ be a homogeneous polynomial of degree d. For any $Y\subseteq[d]$, we define the* relative rank of f w.r.t. $\Pi = (Y, Z)$ *— denoted* rel-rank(f,Π) *— to be*

$$\text{rel-rank}(f,\Pi) := \frac{\text{rank}(M[f,\Pi])}{n^{|Y|}}$$

Clearly, rel-rank$(f,\Pi)\in[0,1]$ for any f and Y as above. Furthermore, note that since rank$(M[f,\Pi])$ is also bounded by $n^{|Z|}$ — the number of columns in the matrix — when $|Y| > d-|Y|$, this measure cannot approach 1 for any choice of f.

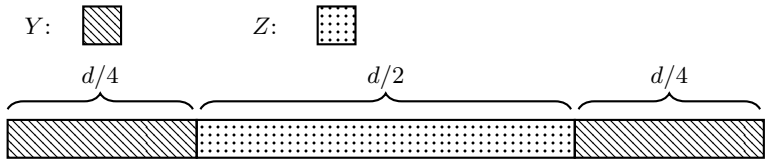


Figure 5: Example of Y for which rel-rank$(\mathsf{PAL}^2_{d/4}, (Y, Z)) = 1$

Notation Fix any homogeneous polynomials $g, h \in \mathbb{F}\langle X\rangle$ of degree d_g and d_h respectively and $f = g \times_j h$, where $j \in [0, d_h]$. Let d denote $\deg(f) = d_g + d_h$ and I denote $[j+1, j+d_g]$. For any partition $\Pi = (Y, Z)$ of $[d]$ we use Y_g to denote $\texttt{Collapse}(Y \cap I, I)$, i.e. the set of ranks of indices that g occupies in $g \times_j h$ which overlap with Y. Similarly, we use Y_h to denote $\texttt{Collapse}(Y \setminus I, [d] \setminus I)$, i.e. the set of ranks of indices that h occupies in $g \times_j h$ which overlap with Y. Also we denote $[d_g] \setminus Y_g$ by Z_g and $[d_h] \setminus Y_h$ by Z_h. Finally, we use Π_g, Π_h to denote partitions (Y_g, Z_g) and (Y_h, Z_h), respectively.

The *non-skew depth* of a non-commutative circuit C is the maximum number of non-skew gates on a path from a variable to the output gate in the DAG underlying C.

4 Preliminaries

We need the following lemmas that are straightforward adaptations of previous work.

Lemma 5 (Homogenization Lemma [9])**.** *Let $f \in \mathbb{F}\langle X\rangle$ be a homogeneous polynomial of degree d computed by a non-commutative circuit C of size s. Then there is a homogeneous non-commutative circuit C' of size at most $O(sd^2)$ computing f. Moreover, if C has non-skew depth at most k, then so does C'. In particular, if C is a skew circuit, then so is C'.*

Lemma 6 (Tensor Lemma)**.** *Let $g, h \in \mathbb{F}\langle X\rangle$ be homogeneous polynomials of degree d_g and d_h respectively and let $f = g \times_j h$ for $j \in [0, d_h]$. Let d denote $\deg(f) = d_g + d_h$. Fix any partition $\Pi = (Y, Z)$ of $[d]$. Then,*

$$\operatorname{rank}(M[f, \Pi]) = \operatorname{rank}(M[g, \Pi_g]) \cdot \operatorname{rank}(M[h, \Pi_h])$$

where Π_g, Π_h are as defined in Section 3.

Proof. We observe that under a suitable labelling of the rows and columns of the matrices, the matrix $M[f, \Pi] = M[g, \Pi_g] \otimes M[h, \Pi_h]$, where $\otimes$ represents the standard tensor (or Kronecker) product of matrices. This will prove the lemma.

Let I denote the interval $[j+1, j+d_g]$.

For each of the matrices $M[f, \Pi_f], M[g, \Pi_g]$ and $M[h, \Pi_h]$, we have labellings from the definitions of these matrices: i.e., the rows and columns of $M[f, \Pi_f]$ are labelled by elements of $\mathcal{M}_{|Y_f|}$ and $\mathcal{M}_{|Z_f|}$ respectively; and similarly for $M[g, \Pi_g]$ and $M[h, \Pi_h]$. For $M[f, \Pi]$, we note that each monomial $m \in \mathcal{M}_{|Y|}$ can be identified with a pair of monomials (m', m'') of degree $|Y_g|$ and $|Y_h|$ respectively using the map $m \mapsto (m_{Y \cap I}, m_{Y \setminus I})$; this map is a bijection and hence, we also have an *alternate* labelling of the rows of $M[f, \Pi]$ by $\mathcal{M}_{|Y_g|} \times \mathcal{M}_{|Y_h|}$; similarly, we also obtain a labelling of

the *columns* of $M[f, Y]$ by $\mathcal{M}_{|Z_g|} \times \mathcal{M}_{|Z_h|}$. Under this alternate labelling for $M[f, \Pi]$, we show that $M[f, \Pi] = M[g, \Pi_g] \otimes M[h, \Pi_h]$.

By the bilinearity of both the $\otimes$ and $\times_j$ maps, it suffices to do this when g and h are both monomials. In this case, $M[g, \Pi_g]$ is a 0-1 matrix with a 1 *only* in the (g_{Y_g}, g_{Z_g})th entry and similarly for $M[h, \Pi_h]$. Since f is also a monomial, the matrix $M[f, \Pi]$ is also a 0-1 matrix with a 1 only in the (f_Y, f_Z)th entry according to the *original* labelling. Under our alternate labelling of $M[f, \Pi]$, this corresponds to the $((f_{Y \cap I}, f_{Y \setminus I}), (f_{Z \cap I}, f_{Z \setminus I}))$th entry of $M[f, \Pi]$. It can be checked from the definition of $\times_j$ that

$$f_{Y \cap I} = g_{Y_g}, f_{Z \cap I} = g_{Z_g}, f_{Y \setminus I} = h_{Y_h}, f_{Z \setminus I} = h_{Z_h}$$

Thus, f has a 1 in only the $((g_{Y_g}, h_{Y_h}), (g_{Z_g}, h_{Z_h}))$th entry and hence, $M[f, \Pi]$ is the tensor product of $M[g, \Pi_g]$ and $M[h, \Pi_h]$ as claimed. This completes the proof of the lemma. □

Corollary 7. *Assume that f, Y, d_g, d_h are as in the statement of Lemma 6. Then*

$$\text{rel-rank}(f, \Pi) = \text{rel-rank}(g, \Pi_g) \cdot \text{rel-rank}(h, \Pi_h) \leq \min\left\{\text{rel-rank}(g, \Pi_g), \text{rel-rank}(h, \Pi_h)\right\}.$$

Moreover, we also have $\text{rank}(M[g, \Pi_g]) \leq n^{|Z_g|}$ *and* $\text{rank}(M[h, \Pi_h]) \leq n^{|Z_h|}$. *Hence,*

$$\text{rel-rank}(f, \Pi) \leq \min\left\{n^{-(|Y_g| - |Z_g|)}, n^{-(|Y_h| - |Z_h|)}\right\}.$$

4.1 Hard polynomials

Let $w = (w_1, w_2, \ldots, w_d)$ be a string in $[n]^d$ and let w^R denote the reverse of the string w, i.e., $(w_d, w_{d-1}, \ldots, w_1)$. Let $\tilde{x}_w$ denote the monomial $x_{w_1} x_{w_2} \ldots x_{w_d}$ over the variable set $X = \{x_1, x_2, \ldots, x_n\}$. We consider the *$n$-variable palindrome polynomial*:

$$\mathsf{PAL}_d(X) = \sum_{w \in [n]^d} \tilde{x}_w \cdot \tilde{x}_{w^R}.$$

Nisan [14] studied the palindrome polynomial for $n = 2$. We denote by $\mathsf{PAL}_d^2(X)$ the *squared palindrome polynomial*.

$$\mathsf{PAL}_d^2(X) = (\mathsf{PAL}_d(X))^2 = \sum_{w_1, w_2 \in [n]^d} \tilde{x}_{w_1} \cdot \tilde{x}_{w_1^R} \cdot \tilde{x}_{w_2} \cdot \tilde{x}_{w_2^R}.$$

5 Lower bound for skew circuits

In this section, we prove an exponential lower bound for skew circuits. We start by giving a decomposition lemma for such circuits. A similar decomposition was given by Nisan [14] for non-commutative ABPs. More recently Hrubeš et al. [9] proved a decomposition lemma for general non-commutative circuits. Our result can be thought of as an interpolation between the decomposition for ABPs and that for general non-commutative circuits.

We then formally define left-right monochromatic (LRM) partitions and prove that any skew circuit of 'small' size has 'small' relative rank with respect to LRM partitions. Finally, we give an explicit polynomial which has full relative rank with respect to a suitably chosen LRM partition. This gives a lower bound for skew circuits.

Lemma 8 (Decomposition Lemma for skew circuits)**.** *Let $f \in \mathbb{F}\langle X\rangle$ be a homogeneous polynomial of degree $d \in \mathbb{N}$ computed by a homogeneous skew circuit C of size s. Fix any $d' \in [d]$. Let $g_1, \ldots, g_t$ ($t \leq s$) be the intermediate polynomials of degree d' computed by C. Then, there exist homogeneous polynomials $h_{i,j}$ ($i \in [t], j \in [0, d-d']$) of degree $d - d'$ such that*

$$f = \sum_{i \in [t]} \sum_{j \in [0, d-d']} g_i \times_j h_{i,j}$$

Furthermore, each $h_{i,j}$ can be computed by a skew circuit of size at most sd.

Proof. Let $v_1, \ldots, v_s$ be a topological ordering of the gates in C that furthermore satisfies $\deg(f_p) \leq \deg(f_q)$ for all $p \leq q$, where f_k is the polynomial computed by gate v_k. Further, let d_k denote $\deg(f_k)$ for any $k \in [s]$.

We show by induction on $k \in [s]$ the following statement that immediately implies the lemma: if $d_k \geq d'$, then there exist polynomials $h_{i,j}^{(k)}$ ($i \in [t], j \in [0, d_k - d']$) of degree $d_k - d'$ such that

$$f_k = \sum_{i \in [t]} \sum_{j \in [0, d_k - d']} g_i \times_j h_{i,j}^{(k)} \tag{1}$$

Furthermore, for each k as above and $i \in [t]$, there is a skew circuit $C_i^{(k)}$ of size at most kd that computes all the polynomials in the set $S_i^{(k)} = \{h_{i,j}^{(\ell)} \mid \ell \in [k], d_\ell \geq d', j \in [d_\ell - d']\}$.

We define the polynomials $h_{i,j}^{(k)}$ by induction on k. The construction of the skew circuits computing $S_i^{(k)}$ will follow directly from the definitions of these polynomials. Since there is nothing to prove for k such that $d_k < d'$, we may take the base case of the induction to be the least k such that $d_k \geq d'$; say this gate is v_{k_0}. We claim that $d_{k_0} = d'$. To see this, observe that if $d' \geq 2$ then v_{k_0} must be a multiplication gate with both inputs being polynomials of degree at least 1 (i.e. neither input is a field element). Since C is a skew circuit we have $f_{k_0} = f_p \cdot f_q$ — with $p, q < k_0$ — where either d_p or d_q is equal to 1. We assume that $d_p = 1$ (the case when $d_q = 1$ is similar). Since $d_q < d'$ by our choice of k_0, we must have $d_{k_0} = d'$. If $d' = 1$, then it is easy to see that v_{k_0} is just a variable and therefore $d_{k_0} = d'$ trivially follows.

Hence, the gate v_{k_0} computes a polynomial g_{i_0} for some $i_0 \in [t]$. In this case, we can take $h_{i_0,0}^{(k_0)} = 1$ and $h_{i,0}^{(k_0)} = 0$ for $i \neq i_0$. Equation (1) clearly holds with this choice of polynomials and each $h_{i,0}^{(k_0)}$ has a skew circuit $C_i^{(k_0)}$ of size 1, which is at most $k_0 \cdot d$. Since $S_i^{(k_0)} = \{h_{i,0}^{(k_0)}\}$, this completes the base case of the induction.

Now for the inductive argument. Fix any $k > k_0$. By our ordering of the vertices, we have $d_k \geq d_{k_0} = d'$. The rest of the argument is based on a case analysis depending on the type of the gate v_k:

- v_k **is a product gate of degree** $d_k = d'$: In this case, the argument proceeds exactly as in the base case and each $h_{i,0}^{(k)}$ can be computed by a skew circuit of size 1. Adding this gate to the circuit $C_i^{(k-1)}$ gives the circuit $C_i^{(k)}$.
- v_k **is a product gate of degree** $d_k > d'$: So $f_k = f_p \cdot f_q$ where we assume that $d_p \leq d_q$ (the other case is similar) and hence $d_p \leq 1$. We divide this case further into subcases depending on whether $d_p = 0$ or 1.

– $d_p = 0$: In this case, f_p is a scalar from the field $\mathbb{F}$ and $d_q = d_k > d'$. By the induction hypothesis, we have polynomials $h_{i,j}^{(q)}$ for each $(i,j) \in [t] \times [0, d_k - d']$ satisfying

$$f_q = \sum_{i \in [t]} \sum_{j \in [0, d_k - d']} g_i \times_j h_{i,j}^{(q)}$$

Thus, we have

$$f_k = f_p \cdot f_q = \sum_{i \in [t]} \sum_{j \in [0, d_k - d']} g_i \times_j (f_p \cdot h_{i,j}^{(q)})$$

It can thus be seen that the polynomials $h_{i,j}^{(k)} := f_p \cdot h_{i,j}^{(q)}$ satisfy (4).

– $d_p = 1$: This case is quite similar, except that we only have $h_{i,j}^{(q)}$ for $(i,j) \in [t] \times [0, d_k - d' - 1]$. Again, we have

$$f_k = f_p \cdot f_q = \sum_{i \in [t]} \sum_{j \in [0, d_k - d' - 1]} f_p \cdot (g_i \times_j h_{i,j}^{(k)}) = \sum_{i \in [t]} \sum_{j \in [0, d_k - d' - 1]} g_i \times_{j+1} (f_p \cdot h_{i,j}^{(q)})$$

where we use Fact 2 for the final equality.
Thus, we define $h_{i,j}^{(k)} := f_p \cdot h_{i,j-1}^{(q)}$ for $j \in [1, d_k - d']$ and $h_{i,0}^{(k)} = 0$.

- v_k **is a sum gate of degree** $d_k \geq d'$: In this case, $v_k = v_p + v_q$ where $d_p = d_q = d_k \geq d'$ (by the homogeneity of C). Hence, the induction hypothesis is applicable to both v_p and v_q. It can now be easily checked that setting $h_{i,j}^{(k)} := h_{i,j}^{(p)} + h_{i,j}^{(q)}$ gives the required polynomials.

To see that the circuits constructed this way have the required properties, it suffices to note that to construct $C_i^{(k)}$ from $C_i^{(k-1)}$ using the above definition of the $h_{i,j}^{(k)}$ polynomials requires us to add exactly $d_k - d' \leq d - 1$ many skew multiplication gates and one variable or constant (in case v_k is a product gate) or $d_k - d' + 1 \leq d$ homogeneous addition gates (in case v_k is a sum gate). Hence, in each case, we add at most d gates to obtain the circuit $C_i^{(k)}$. □

Definition 9. *We say that a partition* $\Pi = (Y, Z)$ *of* $[d]$ *is a* (d_1, d_2)-left right monochromatic partition *((d_1, d_2)-LRM) if* $[d_1] \cup [d - d_2 + 1, d] \subseteq Y$.

Figure 6 gives an illustration of a (d_1, d_2)-LRM partition.

Lemma 10 (Main Lemma: Relative rank of skew circuits)**.** *Let* $f \in \mathbb{F}\langle X \rangle$ *be a homogeneous polynomial of degree* $d \in \mathbb{N}$ *computed by a homogeneous skew circuit* C *of size* s*. For any* (d_1, d_2)*-LRM partition* Π *of* $[d]$ *such that* $d_1 + d_2 \leq d$

$$\text{rel-rank}(f, \Pi) \leq sd \cdot n^{-\min\{d_1, d_2\}}.$$

Proof. Assume that $D = \min\{d_1, d_2\}$. Apply the Decomposition Lemma for skew circuits (Lemma 8) to C with $d' = d - D$ to get polynomials g_i and $h_{i,j}$ for $(i,j) \in [t] \times [0, D]$ as in the statement of the lemma. By the subadditivity of rank, we have

$$\text{rel-rank}(f, \Pi) \leq \sum_{(i,j) \in [t] \times [0,D]} \text{rel-rank}(g_i \times_j h_{i,j}, \Pi) \tag{2}$$

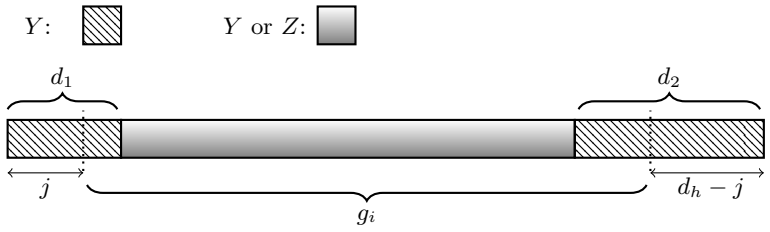


Figure 6: For fixed d_1, d_2, a generic positioning of g_i of degree d' in $g_i \times_j h_{i,j}$

Fix any (i, j) and consider rel-rank$(g_i \times_j h_{i,j}, \Pi)$. By Corollary 7, we have

$$\text{rel-rank}(g_i \times h_{i,j}, \Pi) \leq n^{-(|Y_h|-|Z_h|)}. \tag{3}$$

where $Y_h = \texttt{Collapse}(Y \setminus [j+1, j+d'], [d] \setminus [j+1, j+d'])$ and $Z_h = [D] \setminus Y_h$. Note, however, that since Y contains $[d_1] \cup [d - d_2 + 1, d]$, we have $Y \setminus [j+1, j+d'] = [d] \setminus [j+1, j+d']$ and hence $Y_h = [D]$ and $Z_h = \emptyset$. Using (3), we see rel-rank$(g_i \times h_{i,j}, \Pi) \leq n^{-D}$ and hence by (2), we have the claimed upper bound on rel-rank(f, Π). □

Theorem 11 (Lower bound for skew circuits)**.** *Any skew circuit for* $\mathsf{PAL}^2_{d/4}(X)$ *must have size* $\tilde{\Omega}(n^{d/4})$ *where the* $\tilde{\Omega}(\cdot)$ *hides* poly(d) *factors.*

Proof. Let C be any skew circuit computing $\mathsf{PAL}^2_{d/4}(X)$ and let s denote its size. By Lemma 5, we know that there is a homogeneous circuit of size $s' = O(sd^2)$ computing the same polynomial.

Let $Y = [d/4] \cup [3d/4 + 1, d]$, $Z = [d] \setminus Y$, $\Pi = (Y, Z)$. Note that Π is a $(d/4, d/4)$-LRM partition of $[d]$. Apply Lemma 10 to the circuit C' with $d_1 = d_2 = d/4$. The lemma implies that rel-rank$(\mathsf{PAL}^2_{d/4}(X), \Pi) \leq (s'd) \cdot n^{-d/4}$.

On the other hand, it is easy to verify that $M[\mathsf{PAL}^2_{d/4}(X), \Pi]$ is a square permutation matrix and hence rel-rank$(\mathsf{PAL}^2_{d/4}(X), \Pi) = 1$, which implies the claimed lower bound on s. □

Remark 12. *It is not hard to see that the lower bound of Theorem 11 is close to tight, since* $\mathsf{PAL}^2_{d/4}(X)$ *does have a skew circuit of size* $O(n^{d/4})$.

A similar theorem can be proved for the Lifted Identity polynomial of Hrubeš et al. [9].

6 Lower bounds for circuits with small non-skew depth

We call a gate v in C *top-most* if there is a path from v to the output gate in C that does not pass through any non-skew gates other than possibly v itself.

6.1 A decomposition lemma for circuits of non-skew depth k

Lemma 13 (Decomposition Lemma for circuits with non-skew depth k)**.** *Let* $f \in \mathbb{F}\langle X \rangle$ *be a homogeneous polynomial of degree* d *computed by a homogeneous circuit* C *of non-skew depth at most* k. *Let* $g_1, \ldots, g_t$ $(t \leq s)$ *be the polynomials computed by the top-most non-skew gates in* C *and*

let $d'_i = \deg(g_i)$ for $i \in [t]$. Then, there exist homogeneous polynomials $h_{i,j}$ ($i \in [t], j \in [0, d - d'_i]$) of degree $d - d'_i$ and h_0 of degree d such that

$$f = \sum_{i \in [t]} \sum_{j \in [0, d - d'_i]} g_i \times_j h_{i,j} + h_0.$$

Furthermore, each $h_{i,j}$ and h_0 can be computed by a homogeneous skew circuit of size at most sd.

Proof. The proof is quite similar to that of Lemma 8, so we omit some of the details. As there, we let $v_1, \ldots, v_s$ be a topological ordering of the gates in C that satisfies $\deg(f_p) \leq \deg(f_q)$ for all $p \leq q$, where f_k is the polynomial computed by gate v_k; further, let d_r denote $\deg(f_r)$ for any $r \in [s]$.

Let $T \subseteq V$ be the set of all top-most gates. In particular, T contains the output gate of C and also the set of all the top-most non-skew gates.

We show by induction on $r \in [s]$ the following stronger statement: if $v_r \in T$, then there exist homogeneous polynomials $h^{(r)}_{i,j}$ ($i \in [t], j \in [0, d_r - d'_i]$) of degree $d_r - d'_i$ and $h^{(r)}_0$ of degree d_r such that

$$f_r = \sum_{i \in [t]} \sum_{j \in [0, d_r - d'_i]} g_i \times_j h^{(r)}_{i,j} + h^{(r)}_0 \tag{4}$$

Furthermore, for each r as above and $i \in [t]$, there is a homogeneous skew circuit $C^{(r)}_i$ of size at most rd that computes all the polynomials in the set $S^{(r)}_i = \{h^{(\ell)}_{i,j} \mid \ell \in [r], v_\ell \in T, j \in [0, d_\ell - d'_i]\}$ and also a homogeneous skew circuit $C^{(r)}_0$ of size at most rd computing $S^{(r)}_0 = \{h^{(\ell)}_0 \mid \ell \in [r], v_\ell \in T\}$.

We define the polynomials $h^{(r)}_{i,j}$ and $h^{(r)}_0$ by induction on r. The skew circuits computing $S^{(r)}_i$ ($i \in [0, t]$) can be constructed easily from these definitions.

The base case of the induction $r = 1$ is trivial since the polynomial f_1 is just a variable or a scalar from $\mathbb{F}$: we simply set $h^{(1)}_0 = f_1$ and $h^{(1)}_{i,j} = 0$. We can use this reasoning in general when f_r is a variable or a scalar.

So we consider the inductive case. Say $r > 1$. If $v_r \notin T$, then again there is nothing to prove. So assume $v_r \in T$. Consider the following cases:

- v_r **is a non-skew gate**: Since $v_r \in T$, it must be that v_r is a top-most non-skew gate. Hence, $f_r = g_{i_0}$ for some $i_0 \in [t]$. Hence, we can set $h^{(r)}_{i_0,0} = 1$ and all the other $h^{(r)}_{i,j}$ and $h^{(0)}$ to 0.

- v_r **is a skew multiplication gate**: We have $v_r = v_p \times v_q$ where at least one of f_p and f_q is either a variable or scalar. We assume that f_p is such a polynomial (the other case is similar). Since $v_r \in T$ and skew, we see that $v_q \in T$ as well. Hence, we can apply the induction hypothesis to v_q as well. Having done so, we can choose the $h^{(r)}_{i,j}$ like in Lemma 8 as follows:

$$h^{(r)}_{i,j} = \begin{cases} f_p \cdot h^{(q)}_{i,j} & \text{if } d_p = 0, \\ f_p \cdot h^{(q)}_{i,j-1} & \text{if } d_p = 1 \text{ and } j \in [d_r - d'_i], \\ 0 & \text{if } d_p = 1 \text{ and } j = 0. \end{cases}$$

  We also choose $h^{(r)}_0 = f_p \cdot h^{(q)}_0$.

- v_r **is a sum gate**: Since C is homogeneous, we know that $v_r = v_p + v_q$ for $p, q < r$ such that $d_p = d_q = d_r$. Moreover, since $v_r \in T$ and is not a non-skew gate, we see that $v_p, v_q \in T$ as well. Hence, we can apply the induction hypothesis to these gates and set $h_{i,j}^{(r)} = h_{i,j}^{(p)} + h_{i,j}^{(q)}$ and $h_0^{(r)} = h_0^{(p)} + h_0^{(q)}$.

□

6.2 More partitions w.r.t. which small skew circuits are low rank

For any $n \in \mathbb{N}^+$ and $\theta \in \mathbb{R}$, we use $\exp_n(\theta)$ to denote n^θ.

Definition 14. *We say that a partition $\Pi = (Y, Z)$ of $[d]$ is a $(d_1, d_2, \ell_1, \ell_2)$-*extended left right monochromatic *($(d_1, d_2, \ell_1, \ell_2)$-XLRM)* partition *if $[d_1 + \ell_1] \cup [d - d_2 - \ell_2 + 1, d - \ell_2] \subseteq Y$.*

Given below is an example of a $(d_1, d_2, \ell_1, \ell_2)$-XLRM partition.

Figure 7: Extended left-right monochromatic (XLRM) partitions

Lemma 15 (Generalization of Lemma 10)**.** *Let $f \in \mathbb{F}\langle X \rangle$ be a homogeneous polynomial of degree $d \in \mathbb{N}$ computed by a homogeneous skew circuit C of size s. Let $\Pi = (Y, Z)$ be a $(d_1, d_2, \ell_1, \ell_2)$-XLRM partition, where d_1, d_2, ℓ_1, ℓ_2 are positive integers such that $8|d_1$, $8|d_2$, $\ell_2 \le \ell_1$, and $d \ge d_1 + d_2 + \ell_1 + \ell_2$. Then*

$$\text{rel-rank}(f, \Pi) \le (sd)^{2+O(\frac{\ell_2}{D})} \cdot \exp_n\left\{-\Omega\left(\min\left\{d_1, d_2, \frac{d_1 D}{\ell_2}\right\}\right)\right\}.$$

where D denotes $\min\{d_1, d_2\}$.

We will only apply the above lemma when $d_2 = \Theta(d_1)$ and $\ell_2 = O(d_1)$, in which case the upper bound on rel-rank(f, Π) is $(sd)^{O(1)} \cdot \exp_n(-\Omega(d_1))$.

Proof. Fix d_1, d_2 as in the statement of the lemma. We want to bound the quantity $\rho(d, \ell_1, \ell_2) := \sup_{f,\Pi} \text{rel-rank}(f, \Pi)$ where f, Π are chosen as above.

More generally for (possibly negative) integers p_1, p_2 such that $p_i > -d_i$ (for $i \in [2]$) and $e \in \mathbb{N}$ such that $d_1 + d_2 + p_1 + p_2 \le e$, we can define $\rho(e, p_1, p_2) := \sup_{F,\Pi'} \text{rel-rank}(F, \Pi')$, where F is chosen to be a homogeneous polynomial of degree e computed by a skew circuit of size at most s and $\Pi' = (U, V)$ is a partition of $[e]$ such that $U \subseteq [e]$ satisfies $[d_1 + p_1] \cup [e - d_2 - p_2 + 1, \min\{e, e - p_2\}] \subseteq U$. When $p_2 \ge 0$, this is just the condition that $[d_1 + p_1] \cup [e - d_2 - p_2 + 1, e - p_2] \subseteq U$. Moreover, when $p_2 \le 0$, the partition Π' is an LRM partition.

We will be interested in reducing the problem of bounding rel-rank(f, Π) to bounding $\rho(e, p_1, p_2)$ for $p_2 \le 0$, because in this setting we have an LRM partition and hence by Lemma 10, we immediately get

Claim 16. *Fix e, p_1, p_2 as above and further assume $p_2 \leq 0$ and $d_1 + p_1, d_2 + p_2 > 0$. Then $\rho(e, p_1, p_2) \leq (sd) \cdot \exp_n\{-\Omega(\min\{d_1 + p_1, d_2 + p_2\})\}$.*

To reduce to the above setting, we will prove the following claim:

Claim 17. *Fix any $p_1 \in \mathbb{Z}$ and $e, p_2 \in \mathbb{N}$ such that $p_1 \geq -d_1/2$ and $e \geq d_1 + d_2 + p_1 + p_2$. For any $b, a \in \mathbb{N}$ such that $b \leq a \leq D/2$ we have*

$$\rho(e, p_1, p_2) \leq s \cdot (a+1) \cdot n^{-b} + s \cdot (a+1) \cdot \max_{\substack{a_1, a_2: \\ a_1 + a_2 = a \\ a_1 \leq a_2 + b}} \rho(e - a, p_1 - a_1, p_2 - a_2).$$

We will prove the claim below. First, we use the claim to finish the proof of the lemma. The idea is to apply the above claim repeatedly until we can apply Lemma 10.

We start by defining a few parameters. Let $a = D/2$, $N = 1 + \lfloor 4\ell_2/a \rfloor$, and $b = \min\{\lceil d_1/4N \rceil, a/4\}$.

Define $e^{(0)} = d$, $p_1^{(0)} = \ell_1$ and $p_2^{(0)} = \ell_2$. It can be checked that Claim 17 can be applied with $e = e^{(0)}, p_1 = p_1^{(0)}$, and $p_2 = p_2^{(0)}$. We thus can apply Claim 17 and use the fact that $a + 1 \leq d$ to obtain

$$\rho(e^{(0)}, p_1^{(0)}, p_2^{(0)}) \leq sd \cdot n^{-b} + sd \cdot \rho(e^{(0)} - a, p_1^{(0)} - a_1^{(0)}, p_2^{(0)} - a_2^{(0)}).$$

for some $a_1^{(0)}, a_2^{(0)} \in \mathbb{N}$ such that $a_1^{(0)} + a_2^{(0)} = a$ and $a_1^{(0)} \leq a_2^{(0)} + b$. Define $e^{(1)} = e^{(0)} - a$, $p_1^{(1)} = p_1^{(0)} - a_1^{(0)}$, and $p_2^{(1)} = p_2^{(0)} - a_2^{(0)}$. Note that we still have $e^{(1)} \geq d_1 + d_2 + p_1^{(1)} + p_2^{(1)}$. Also note that $p_1^{(1)} \geq p_2^{(1)} - b$.

We repeat the above process to define $e^{(i+1)}, p_1^{(i+1)}, p_2^{(i+1)}$ from $e^{(i)}, p_1^{(i)}, p_2^{(i)}$ as long as $p_1^{(i)} \geq -d_1/2$ and $p_2^{(i)} > 0$. After $i \geq 1$ such iterations, we have

$$\rho(e^{(0)}, p_1^{(0)}, p_2^{(0)}) \leq \left(\sum_{j=1}^{i} (sd)^{j+1} \right) \cdot n^{-b} + (sd)^i \cdot \rho(e^{(i)}, p_1^{(i)}, p_2^{(i)}) \tag{5}$$

Note, moreover, that we also have the following

$$e^{(i)} = e^{(i-1)} - a, \qquad p_1^{(i)} \geq p_2^{(i)} - ib, \qquad p_2^{(i-1)} - a \leq p_2^{(i)} = p_2^{(0)} - \sum_{j<i} a_2^{(j)} \leq p_2^{(0)} - ia/4 \tag{6}$$

For the last two inequalities, we have used the fact that for any $j < i$, $a_2^{(j)} \leq a$; and also $a = a_1^{(j)} + a_2^{(j)} \leq 2a_2^{(j)} + b$ and since $b \leq a/4$, we have $a_2^{(j)} \geq a/4$.

We terminate the process when either $p_1^{(i)} < -d_1/2$ or $p_2^{(i)} \leq 0$. Let i_0 be the number of iterations in the above procedure. It can be verified using the last inequality in (6) that $p_2^{(i)} \leq 0$ after at most N iterations and hence $i_0 \leq N$. Furthermore, we have

$$p_1^{(i_0)} = p_1^{(i_0-1)} - a_1^{(i_0)} \geq p_2^{(i_0-1)} - (i_0 - 1)b - a \geq -(N-1)b - a \geq -d_1/4 - d_1/4 \geq -d_1/2$$

where the first inequality follows from (6) and the fact that $a_1^{(i_0)} \leq a$, the second from the fact that $p_2^{(i_0-1)} \geq 0$ (by the definition of i_0) and $i_0 \leq N$ and the third by the definitions of a, b, and N above.

Thus, after $i_0 \leq N$ iterations, we will have $p_2^{(i_0)} \leq 0$ and $p_1^{(i_0)} \geq -d_1/2$; furthermore, by (6), we have $p_2^{(i_0)} \geq p_2^{(i_0-1)} - a \geq -a \geq -d_2/2$. By Claim 16, we have $\rho(e^{(i_0)}, p_1^{(i_0)}, p_2^{(i_0)}) \leq (sd) \cdot \exp_n\{-\Omega(\min\{d_1 + p_1^{(i_0)}, d_2 + p_2^{(i_0)}\})\}$. Since $p_1^{(i_0)} \geq -d_1/2$ and $p_2^{(i_0)} \geq p_2^{(i_0-1)} - a \geq -d_2/2$, we have $d_1 + p_1^{(i_0)} \geq d_1/2$. Hence, we have shown that

$$\rho(e^{(i_0)}, p_1^{(i_0)}, p_2^{(i_0)}) \leq (sd) \cdot \exp_n\{-\Omega(\min\{d_1, d_2\})\}$$

Plugging the above into (5) with $i = i_0$, we have

$$\begin{aligned}
\rho(e^{(0)}, p_1^{(0)}, p_2^{(0)}) &\leq \left(\sum_{j=1}^{i_0} (sd)^{j+1}\right) \cdot n^{-b} + (sd)^{i_0+1} \cdot \exp_n\{-\Omega(\min\{d_1, d_2\})\} \\
&\leq (sd)^{N+1} \cdot \exp_n\{-\Omega(\min\{d_1, d_2, b\})\} \\
&\leq (sd)^{2+O(\ell_2/D)} \cdot \exp_n\left\{-\Omega\left(\min\left\{d_1, d_2, \frac{d_1 D}{\ell_2}\right\}\right)\right\}.
\end{aligned}$$

as claimed. For the last inequality, we have used the fact that $b = \min\{a/4, \lceil d_1/4N \rceil\} = \Omega(\min\{d_1, d_2, (d_1 D/\ell_2)\})$. □

6.2.1 Proof of Claim 17

We have $\rho(e, p_1, p_2) = \sup_{F,\Pi'} \text{rel-rank}(F, \Pi')$ where F and $\Pi' = (U, V)$ are as mentioned above. We apply Lemma 8 to F with $d' = e - a$ to get a decomposition of the form $F = \sum_{i=1}^{t} \sum_{j \in [0,a]} g_i \times_j h_{i,j}$, where $t \leq s$ and the polynomials g_i ($i \in [t]$) are intermediate polynomials computed by the circuit computing F and hence have skew circuits of size at most s. By the subadditivity of matrix rank, we have

$$\text{rel-rank}(F, \Pi') \leq \sum_{i=1}^{t} \sum_{j \in [0,a]} \text{rel-rank}(g_i \times_j h_{i,j}, \Pi') \tag{7}$$

Fix any $(i, j) \in [t] \times [0, a]$. We upper bound $\text{rel-rank}(g_i \times_j h_{i,j}, \Pi')$ based on j as follows.

We first consider the case when j is *large*: more precisely, we assume that $j \geq (a - j) + b$. In this case, we use the following consequence of Corollary 7:

$$\text{rel-rank}(g_i \times_j h_{i,j}, \Pi') \leq n^{-(|U_h| - |V_h|)}, \tag{8}$$

where $U_h = \texttt{Collapse}(U \setminus [j+1, j+d'], [e] \setminus [j+1, j+d'])$ and $V_h = [a] \setminus U_h$. Since $U \supseteq [d_1 + p_1] \supseteq [d_1/2] \supseteq [D/2] \supseteq [a] \supseteq [j]$, we have $|U_h| \geq j$ and consequently, $|V_h| = a - |U_h| \leq (a - j) \leq j - b$ by our assumption that j is large. Thus, $|U_h| - |V_h| \geq b$ and hence (8) implies that $\text{rel-rank}(g_i \times_j h_{i,j}, \Pi') \leq n^{-b}$.

Now consider the case when j is small: i.e., $j \leq (a - j) + b$. In this case, by Corollary 7, we have

$$\text{rel-rank}(g_i \times_j h_{i,j}, \Pi') \leq \text{rel-rank}(g_i, \Pi'_g) \tag{9}$$

for $\Pi'_g = (U_g, V_g)$ where $U_g = \texttt{Collapse}(U \cap [j+1, j+d'], [j+1, j+d'])$. Note that g_i is a homogeneous polynomial of degree $e - a$ computed by a skew circuit of size at most s. Furthermore, since $U \supseteq [d_1 + p_1] \cup [e - d_2 - p_2 + 1, e - p_2]$, it follows from some simple analysis that

$$U_g \supseteq [d_1 + p_1 - j] \cup [(e - a) - d_2 - (p_2 - (a - j)) + 1, \min\{e - a, e - a - (p_2 - (a - j))\}]$$

Thus, it follows that

$$\text{rel-rank}(g_i, \Pi'_g) \leq \rho(e-a, p_1-j, p_2-(a-j)) \leq \max_{\substack{a_1,a_2: \\ a_1+a_2=a \\ a_1 \leq a_2+b}} \rho(e-a, p_1-a_1, p_2-a_2)$$

where for the last inequality we have used the fact that j is small.

Hence, we have shown that for every i, j,

$$\text{rel-rank}(g_i \times_j h_{i,j}, \Pi') \leq n^{-b} + \max_{\substack{a_1,a_2: \\ a_1+a_2=a \\ a_1 \leq a_2+b}} \rho(e-a, p_1-a_1, p_2-a_2).$$

The claim now follows from (7).

6.3 The candidate hard partition for circuits of non-skew depth at most k

Throughout, let $d_0 \in \mathbb{N}^+$ be a fixed parameter.

Let $d \in \mathbb{N}$. Given an (ordered) partition $\Pi = (Y, Z)$ of $[d]$, we define the *signature of* Π to be the sequence $\text{sgn}(\Pi) = \sigma = (i_1, i_2, \ldots, i_p)$ of non-negative integers such that the first i_1 elements of $[d]$ belong to Y, the next i_2 elements belong to Z, the next i_3 again to Y, and so on. Formally,

$$Y = \bigcup_{q \text{ odd}} [\sum_{j<q} i_j + 1, \sum_{j \leq q} i_j].$$

We denote by $|\sigma|$ the quantity $\sum_{q \leq p} i_q = d$ and use $|\sigma|_0$ to denote p.

Given two signatures $\sigma_1 \in \mathbb{N}^n$ and $\sigma_2 \in \mathbb{N}^m$, we use $\sigma_1 \circ \sigma_2 \in \mathbb{N}^{m+n}$ to denote their concatenation. We also use σ_1^r to denote the r-fold repetition of σ_1.

Given a signature $\sigma = (i_1, \ldots, i_p)$, we say that a signature τ is a *prefix of* σ if $\tau = (i'_1, \ldots, i'_q)$ for $q \leq p$, where $i'_j = i_j$ for $j < q$ and $i'_q \leq i_q$.

Clearly, we may define a partition Π of $[d]$ using its signature. For any $k \in \mathbb{N}$, we now define a partition $\Pi_k = (Y_k, Z_k)$ of $[d]$ (for suitable d) such that small circuits of non-skew depth at most k computing a homogeneous polynomial of degree d have low rank w.r.t. Π_k.

Fix any $k \in \mathbb{N}$ and let $D_k = 8d_0 + 12d_0k$. We define the partition $\Pi_k = (Y_k, Z_k)$ of $[D_k]$ so that

$$\text{sgn}(\Pi_k) = (3(k+1)d_0, 2d_0) \circ (d_0, 2d_0)^{1+3k}$$

Note that $|Y_k| = |Z_k| = D_k/2$. Figure 8 illustrates the partition Π_0 and also the relation between the partitions Π_k and Π_{k-1}, which will be important in our lower bound.

We will later show that small circuits of non-skew depth at most k computing a homogeneous polynomial of degree D_k cannot compute a polynomial that has high relative rank w.r.t. Π_k. In the remainder of this section, we show that there are small circuits of non-skew depth $O(k)$ (in fact, circuits using only $O(k)$ many non-skew gates) that can compute a homogeneous polynomial f_k of degree D_k that has *full rank* w.r.t. Π_k. The basic 'gadget' in this construction is the palindrome polynomial, and the construction of f_k involves 'wrapping' a copy of $\mathsf{PAL}_{D_k/4}(X)$ around $O(k)$ copies of $\mathsf{PAL}_{d_0}(X)$.

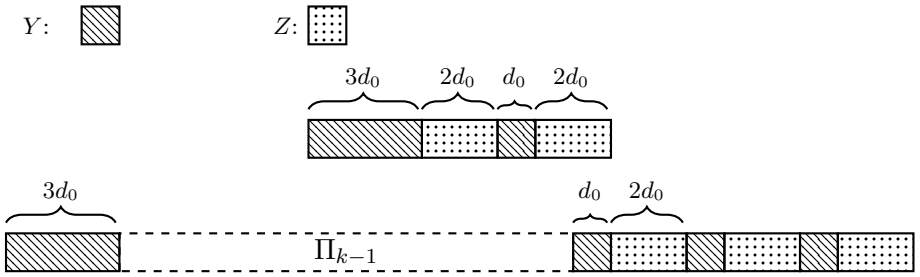


Figure 8: The partition Π_0 (above) and constructing Π_k from Π_{k-1} (below)

Lemma 18. *Fix any positive integers k, d_0 and let D_k be as above. Then, there is a homogeneous polynomial $f_k \in \mathbb{F}\langle X\rangle$ of degree D_k that is computable by a non-commutative arithmetic circuit of size $O(nD_k)$ with $O(k)$ many non-skew gates and s.t.* rel-rank$(f_k, \Pi_k) = 1$.

Proof. We define the polynomials f_k inductively. For $k = 0$, we define

$$f_0 := (\mathsf{PAL}_{2d_0}(X) \cdot \mathsf{PAL}_{d_0}(X)) \times_{d_0} \mathsf{PAL}_{d_0}(X)$$

In the notation of Section 4.1, we can write f_0 as

$$f_0 = \sum_{w_1,w_2,w_3,w_4 \in [n]^{d_0}} \tilde{x}_{w_1} \cdot \tilde{x}_{w_2} \cdot \tilde{x}_{w_3} \cdot \tilde{x}_{w_3^R} \cdot \tilde{x}_{w_2^R} \cdot \tilde{x}_{w_4} \cdot \tilde{x}_{w_4^R} \cdot \tilde{x}_{w_1^R}$$

Figure 9 illustrates the positioning of the segments of the monomial corresponding to w_1, w_2, w_3, and w_4 w.r.t. the partition Π_0.

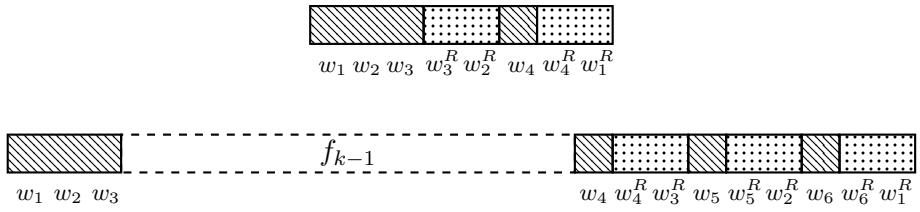


Figure 9: The construction of polynomials f_0 (above) and f_k from f_{k-1} (below)

We observe that f_0 can be computed by a homogeneous non-commutative arithmetic circuit of size $O(nD_0) = O(nd_0)$ with exactly one non-skew gate. To see this, note that $g_0 := (\mathsf{PAL}_{2d_0}(X) \cdot \mathsf{PAL}_{d_0}(X))$ can be computed by first computing each of the terms of the product using homogeneous skew circuits of size $O(nd_0)$ and then multiplying them using exactly one non-skew gate. We can then compute f_0 by using g_0 and only homogeneous skew multiplication gates by using the following inductive definitions:

$$g_0^{(0)} := g_0$$
$$g_0^{(i+1)} := \sum_{j=1}^{n} x_j \cdot g_0^{(i)} \cdot x_j$$

The polynomial $g_0^{(d_0)}$ is exactly f_0. Note that computing $g_0^{(i+1)}$ from $g_0^{(i)}$ requires only $O(n)$ additional gates. Thus, the size of the circuit computing f_0 is $O(nd_0)$.

For $k > 0$, we define the polynomial f_k inductively as follows. The construction is illustrated in Figure 9.

$$f_k := \sum_{w_1,w_2,w_3,w_4,w_5,w_6 \in [n]^{d_0}} (\tilde{x}_{w_1}\tilde{x}_{w_2}\tilde{x}_{w_3}) \cdot f_{k-1} \cdot (\tilde{x}_{w_4}\tilde{x}_{w_4^R}) \cdot \tilde{x}_{w_3^R} \cdot (\tilde{x}_{w_5}\tilde{x}_{w_5^R}) \cdot \tilde{x}_{w_2^R} \cdot (\tilde{x}_{w_6}\tilde{x}_{w_6^R}) \cdot \tilde{x}_{w_1^R}$$

It can be easily checked that the matrix $M[f_k, \Pi_k]$ is an $n^{D_k/2} \times n^{D_k/2}$ permutation matrix and hence rel-rank$(f_k, \Pi_k) = 1$.

We need to check that f_k defined as above has a small non-commutative circuit with $O(k)$ many non-skew gates. For $k \geq 1$, we define

$$h_k := (f_{k-1} \cdot \mathsf{PAL}_{d_0}(X)) \times_{d_0} \mathsf{PAL}_{d_0}(X)$$
$$g_k := (h_k \cdot \mathsf{PAL}_{d_0}(X)) \times_{d_0} \mathsf{PAL}_{d_0}(X)$$

Note that

$$f_k = (g_k \cdot \mathsf{PAL}_{d_0}(X)) \times_{d_0} \mathsf{PAL}_{d_0}(X)$$

The circuit for h_k is obtained from the circuit for f_{k-1} in a manner similar to the construction of the circuit for f_0, and similarly, we can obtain a circuit for g_k and then a circuit for f_k. We omit the details. It is easy to check that only 3 additional non-skew multiplication gates are used by the above procedure and hence the number of non-skew gates used overall is $O(k)$. □

6.4 The lower bound for circuits of non-skew depth k

In this section, we show that small non-commutative circuits of non-skew depth k computing a homogeneous polynomial of degree D_k cannot compute a polynomial that has high relative rank w.r.t. Π_k. Throughout, let $d_0 \in \mathbb{N}$ be a fixed parameter.

For $\ell \in \mathbb{N}^+$, we say that a pair (g, Π) is *ℓ-good* if $g \in \mathbb{F}\langle X \rangle$ is a homogeneous polynomial with $\deg(g) = D \geq D_\ell$ and $\Pi = (Y, Z)$ is a partition of $[D]$ such that $\mathrm{sgn}(\Pi) = (a, 2d_0) \circ (d_0, 2d_0)^{1+3\ell+r} \circ (b, c)$ where

- $a \geq 3(\ell+1)d_0$, $r \geq 0$, and
- either $c = 0$ and $b \in [d_0]$ or $b = d_0$ and $c \in [2d_0 - 1]$.

Intuitively, the (g, Π) being ℓ-good means that $D \geq D_\ell$ and Π 'contains' a copy of Π_ℓ as a sub-segment and Π is furthermore similarly contained in $\Pi_{\ell'}$ for some $\ell' \geq \ell$. See Figure 10, where the top partition corresponds to the case $c = 0$ and the bottom one to the case $b = d_0$ as mentioned above.

The main lemma is the following:

Lemma 19 (Main Lemma for circuits of non-skew depth k)**.** *Assume $k, d_0 \in \mathbb{N}$ such that $64|d_0$. Let $f \in \mathbb{F}\langle X \rangle$ be any homogeneous polynomial of degree D_k computed by a non-commutative circuit C of size at most s with non-skew depth at most k and let $\Pi_k = (Y_k, Z_k)$ be the partition defined above. Then,* rel-rank$(f, \Pi_k) \leq (sD_k)^{O(1)} \cdot n^{-\Omega(d_0)}$.

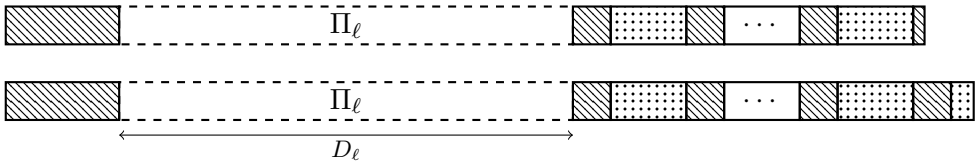


Figure 10: Partitions that arise in ℓ-good pairs

The basic idea of the proof is to repeatedly use Lemma 13 to decompose the polynomial f as a sum of polynomials computed by circuits with smaller non-skew depth. When we apply Lemma 13, we repeatedly obtain polynomials of the form $g \times_j h$ where g and h are homogeneous polynomials of degree d_g and $D_k - d_g$ respectively and $j \in [0, D_k - d_g]$. Given a polynomial $g \in \mathbb{F}\langle X \rangle$, $j \in [0, D_k - d_g]$, and $\ell \in [0, k]$, we say that the pair (g, j) is *ℓ-admissible* if the pair (g, Π_g) is ℓ-good, where $\Pi_g = (Y_g, Z_g)$ for $Y_g := \texttt{Collapse}(Y_k \cap [j+1, j+d_g], [j+1, j+d_g])$ and $Z_g := [d_g] \setminus Y_g$. See Figure 11.

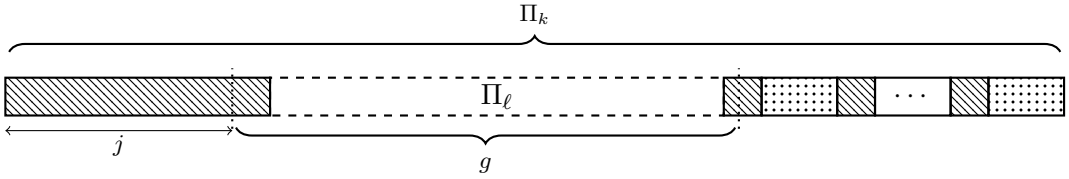


Figure 11: Example of an ℓ-admissible pair (g, j)

Proof. First let us introduce some notation. Let the non-skew depth of a node v of C be the maximum number of non-skew gates on any path from a leaf to v. For $\ell \in [k]$, let G_ℓ (resp. $G_{=\ell}$) be the set of all polynomials computed by gates in the circuit that have non-skew depth at most ℓ (resp. exactly ℓ); note that $|G_{=\ell}| \leq |G_\ell| \leq s$. We also denote by A_ℓ the set $\{(g, j) \mid g \in G_\ell \text{ and } (g, j) \text{ is } \ell\text{-admissible}\}$. Finally, we define V_ℓ by

$$V_\ell = \{ \sum_{(g,j) \in A_\ell} g \times_j H_j^g \mid H_j^g \in \mathbb{F}\langle X \rangle \text{ homogeneous of degree exactly } (D_k - \deg(g)) \}$$

Note that $V_\ell \subseteq \mathbb{F}\langle X \rangle$ is a vector space over $\mathbb{F}$.

Our proof proceeds in two steps:

1. We first show that for each $\ell \in [0, k]$, the polynomial f can be decomposed as $f = p_\ell + e_\ell$ where $p_\ell \in V_\ell$ and e_ℓ is such that rel-rank(e_ℓ, Π_k) is small. The proof is by downward induction on ℓ.

2. We then show that rel-rank(p_0, Π_k) is small for each $p_0 \in V_0$. Along with the above decomposition, this will finish the proof.

We start with 1. above. Formally, we prove that there are absolute constants $\alpha, \beta > 0$ such that for each $\ell \in [0, k]$, the polynomial f can be written as

$$f = p_\ell + e_\ell \tag{10}$$

where $p_\ell \in V_\ell$ and $e_\ell \in \mathbb{F}\langle X \rangle$ is homogeneous of degree D_0 and satisfies

$$\text{rel-rank}(e_\ell, \Pi_k) \leq (sD_k)^\alpha \cdot (k - \ell) \cdot n^{-\beta d_0}. \tag{11}$$

The proof is by downward induction on ℓ. We will choose α, β so that they satisfy some constraints that come up during the course of the proof. The base case when $\ell = k$ is trivial, since we can choose $p_k = f \in V_k$ and e_k to be the zero polynomial. Both (10) and (11) are thus satisfied for any choice of α, β.

Now for the induction case. Say that $\ell \in [0, k-1]$. By the induction hypothesis we have $f = p_{\ell+1} + e_{\ell+1}$, where $p_{\ell+1} \in V_{\ell+1}$ and $\text{rel-rank}(e_{\ell+1}, \Pi_k) \leq (sD_k)^\alpha \cdot (k - \ell - 1) \cdot n^{-\beta d_0}$. By the definition of $V_{\ell+1}$, we know that

$$\begin{aligned} p_{\ell+1} &= \sum_{(g,j) \in A_{\ell+1}} g \times_j H_j^g \\ &= \sum_{(g,j) \in A'_{\ell+1}} g \times_j H_j^g + \underbrace{\sum_{(g,j) \in A_\ell} g \times_j H_j^g}_{p'_{\ell+1} \in V_\ell} \end{aligned} \tag{12}$$

where $A'_{\ell+1} := A_{\ell+1} \setminus A_\ell = \{(g,j) \mid (g,j) \text{ is } \ell+1\text{-admissible and } g \in G_{=\ell+1}\}$. (Here, we have used the fact that if (g,j) is $(\ell+1)$-admissible and $g \in G_\ell$, then (g,j) is also ℓ-admissible.)

As noted above, the terms corresponding to $(g,j) \in A_\ell$ already sum to a polynomial $p'_{\ell+1} \in V_\ell$. To prove the induction statement (10), it therefore suffices to decompose each polynomial $g \times_j H_j^g$ where $(g,j) \in A'_{\ell+1}$. To do this, we need the following claim, whose proof is deferred:

Claim 20. *Fix any $\ell \in [k]$. Also fix any $g \in G_{=\ell}$ of degree $d_g \in [D_\ell, D_k]$, any homogeneous polynomial $H \in \mathbb{F}\langle X \rangle$ of degree $D_k - d_g$, and j such that (g,j) is ℓ-admissible. Then, the polynomial $g \times_j H$ can be decomposed as*

$$g \times_j H = p + e$$

where $p \in V_{\ell-1}$ and $e \in \mathbb{F}\langle X \rangle$ is homogeneous of degree D_k and satisfies $\text{rel-rank}(e, \Pi_k) \leq (sD_k)^{O(1)} \cdot n^{-\Omega(d_0)}$.

Applying the above claim (with ℓ replaced by $\ell+1$) to each pair $(g,j) \in A'_{\ell+1}$ from the right hand side of (12), we obtain for each such (g,j) that

$$g \times_j H_j^g = p_j^g + e_j^g$$

where $p_j^g \in V_\ell$ and $\text{rel-rank}(e_j^g, \Pi_k) \leq (sD_k)^{\alpha_1} \cdot n^{-\beta_1 d_0}$ for suitably large $\alpha_1 > 0$ and small $\beta_1 > 0$. Substituting in (12), we get

$$p_{\ell+1} = \underbrace{p'_{\ell+1} + \sum_{(g,j) \in A'_{\ell+1}} p_j^g}_{p_\ell} + \underbrace{\sum_{(g,j) \in A'_{\ell+1}} e_j^g}_{e'_\ell}.$$

Note that $p_\ell \in V_\ell$ (since V_ℓ is a vector space). Also, as $|A'_{\ell+1}| \leq (sD_k)$, we have rel-rank$(e'_\ell, \Pi_k) \leq (sD_k)^{\alpha_1+1} \cdot n^{-\beta_1 d_0} \leq (sD_k)^{\alpha} \cdot n^{-\beta d_0}$ for $\alpha \geq \alpha_1 + 1$ and $\beta \leq \beta_1$.

Setting p_ℓ as above and $e_\ell = e_{\ell+1} + e'_\ell$, we have the required decomposition. The inequality (11) follows since rel-rank$(e_\ell, \Pi_k) \leq$ rel-rank$(e_{\ell+1}, \Pi_k)$ + rel-rank(e'_ℓ, Π_k). This finishes the proof of the induction.

Thus, for $\ell = 0$, we have

$$f = p_0 + e_0$$

for some $p_0 \in V_0$ and rel-rank$(e_0, \Pi_0) \leq k \cdot (sD_k)^{\alpha} \cdot n^{-\beta d_0} \leq (sD_k)^{\alpha+1} \cdot n^{-\beta d_0}$. To bound rel-rank$(f, \Pi_k)$, we only need to bound rel-rank(p_0, Π_k). Since $p_0 \in V_0$, we have

$$p_0 = \sum_{(g,j) \in A_0} g \times_j H_j^g \tag{13}$$

To analyze rel-rank(p_0, Π_k), we will need the following claim, the proof of which is also deferred:

Claim 21. *Assume that $h \in \mathbb{F}\langle X \rangle$ of degree $d_h \in [D_0, D_k]$ is computed by a homogeneous skew circuit of size s_1.*

(a) Let $\Pi_h = (Y_h, Z_h)$ be any partition of $[d_h]$ such that (h, Π_h) is 0-good. Then rel-rank$(h, \Pi_h) \leq (s_1 D_k)^{O(1)} \cdot n^{-\Omega(d_0)}$.

(b) Let $H \in \mathbb{F}\langle X \rangle$ be a homogeneous polynomial of degree $d_H = D_k - d_h$. Given $j \in [0, d_H]$ is such that (h, j) is 0-admissible, we have rel-rank$(h \times_j H, \Pi_k) \leq (s_1 D_k)^{O(1)} \cdot n^{-\Omega(d_0)}$.

Fix $(g, j) \in A_0$ and consider the polynomial $g \times_j H_j^g$ in the right hand side of (13). By Claim 21 and using the fact that g is computable by a skew circuit of size at most s, we know that rel-rank$(g \times_j H_j^g, \Pi_k) \leq (sD_k)^{O(1)} \cdot n^{-\Omega(d_0)}$. Thus, we have

$$\begin{aligned}
\text{rel-rank}(f, \Pi_k) &\leq \text{rel-rank}(p_0, \Pi_k) + \text{rel-rank}(e_0, \Pi_k) \\
&\leq \sum_{(g,j) \in A_0} \text{rel-rank}(g \times_j H_j^g, \Pi_k) + \text{rel-rank}(e_0, \Pi_k) \\
&\leq (sD_k)^{O(1)} \cdot n^{-\Omega(d_0)}
\end{aligned}$$

which finishes the proof of the lemma. □

It remains to prove the two claims used in the proof of Lemma 19. We prove Claim 21 first and then Claim 20.

Proof of Claim 21. We first prove Part (a) of the claim. Since (h, Π_h) is 0-good, we have $\text{sgn}(\Pi_h) = (a, 2d_0) \circ (d_0, 2d_0)^{1+r} \circ (b, c)$, for $a \geq 3d_0, r \geq 0$ and b, c such that either $c = 0$ and $b \in [d_0]$ or $b = d_0$ and $c \in [2d_0 - 1]$.

We need to show that

$$\text{rel-rank}(h, \Pi_h) \leq (s_1 D_k)^{O(1)} \cdot n^{-\Omega(d_0)}, \tag{14}$$

We divide the analysis into the following cases (see also Figure 12).

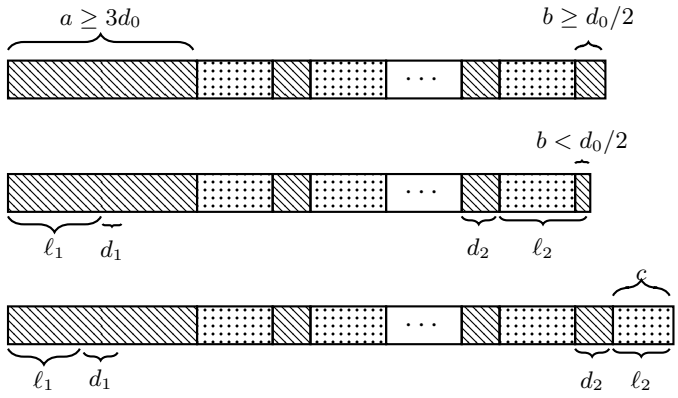


Figure 12: Cases from Claim 21

- $c = 0$ and $b \geq d_0/2$: In this case, we can apply Lemma 10 with $d_1 = 3d_0$ and $d_2 = d_0/2$ to get (14).
- $c = 0$ and $b < d_0/2$: In this case, we apply Lemma 15 with $d_1 = d_0/2$, $d_2 = d_0$, $\ell_1 = 5d_0/2$, and $\ell_2 = b + 2d_0 < 5d_0/2$. Note that $Y \supseteq [3d_0] \cup [d - b - 3d_0 + 1, d - b - 2d_0] = [d_1 + \ell_1] \cup [d - d_2 - \ell_2 + 1, d - \ell_2]$ and hence Lemma 15 implies (14).
- $b = d_0$ and $c > 0$: We apply Lemma 15 with parameters $d_1 = d_2 = d_0$, $\ell_1 = 2d_0$, and $\ell_2 = c < 2d_0$, which gives (14).

Part (b) of the claim follows from Part (a) as follows. Let $Y_h := \texttt{Collapse}(Y_k \cap [j+1, j+d_h], [j+1, j+d_h])$, $Z_h := [d_h] \setminus Y_h$, and $\Pi_h := (Y_h, Z_h)$. Since (h, j) is 0-admissible, we know that (h, Π_h) is 0-good. By Corollary 7, we have

$$\text{rel-rank}(h \times_j H, \Pi_k) \leq \text{rel-rank}(h, \Pi_h) \leq (s_1 D_k)^{O(1)} \cdot n^{-\Omega(d_0)}$$

where the last inequality follows from Part 1. □

Proof of Claim 20. Let $Y_g := \texttt{Collapse}(Y_k \cap [j+1, j+d_g], [j+1, j+d_g])$. Also define $Z_g := [d_g] \setminus Y_g$ and $\Pi_g := (Y_g, Z_g)$. Since (g, j) is ℓ-admissible, we know that (g, Π_g) is ℓ-good.

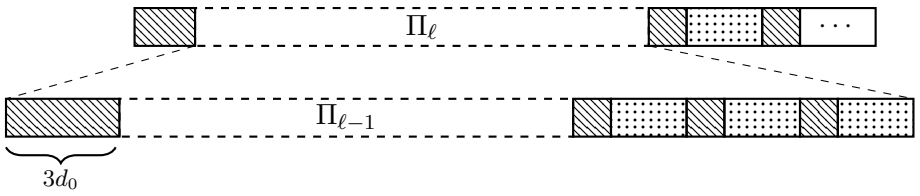


Figure 13: The partition Π_g (above) and the relation between Π_ℓ and $\Pi_{\ell-1}$ (below)

To do this, consider the subcircuit C_g of C that computes g. Since g is at non-skew depth ℓ, we may assume that C_g has non-skew depth ℓ also by removing gates at larger non-skew depths. Recall that C and hence C_g has size at most s.

By applying Lemma 13 to the polynomial g, we can see that

$$g = \sum_{i \in [t]} \sum_{m \in [0, d_g - d_i]} g_i \times_m h_{i,m} + h_0$$

where $g_1, \ldots, g_t$ are the polynomials computed by the top-most non-skew gates in C_g and $d_i = \deg(g_i)$. Further, each of the $h_{i,m}$ and h_0 have skew circuits of size at most $sd_g \leq sD_k$. Thus, we have

$$g \times_j H = \sum_i \sum_{j,m} (g_i \times_m h_{i,m}) \times_j H + h_0 \times_j H \tag{15}$$

We argue that polynomial on the right hand side of (15) either belongs to $V_{\ell-1}$ or has relative rank at most $(sD_k)^{O(1)} \cdot n^{-\Omega(d_0)}$ w.r.t. Y_k. Since $V_{\ell-1}$ is a vector space and rel-rank$(\cdot, \Pi_k)$ is subadditive, this will complete the proof.

We consider first the polynomial $h_0 \times_j H$. Note that (h_0, j) is ℓ-admissible — since (g, j) is — and hence it is also 0-admissible. Moreover, h_0 is computable by a skew circuit of size at most sD_k. Hence, by Claim 21, we have

$$\text{rel-rank}(h_0 \times_j H, \Pi_k) \leq (sD_k)^{O(1)} \cdot n^{-\Omega(d_0)}, \tag{16}$$

which completes the analysis of this term.

Now consider any polynomial $q_{i,m} := (g_i \times_m h_{i,m}) \times_j H$ appearing in (15). For notational simplicity, we let $d'_g := d_i = \deg(g_i)$ and $d'_h := \deg(h_{i,m}) = d_g - d_i$. We will show that either $q_{i,m} \in V_{\ell-1}$ or rel-rank$(q_{i,m}, \Pi_k)$ is small; to prove the latter, we will use the following inequalities which follow from Lemma 6 and Corollary 7:

$$\begin{aligned} \text{rel-rank}(q_{i,m}, \Pi_k) \leq \text{rel-rank}(g_i \times_m h_{i,m}, \Pi_g) &\leq \min\{\text{rel-rank}(g_i, \Pi'_g), \text{rel-rank}(h_{i,m}, \Pi'_h)\} \\ &\leq \min\{n^{-(|Y'_g| - |Z'_g|)}, n^{-(|Y'_h| - |Z'_h|)}\} \end{aligned} \tag{17}$$

where $\Pi'_g = (Y'_g, Z'_g)$ and $\Pi'_h = (Y'_h, Z'_h)$ are the natural restrictions of Π_g to g_i and $h_{i,m}$ respectively. That is, $Y'_g := \texttt{Collapse}(Y_g \cap [m+1, m+d'_g], [m+1, m+d'_g])$, $Y'_h := \texttt{Collapse}(Y_g \setminus [m+1, m+d'_g], [d_g] \setminus [m+1, m+d'_g])$, and Z'_g and Z'_h denote $[d_i] \setminus Y'_h$ and $[d_{i,m}] \setminus Z'_h$ respectively.

Since (g, Π_g) is ℓ-good, we know that $d_g \geq D_\ell$ and, furthermore, we have $\text{sgn}(\Pi_g) = (a, 2d_0) \circ (d_0, 2d_0)^{1+3\ell+r} \circ (b, c)$ where $a \geq 3(\ell+1)d_0, r \geq 0$ and b, c such that either $c = 0$ and $b \in [d_0]$ or $b = d_0$ and $c \in [2d_0 - 1]$.

The upper bound on rel-rank$(q_{i,m}, \Pi_k)$ is based on a case analysis.

1. $m < 5d_0/2$ and $d_g - m - d'_g < b + c + 3rd_0 + 9d_0$: In this case $\text{sgn}(\Pi'_g) = (a_g, 2d_0) \circ (d_0, 2d_0)^{1+3(\ell-1)} \circ \sigma$, where $a_g \geq (3\ell + 1/2)d_0$ and σ is some signature: in particular, $d'_g \geq D_{\ell-1} + d_0/2$. In what follows, we will argue that either g_i has low relative rank w.r.t. Π'_g or $q_{i,m} \in V_{\ell-1}$.

   Since g_i is computed by a top-most non-skew gate in the circuit C_g, we can write $g_i = g_{i,1} \cdot g_{i,2}$ where $g_{i,1}$ and $g_{i,2}$ are homogeneous polynomials computed by homogeneous circuits of size at most s and non-skew depth at most $\ell - 1$. Let e_1 and $e_2 = d'_g - e_1$ denote the degrees of $g_{i,1}$ and $g_{i,2}$ respectively. Let $\Pi'_{g,1} = (Y'_{g,1}, Z'_{g,1})$ and $\Pi'_{g,1} = (Y'_{g,1}, Z'_{g,1})$ be the

induced partitions on $g_{i,1}$ and $g_{i,2}$ respectively: i.e., $Y'_{g,1} = \texttt{Collapse}(Y'_g \cap [e_1], [e_1])$ and $Y'_{g,2} = \texttt{Collapse}(Y'_g \cap [e_1+1, d'_g], [e_1+1, d_g])$.

Our analysis is further divided into two cases depending on e_1:

(i) $e_1 < d_0/2$: In this case, we see that $e_2 = d'_g - e_1 \geq D_{\ell-1}$ and also $\mathrm{sgn}(\Pi'_{g,2}) = (a_g - e', 2d_0) \circ (d_0, 2d_0)^{1+3(\ell-1)} \circ \sigma$. Hence, $(g_{i,2}, \Pi'_{g,2})$ is $(\ell-1)$-good. Thus, the polynomial $q_{i,m}$ — which by Fact 2 can be written as $g_{i,2} \times_{j_2} H_2$ for some homogeneous polynomial H_2 of degree $D_k - d'_{g,2}$ and some j_2 — belongs to $V_{\ell-1}$ and hence we are done.

GM: $(a_g - e', 2d_0)$ should be $(a_g - e_1, 2d_0)$?

(ii) $e_1 \geq d_0/2$: If $\mathrm{sgn}(\Pi'_{g,1}) = (a_g, 2d_0) \circ (d_0, 2d_0)^{1+3(\ell-1)} \circ \sigma'$ for some signature σ', then as in the previous case, we have $q_{i,m} = g_{i,1} \times_{j_1} H_1$ for some suitable H_1 and j_1, and hence $q_{i,m} \in V_{\ell-1}$.

Otherwise, we can use the fact that $\mathrm{sgn}(\Pi'_{g,1})$ must be a prefix of $(a_g, 2d_0) \circ (d_0, 2d_0)^{1+3(k-1)}$ and using the fact that $|\mathrm{sgn}(\Pi'_{g,1})| = e_1 \geq d_0/2$, we see that $|Y'_{g,1}| - |Z'_{g,1}| \geq d_0/2$ and therefore, we have $\text{rel-rank}(g_{i,1}, \Pi'_{g,1}) \leq n^{-(|Y'_{g,1}| - |Z'_{g,1}|)} \leq n^{-\Omega(d_0)}$. By Lemma 6, the same bound holds for $\text{rel-rank}(q_{i,m}, \Pi_k)$ as well.

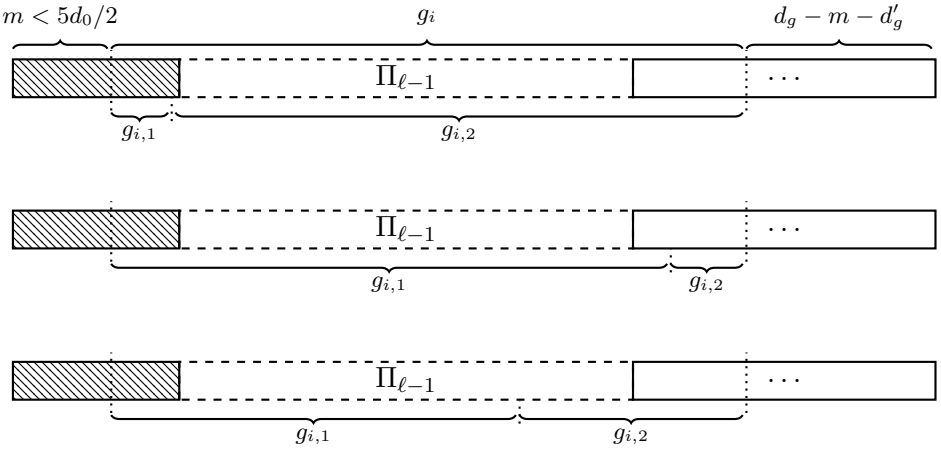


Figure 14: The subcases in Case 1: The first figure represents Case 1(i), and the second and third represent Case 1(ii).

2. $m < 5d_0/2$ but $d_g - m - d'_g \geq b + c + 3rd_0 + 3d_0$: In this case, it can be checked that $\mathrm{sgn}(\Pi_g)$ is a prefix of $(a_g, 2d_0) \circ (d_0, 2d_0)^{1+3(\ell-1)}$ for some $a_g \geq (3\ell + 1/2)d_0$. We analyze in two different ways depending on whether d'_g is reasonably large or not.

   (i) $d'_g \geq d_0/2$: In this case, it follows that no matter what exactly $\mathrm{sgn}(\Pi_g)$ is, we will always have $|Y'_g| - |Z'_g| \geq d_0/2$ and hence by (17), we have $\text{rel-rank}(g_i \times_m h_{i,m}, \Pi'_g) \leq n^{-\Omega(d_0)}$.

   (ii) $d'_g < d_0/2$: In this case, it can be checked that $d'_h \geq D_{\ell-1}$ and $(h, \mathrm{sgn}(\Pi'_h))$ is $(\ell-1, d'_h)$-good and hence also $(0, d'_h)$-good. Thus, we have

$$\begin{aligned}\text{rel-rank}(q_{i,m}, \Pi_k) &= \text{rel-rank}((g_i \times_m h_{i,m}) \times_j H, \Pi_k) = \text{rel-rank}(g_i \times_{j+m} (h_{i,m} \times_j H), \Pi_k) \\ &\leq \text{rel-rank}(h_{i,m}, \Pi'_h) \leq (sD)^{O(1)} \cdot n^{-\Omega(d_0)}\end{aligned}$$

where the second equality uses Fact 2, the first inequality uses two applications of Corollary 7, and the last inequality follows from Part 1 of Claim 21.

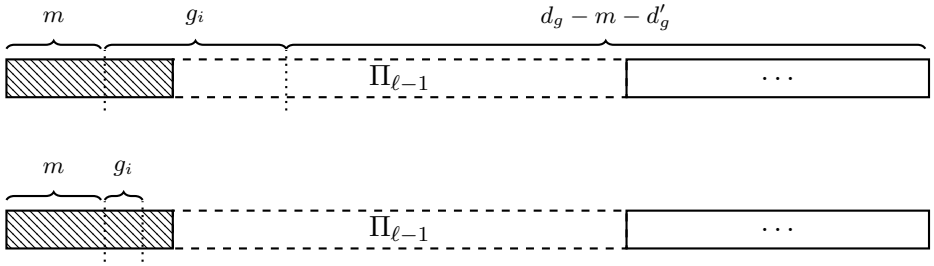


Figure 15: The subcases in Case 2: Case 2(i) above and Case 2(ii) below.

3. $m \in [5d_0/2, a]$: In this case, we show that rel-rank$(h_{i,m}, \Pi'_h) \leq (sD_k)^{O(1)} \cdot n^{-\Omega(d_0)}$. By (17), the same upper bound holds for rel-rank$(q_{i,m}, \Pi_k)$.

   (i) $d_g - d'_g - m < (5/2 - 1/8)d_0$: In this case, we have $|Y'_h| \geq m \geq 5d_0/2$ and $|Z'_h| \leq d_g - d'_g - m < (5/2 - 1/8)d_0$. Thus, we have $|Y'_h| - |Z'_h| \geq d_0/8$ and hence by Corollary 7, rel-rank$(h_{i,m}, \Pi'_h) \leq n^{-\Omega(d_0)}$. So, from now on, we assume that $d_g - d'_g - m \geq (5/2 - 1/8)d_0$. In particular, $d'_h \geq m + (5/2 - 1/8)d_0 \geq (5 - 1/8)d_0$.

   (ii) $c = 0$ and $b < d_0/4$: In this case, it can be checked that $Y'_h \supseteq [5d_0/2] \cup [d'_h - b - 2d_0 - d_0/8 + 1, d'_h - b - 2d_0]$. We apply Lemma 15 with parameters $\ell_1 = (5/2 - 1/4)d_0, d_1 = d_0/4, \ell_2 = 2d_0 + b < \ell_1$, and $d_2 = d_0/8$ to get rel-rank$(h_{i,m}, \Pi'_h) \leq (sD_k)^{O(1)} \cdot n^{-\Omega(d_0)}$.

   (iii) $c = 0$ and $b \geq d_0/4$: In this case, it can be checked that $Y'_h \supseteq [d_0/4] \cup [d'_h - d_0/4 + 1, d'_h]$, and hence Lemma 10 implies that rel-rank$(h_{i,m}, \Pi'_h) \leq (s(d'_h)^2) \cdot n^{-\Omega(d_0)} \leq (sD_k)^2 \cdot n^{-\Omega(d_0)}$.

   (iv) $b = d_0$ and $c \in [2d_0 - 1]$: In this case, we apply Lemma 15 with parameters $\ell_1 = (2 + 1/4)d_0, d_1 = d_0/4, \ell_2 = c < 2d_0 < \ell_1, d_2 = b = d_0$ to get rel-rank$(h_{i,m}, \Pi'_h) \leq (sD)^{O(1)} \cdot n^{-\Omega(d_0)}$.

4. $m > a$: Here again we show that rel-rank$(h_{i,m}, \Pi'_h) \leq (sD_k)^{O(1)} \cdot n^{-\Omega(d_0)}$. Consider the partition $\Pi'_h = (a_1, a_2, \ldots, a_r)$. Each a_t ($t \in [r]$) corresponds to a segment of the partition Π_g. Observe that $a_1 = a \geq 6d_0$ (see Figure 17).

   We will be interested in finding the largest $t \in [r]$ to be the largest odd integer such that $a_t \geq d_0/4$; denote this t by t_0 (note that t_0 may even be 1); in other words, t_0 indexes the right most Y segment that has length at least $d_0/4$. Let $A = \sum_{p > p_0} a_t$. To show that rel-rank$(h_{i,m}, \Pi'_h)$ is small, we apply Lemma 15 with parameters $\ell_1 = 5d_0, d_1 = d_2 = d_0/4$, and $\ell_2 = A$. It can be checked by a further small case analysis that irrespective of the exact value of m, d'_g, b and c, we always have $A < 5d_0$ and hence Lemma 15 is applicable. The lemma gives us rel-rank$(h_{i,m}, \Pi'_h) \leq (sD_k)^{O(1)} \cdot n^{-\Omega(d_0)}$, as required.

□

The main lower bound for non-commutative circuits of small non-skew depth follows.

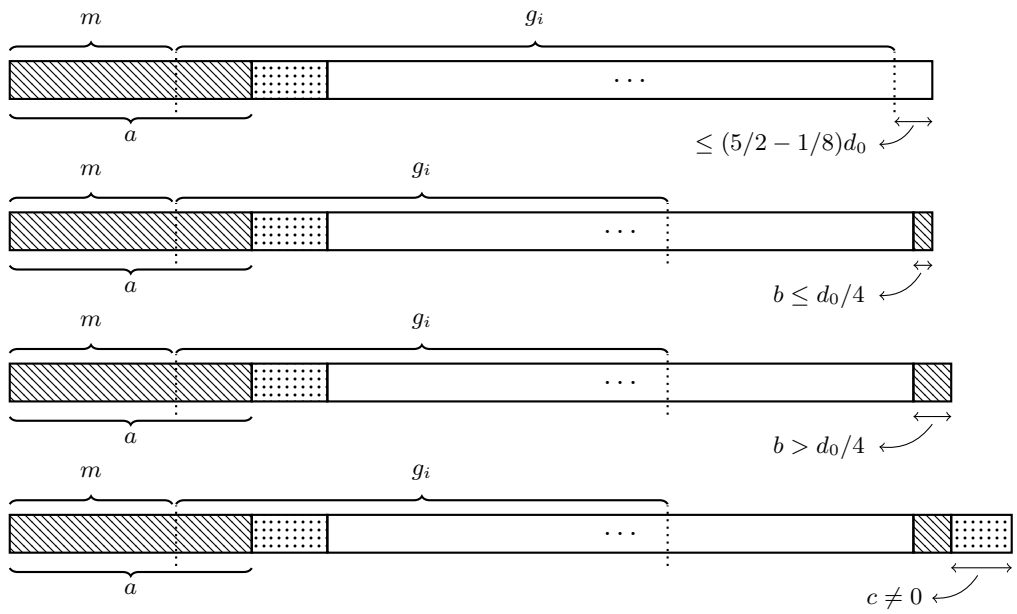


Figure 16: The subcases of Case 3: Case 3(i),3(ii),3(iii), and 3(iv) in order.

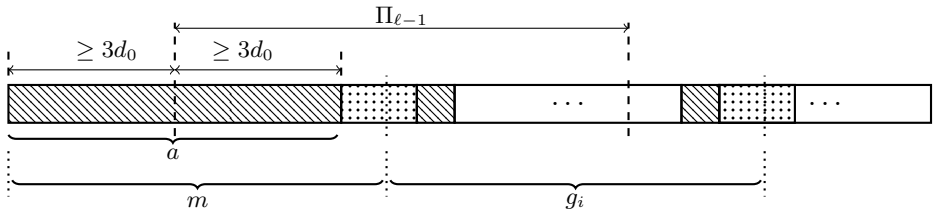


Figure 17: Case 4

Theorem 22 (Lower bound for circuits of non-skew depth k)**.** *Let $k, d \in \mathbb{N}$ be any parameters such that $(8k+12)|d$ and $64|(d/k)$. There is a homogeneous polynomial $f \in \mathbb{F}\langle X\rangle$ of degree d such that f is computable by a homogeneous circuit of size $O(nd)$ with $O(k)$ non-skew gates but any non-commutative circuit of skew depth at most k computing f must have size at least $\tilde{\Omega}(n^{\Omega(d/k)})$, where the $\tilde{\Omega}(\cdot)$ hides* $\mathrm{poly}(d)$ *factors.*

Proof. We let $f = f_k$ as defined above with $d_0 := d/(8k+12)$ (and hence $\deg(f_k) = D_k = d$). By Lemma 18, we know that f is computable by a homogeneous circuit of size $O(nd)$ with $O(k)$ non-skew gates. Moreover, $\text{rel-rank}(f, \Pi_k) = 1$, where $\Pi_k = (Y_k, Z_k)$ is the partition defined in Section 6.3.

Let C be any non-commutative circuit of non-skew depth at most k computing f and let s denote the size of C. By Lemma 5, we know that there is also a homogeneous circuit C' of non-skew depth at most k and size at most $sd^{O(1)}$ computing f. Thus, Lemma 19 implies that $\text{rel-rank}(f, \Pi_k) \leq (sd)^{O(1)} n^{-\Omega(d_0)} = (sd)^{O(1)} n^{-\Omega(d/k)}$. As $\text{rel-rank}(f, \Pi_k) = 1$, we have the required lower bound on s. □

7 Lower bound for the determinant and permanent

Nisan's lower bounds from [14] held not only for the palindrome polynomial seen above, but also for the permanent and the determinant polynomials, because it is easy to see that their partial derivative matrices have high rank. In our case, we could also try to study the rank of the permanent or the determinant, using our version of the partial derivative matrix. However it is simpler to use the fact that the permanent and determinant can easily express the palindrome polynomial.

Recall the definitions of the non-commutative (Cayley) determinant and permanent of an $n \times n$ matrix of variables $X = (X_{i,j})_{i,j \in [n]}$:

$$\det(X) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) X_{1,\sigma(1)} \cdot X_{2,\sigma(2)} \cdots X_{n,\sigma(n)} \qquad \operatorname{per}(X) = \sum_{\sigma \in S_n} X_{1,\sigma(1)} \cdot X_{2,\sigma(2)} \cdots X_{n,\sigma(n)}$$

That is, we just take the commutative determinant and permanent and make it non-commutative by ordering the variables in each monomial in increasing order of the rows in which they appear.

Lemma 23. *Let P_d be the $2d \times 2d$ matrix with x_0 on the diagonal, x_1 on the anti-diagonal, and 0 everywhere else. Let D_d be the $2d \times 2d$ matrix with x_0 on the diagonal, x_1 on the first d positions of the anti-diagonal and $-x_1$ on the last d positions of the anti-diagonal. Then $\mathsf{PAL}_d(x_0, x_1) = \operatorname{per} P_d = \det D_d$.*

Proof. The permanent of P_d can be obtained by choosing in each row of P_d a column index, while ensuring that each column index is taken only once; multiplying the values obtained; and then adding the results for all possible choices. Since there are only two non-zero values per row, for the row i (with $1 \leq i \leq d$), we can either choose the index i with value x_0 or the index $2d+1-i$ with value x_1. In the first case, the column of index i is now forbidden and therefore for the row $2d+1-i$ the only available non-zero value is x_0 with the column index $2d+1-i$. In the the second case, the column of index $2d+1-i$ is now forbidden and therefore for the row $2d+1-i$ the only available non-zero value is x_1 with column index i.

For the determinant, note that the above reasoning shows that a permutation yielding a non-zero value is a combination of fixed points (when choosing the value x_0 at row i in column i one must then choose value x_0 at row $2d+1-i$ in column $2d+1-i$) and transpositions (when choosing the value x_1 at row $2d+1-i$ in column i one must then choose value x_1 at row i in column $2d+1-i$). Therefore adding a minus sign to the last d values x_1 cancels out the sign of the permutation in the determinant. □

Corollary 24. *Let $k, d \in \mathbb{N}$ be any parameters such that $(8k+12)|d$ and $64|(d/k)$. Any circuit of non-skew depth k for the permanent or the determinant of an $d \times d$ matrix must have size $2^{\Omega(d/k)}$.*

Proof. Let us show the corollary for the permanent only, since the case for the determinant is similar. We will show that there exists a matrix P_k such that the permanent of P_k is f'_k, where f'_k is f_k but built with the 2-variable palindrome polynomial ($n = 2$). We will follow the construction of f_k from the proof of Lemma 18. Lemma 23 shows that there exists a matrix of order d_0 whose permanent is $\mathsf{PAL}_{d_0}(x_0, x_1)$. To get f'_0 from this polynomial, or to go from f'_{k-1} to f'_k we basically need two types of steps.

1. Computing the product of two previously obtained polynomials. If we have already built two matrices M and N whose permanents are f and g respectively, then clearly $f \cdot g$ is the permanent of the block diagonal matrix with M and N on the diagonal. The order of the block matrix is the sum of the orders of M and N.

2. Computing a j-product of a previously computed polynomial with a palindrome polynomial. If we have already built a matrix M whose permanent is the polynomial f, then we can build a matrix whose permanent is $f \times_{d_0} \mathsf{PAL}_{d_0}(x_0, x_1)$ by considering the block matrix $\left(\begin{smallmatrix} D & 0 & A \\ 0 & M & 0 \\ A & 0 & D \end{smallmatrix}\right)$, where D is the order-d_0 matrix with x_0 on the diagonal and A is the order-d_0 matrix with x_1 on the anti-diagonal (the reasoning is similar to the one in the proof of Lemma 23). The order of this matrix is the order of M plus $2d_0$.

Thus f'_0 is the permanent of a matrix of order $8d_0$ and going from f'_{k-1} to f'_k increases the size of the matrix by $12d_0$ (refer once again to the proof of Lemma 18). The order of the matrix P_k whose permanent is f'_k is thus $d := D_k = (8 + 12k)d_0$. By Theorem 22, any circuit of non-skew depth k for the permanent must have size $2^{\Omega(d_0)} = 2^{\Omega(d/k)}$. □

8 Full-rank with respect to all partitions

Our lower bound proofs have been based on showing that any arithmetic circuit of non-skew depth at most k cannot compute a polynomial that has large rank w.r.t. some fixed partition Π_k. We can ask if this strategy can yield lower bounds for general non-commutative arithmetic circuits (i.e., with no restrictions on non-skew depth) as well. Our aim in this section is to show that the answer to this question is possibly no: we show that over any large enough field $\mathbb{F}$ and any set of n variables X, there is a polynomial $p \in \mathbb{F}\langle X\rangle$ that has non-commutative arithmetic circuits of polynomial size, but which furthermore satisfies the property that for *all* partitions $\Pi = (Y, Z)$ with $|Y| \leq |Z|$, rel-rank$(p, \Pi) = 1$. This shows that we cannot even hope to prove that for any polynomial p computed by a polynomial-sized non-commutative circuit, there exists *some* partition with respect to which p has small rank.

The proof follows closely a very similar construction due to Raz and Yehudayoff from [18] in the context of commutative *multilinear circuits*.

Notation. We first introduce some notation. Given a finite set S of even cardinality, we define an *S-matching* to be an unordered partition of S into sets of size two: i.e., M is an S-matching if $M \subseteq \binom{S}{2}$ and the sets in M partition S.

Fix any degree parameter $d \in \mathbb{N}$ that is *even*. For any $i, j \in [d]$ with $i < j$ and $|[i, j]| = j - i + 1$ even, we define a set $\mathcal{M}_{i,j}$ of $[i, j]$-matchings as follows. The set $\mathcal{M}_{i,j}$ is defined by induction on $|[i, j]|$. The base case is when $j = i + 1$ and in this case, we set $\mathcal{M}_{i,j} = \{\{i, i + 1\}\}$. In the case that $j - i + 1 = 2\ell$ for $\ell > 1$, we define the set $\mathcal{M}_{i,j}$ as follows:

$$\begin{aligned}\mathcal{M}_{i,j} =& \{M \cup M' \mid M \in \mathcal{M}_{i,j'}, M' \in \mathcal{M}_{j'+1,j} \text{ for some } j' \in \{i + 1, i + 3, \ldots, j - 2\}\} \\ &\cup \{M \cup \{\{i, j\}\} \mid M \in \mathcal{M}_{i+1,j-1}\}\end{aligned}$$

Now, fix any $\lambda_e \in \mathbb{F}$ for each $e \in \binom{[d]}{2}$. Given any set $M \subseteq \binom{[d]}{2}$, we denote by λ_M the product $\prod_{e \in M} \lambda_e$. Finally, we define the polynomial $p^{\overline{\lambda}}$ (where $\overline{\lambda}$ denotes the tuple $(\lambda_{1,2}, \ldots, \lambda_{d-1,d})$) to be

$$p^{\overline{\lambda}}(X) = \sum_{M \in \mathcal{M}_{1,d}} \lambda_M \cdot p_M(X) \tag{18}$$

where p_M is defined as follows.

$$p_M(X) = \sum_{w \in [n]^d : w_i = w_j \forall \{i,j\} \in M} \tilde{x}_w$$

(Above, $\tilde{x}_w = x_{w_1} \cdots x_{w_d}$ as defined in Section 4.1.) [7]

We will show that for any choice of λ_e ($e \in \binom{[d]}{2}$), the polynomial $p^{\overline{\lambda}}$ has a non-commutative circuit of size $\mathrm{poly}(n, d)$. On the other hand, if the field $\mathbb{F}$ is large enough, then *there exists* a choice of λ_e ($e \in \binom{[d]}{2}$) such that for any partition $\Pi = (Y, Z)$ with $|Y| \leq |Z|$, $\mathrm{rank}(M[p^{\overline{\lambda}}, \Pi]) = n^{|Y|}$ (i.e., $\text{rel-rank}(p^{\overline{\lambda}}, \Pi) = 1$.

The first lemma gives us the circuit upper bound.

Lemma 25. *Fix any field $\mathbb{F}$ and $d, n \in \mathbb{N}$ such that d is even. For any choice of field elements $\lambda_e \in \mathbb{F}$ ($e \in \binom{[d]}{2}$), the polynomial $p^{\overline{\lambda}}$ has a non-commutative arithmetic circuit of size* $\mathrm{poly}(n, d)$.

Proof. We first define several intermediate polynomials that are computed in the course of computing the polynomial $p^{\overline{\lambda}}$. For any $i, j \in [d]$ such that $i < j$ and $\ell := j - i + 1$ is even, define the polynomial $p^{\overline{\lambda}}_{i,j}$ to be

$$p^{\overline{\lambda}}_{i,j}(X) = \sum_{M \in \mathcal{M}_{i,j}} \lambda_M \cdot p_M(X)$$

where p_M, for $M \in \mathcal{M}_{i,j}$ is defined as

$$p_M(X) = \sum_{w \in [n]^\ell : w_{s-(i-1)} = w_{t-(i-1)} \forall \{s,t\} \in M} \tilde{x}_w.$$

Note that $p^{\overline{\lambda}}$ is the same as $p^{\overline{\lambda}}_{1,d}$. Our circuit for $p^{\overline{\lambda}}$ computes $p^{\overline{\lambda}}_{i,j}$ for each $i, j \in [d]$. The construction is increasing order of the parameter ℓ.

When $\ell = 2$ (the smallest value possible), the polynomial is simply $p^{\overline{\lambda}}_{i,i+1} = \lambda_{\{i,i+1\}} \sum_{x \in X} xx$, which can be computed by a circuit of size $O(n)$.

Now say we have a circuit C of size S that computes $p^{\overline{\lambda}}_{s,t}$ when $t - s + 1 < \ell$. To compute $p^{\overline{\lambda}}_{i,j}$ where $j - i + 1 = \ell$, we use the following simple identity, which follows from the definition of $\mathcal{M}_{i,j}$

$$p^{\overline{\lambda}}_{i,j} = \left(\sum_{j' \in \{i+1, i+3, \ldots, j-2\}} p^{\overline{\lambda}}_{i,j'} \cdot p^{\overline{\lambda}}_{j'+1,j} \right) + \lambda_{i,j} \sum_{x \in X} x \cdot p^{\overline{\lambda}}_{i+1,j-2} \cdot x$$

[7] The reader may find it instructive to note that each polynomial for which we have proved a lower bound so far has been of the form p_M for some $[d]$-matching M.

Since each of the polynomials $p^{\overline{\lambda}}_{i,j'}$, $p^{\overline{\lambda}}_{j'+1,j}$, and $p^{\overline{\lambda}}_{i+1,j-2}$ have already been computed by the circuit C, the additional size required to compute $p^{\overline{\lambda}}_{i,j}$ is $O(d+n)$. We continue this way until we have computed all the $p^{\overline{\lambda}}_{i,j}$.

The total number of pairs i,j is $O(d^2)$ and hence the size of the circuit thus constructed is $O(d^2(d+n)) = \text{poly}(n,d)$. □

The second lemma tells us that it suffices to consider only *balanced* partitions (Y,Z): i.e., partitions such that $|Y| = |Z| = d/2$.

Lemma 26. *Let $d \in \mathbb{N}$ be even. Let $f \in \mathbb{F}\langle X\rangle$ be any homogeneous polynomial of degree d. If there is a partition $\Pi = (Y,Z)$ with $|Y| \leq |Z|$ such that* rel-rank$(f,\Pi) < 1$*, then for any balanced partition $\Pi' = (Y',Z')$ such that $Y' \supseteq Y$, we have* rel-rank$(f,\Pi') < 1$.

Proof. Consider the matrix $M[f,\Pi']$. Each row is labelled by a monomial m of degree $|Y'|$, which can be identified with a pair (m',m'') where m' is the natural restriction of m to the locations in Y and m'' is the restriction to the locations in $Y' \setminus Y$.

Fix any m'' and consider all the monomials m that give rise to this particular m''. The resulting matrix has exactly $n^{|Y|}$ rows and $n^{|Z'|}$ columns. Each column is labelled by a monomial m''' of degree $|Z'|$ and each row by a monomial m' of degree $|Y|$. The (m',m''')th entry of the the matrix is the coefficient — in the polynomial f — of the monomial m which equals m' when restricted to Y, equals m'' when restricted to $Y' \setminus Y$, and equals m''' when restricted to Z'. It is not hard to check that this matrix is a submatrix of the matrix $M[f,\Pi]$ (obtained by removing some columns). Since rel-rank$(f,\Pi) < 1$, we have rank$(M[f,\Pi]) < n^{|Y|}$.

Thus, for any fixed m'', the rank of the submatrix obtained as above has rank $< n^{|Y|}$. Since there are $n^{|Y'|-|Y|}$ such matrices, the rank of $M[f,\Pi']$ is strictly less than $n^{|Y'|-|Y|} \cdot n^{|Y|} = n^{|Y'|}$. Hence, we have rel-rank$(f,\Pi') < 1$. □

Lemma 27. *Let $d \in \mathbb{N}$ be even and $\mathbb{F}$ be any field such that $\mathbb{F}$ is either infinite or $|\mathbb{F}| > d2^{2d}$. Then, there is a choice of field elements $\lambda_e \in \mathbb{F}$ ($e \in \binom{[d]}{2}$) such that for any balanced partition Π, we have* rel-rank$(p^{\overline{\lambda}},\Pi) = 1$.

Proof. We fix any finite subset $F \subseteq \mathbb{F}$ of size at least $d2^{2d}+1$ and choose each λ_e ($e \in \binom{[d]}{2}$) independently and uniformly at random from F. We will show that $p^{\overline{\lambda}}(X)$ has the required property with non-zero probability over the choice of the λ_e.

Fix any balanced partition $\Pi = (Y,Z)$. We say that a $[d]$-matching M is *good* for Π if, for each $i \in Y$, there is a $j \in Z$ such that $\{i,j\} \in M$.

We use the following simple fact about the set of matchings $\mathcal{M}_{1,d}$.

Fact 28. *For any balanced partition $\Pi = (Y,Z)$, there is a matching $M \in \mathcal{M}_{1,d}$ that is good for Π.*

By Fact 28, there is a matching $M_0 \in \mathcal{M}_{1,d}$ such that M_0 is good for Π. It follows then from the definition of p_{M_0} above that the matrix $M[p_{M_0},\Pi]$ is a permutation matrix and hence rank$(M[p_{M_0},\Pi]) = n^{d/2}$. We argue that, with high probability over the choice of $\overline{\lambda}$, this is true of the polynomial $p^{\overline{\lambda}}$ as well.

In order to do this, we consider $\det(M[p^{\overline{\lambda}}, \Pi])$. By the definition of $p^{\overline{\lambda}}$, we have

$$M[p^{\overline{\lambda}}, \Pi] = \sum_{N \in \mathcal{M}_{1,d}} \lambda_N M[p_N, \Pi] = \lambda_{M_0} M[p_{M_0}, \Pi] + \sum_{N \in \mathcal{M}_{1,d} \setminus \{M_0\}} \lambda_N M[p_N, \Pi]$$

Since $M[p_N, \Pi]$ is a 0-1 matrix for each N, we see that $\det(M[p^{\overline{\lambda}}, \Pi])$ is a polynomial in λ_e ($e \in \binom{[d]}{2}$) of degree at most $d2^d$. We claim that this polynomial is in fact non-zero: to see this, note that if we substitute $\lambda_e = 1$ for $e \in M_0$ and 0 for $e \notin M_0$ in the above expression for $M[p^{\overline{\lambda}}, \Pi]$, we obtain the matrix $M[p_{M_0}, \Pi]$; hence, under this substitution, the polynomial $\det(M[p^{\overline{\lambda}}, \Pi])$ takes the value $\det(M[p_{M_0}, \Pi])$ which is non-zero since M_0 is a permutation matrix. We have thus shown that $\det(M[p^{\overline{\lambda}}, \Pi])$ is a non-zero polynomial in λ_e ($e \in \binom{[d]}{2}$). Since the degree of this polynomial is at most $d2^d$, for λ_e uniformly randomly chosen from F, we have by the Schwartz-Zippel lemma [20]

$$\Pr_{\overline{\lambda}}[\det(M[p^{\overline{\lambda}}, \Pi]) = 0] \leq \frac{d2^d}{|F|} < \frac{1}{2^d}$$

since $|F| > d2^{2d}$. Union bounding over the $\binom{d}{d/2} \leq 2^d$ choices for Π, we see that with probability greater than 0 over the choice of $\overline{\lambda}$, we have $\det(M[p^{\overline{\lambda}}, \Pi]) \neq 0$ for each balanced partition Π and hence, $\text{rel-rank}(p^{\overline{\lambda}}, \Pi) = 1$ for every balanced partition Π. □

Theorem 29. *Let $d \in \mathbb{N}$ be even and $\mathbb{F}$ be any field such that $\mathbb{F}$ is either infinite or $|\mathbb{F}| > d2^{2d}$. Let X be any set of n variables. Then, there is a homogeneous polynomial $p \in \mathbb{F}\langle X \rangle$ of degree d such that p has a circuit of size* $\text{poly}(n, d)$ *but given any partition $\Pi = (Y, Z)$ such that $|Y| \leq |Z|$, we have* $\text{rel-rank}(p, \Pi) = 1$.

Proof. Follows directly from Lemmas 25, 26, and 27. □

References

[1] E. Allender, J. Jiao, M. Mahajan, and V. Vinay. Non-commutative arithmetic circuits: Depth reduction and size lower bounds. *Theor. Comput. Sci.*, 209(1-2):47–86, 1998.

[2] A. Barvinok. New permanent estimators via non-commutative determinants, 2000.

[3] R. Beigel. When do extra majority gates help? polylog(n) majority gates are equivalent to one. *Computational Complexity*, 4:314–324, 1994.

[4] P. Bürgisser, J. M. Landsberg, L. Manivel, and J. Weyman. An overview of mathematical issues arising in the geometric complexity theory approach to **VP** $\neq$ **VNP**. *SIAM J. Comput.*, 40(4):1179–1209, Aug. 2011.

[5] A. Chattopadhyay and K. A. Hansen. Lower bounds for circuits with few modular and symmetric gates. In L. Caires, G. F. Italiano, L. Monteiro, C. Palamidessi, and M. Yung, editors, *ICALP*, volume 3580 of *Lecture Notes in Computer Science*, pages 994–1005. Springer, 2005.

[6] X. Chen, N. Kayal, and A. Wigderson. Partial derivatives in arithmetic complexity and beyond. *Foundations and Trends in Theoretical Computer Science*, 6(1-2):1–138, 2011.

[7] S. Chien, L. E. Rasmussen, and A. Sinclair. Clifford algebras and approximating the permanent. *J. Comput. Syst. Sci.*, 67(2):263–290, 2003.

[8] Z. Dvir, G. Malod, S. Perifel, and A. Yehudayoff. Separating multilinear branching programs and formulas. In H. J. Karloff and T. Pitassi, editors, *STOC*, pages 615–624. ACM, 2012.

[9] P. Hrubes, A. Wigderson, and A. Yehudayoff. Non-commutative circuits and the sum-of-squares problem. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:21, 2010.

[10] L. Hyafil. The power of commutativity. In *FOCS*, pages 171–174. IEEE Computer Society, 1977.

[11] S. Lovett and S. Srinivasan. Correlation bounds for poly-size ac circuits with $n^{1-o(1)}$ symmetric gates. In L. A. Goldberg, K. Jansen, R. Ravi, and J. D. P. Rolim, editors, *APPROX-RANDOM*, volume 6845 of *Lecture Notes in Computer Science*, pages 640–651. Springer, 2011.

[12] M. Mahajan and B. V. R. Rao. Small space analogues of valiant's classes and the limitations of skew formulas. *Computational Complexity*, 22(1):1–38, 2013.

[13] G. Malod and N. Portier. Characterizing valiant's algebraic complexity classes. *J. Complex.*, 24(1):16–38, 2008.

[14] N. Nisan. Lower bounds for non-commutative computation (extended abstract). In C. Koutsougeras and J. S. Vitter, editors, *STOC*, pages 410–418. ACM, 1991.

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

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

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

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

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

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

[21] 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.

[22] S. Toda. Classes of arithmetic circuits capturing the complexity of computing the determinant. *TIEICE: IEICE Transactions on Communications/Electronics/Information and Systems*, 1992.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de