Improved lower bound for multi-r-ic depth four circuits as a function of the number of input variables

SUMANT HEGDE[1] and CHANDAN SAHA[1] *

[1]*Department of Computer Science and Automation, Indian Institute of Science, Bangalore - 560012, India*

An arithmetic circuit (respectively, formula) is a rooted graph (respectively, tree) whose nodes are addition or multiplication gates and input variables/nodes. It computes a polynomial in a natural way. The formal degree of an addition (respectively, multiplication) gate with respect to a variable x is defined as the maximum (respectively, sum) of the formal degrees of its children, with respect to x. The formal degree of an input node with respect to x is 1 if the node is labelled with x, and 0 otherwise. In a *multi-r-ic* formula, the formal degree of every gate with respect to every variable is at most r. Multi-r-ic formulas make an intermediate model between multilinear formulas (the $r = 1$ case), for which lower bounds are relatively well-understood, and general formulas (the unbounded-r case), which are conjectured to have superpolynomial size lower bound.

On depth four multi-r-ic formulas/circuits computing $\mathrm{IMM}_{n,d}$ – the product of d symbolic matrices of size $n \times n$ each, Kayal, Saha and Tavenas (Kayal et al., 2016b) showed a lower bound of $(\frac{N}{dr^2})^{\Omega(\sqrt{d/r})}$ (where $N \approx n^2 d$, the total number of underlying variables). As a function of N and r, the lower bound is at most $2^{\Omega(\sqrt{N/r^3})}$ when $d = \Theta(N/r^2)$, and so for the bound to remain superpolynomial (as a function of N), r can be at most $N^{1/3}$. Our work proves a superpolynomial lower bound (as a function of N) on the same model (but computing a VNP-polynomial), for r as high as $(N \log N)^{0.9}$. It also yields a better lower bound than that of (Kayal et al., 2016b), when viewed as a function of N and r.

Theorem. Let N, d, r be positive integers such that $0.51N \leq d \leq 0.9N$ and $r \leq (N \log N)^{0.9}$. There is an explicit N-variate degree-d multilinear polynomial in VNP such that any multi-r-ic depth four circuit computing it has size $2^{\Omega\left(\sqrt{\frac{N\log N}{r}}\right)}$.

Keywords: **Arithmetic circuits, Multi-r-ic formulas, Lower bounds, Shifted Partial Derivatives, Nisan-Wigderson polynomial**

I. Introduction

In the recent years, algebraic computation has been attracting wide attention. Algebraic computation is a recurring feature in algorithms for problems such as matrix multiplication, determinant computation, fast Fourier transform, factoring polynomials (and integers), computing gcd etc., which have practical applications in various technological and scientific fields. Unsurprisingly, theoretical computer scientists have closely investigated both the algorithmic and the complexity theoretic aspects of algebraic operations. The latter has resulted in the emergence of algebraic complexity theory – a branch of computational complexity theory.

Arithmetic circuits and formulas

In algebraic complexity theory, many interesting questions have connection with the efficiency of computation of polynomials. Of the many models that are defined to capture the computation of

[*] Author for correspondence: chandan@iisc.ac.in

polynomials in a step-by-step and succinct fashion, *arithmetic circuits* seem to be natural and appealing. Section III gives a formal description of arithmetic circuits (and formulas). An example of an arithmetic circuit (which is also a formula) computing the polynomial $x^2+xy+xz+yz$ is shown in Figure 1. Two parameters associated with an arithmetic circuit are *size* and *depth*: they are defined respectively as the number of edges and the length of the longest directed path in the circuit.

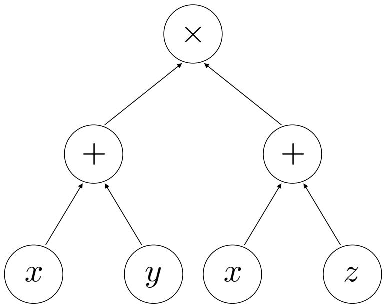

Figure 1: *An arithmetic circuit computing $x^2+xy+xz+yz$*

Under this model, broadly three kinds of problems are studied, namely lower bounds, polynomial identity testing (PIT), and circuit reconstruction. Roughly, a lower bound problem seeks to show that every arithmetic circuit computing an explicit polynomial f must be of at least certain size. In PIT, the problem is to efficiently check whether a given arithmetic circuit computes the identically zero polynomial. It is a highly important derandomization problem. The circuit reconstruction/learning problem is as follows: A polynomial f is given as a 'blackbox' which has the ability to take as input a field element t (a query) and output $f(t)$. The goal is to efficiently design a circuit computing f using few queries to the blackbox. These three problems have some fascinating connections among them.

Lower bounds

Lower bounds are more interesting when the polynomial f in question is a 'naturally occurring' one, such as the Det_n. Det_n is the determinant of an $n \times n$ matrix whose entries are distinct symbolic variables, making Det_n an n^2-variate degree-n polynomial. It is believed that every arithmetic formula computing Det_n requires a superpoly(n) size. In comparison, there is an efficient – i.e. poly(n)-sized – arithmetic circuit that computes Det_n. On the other hand, consider Perm_n, the *permanent* polynomial. (Perm_n is obtained by replacing every every -1 coefficient with $+1$ in the polynomial Det_n.) It is known due to (Bürgisser, 2000) that if the famous conjecture $\mathsf{P} \neq \mathsf{NP}$ is true (in the nonuniform setting) then every arithmetic circuit over $\mathbb{C}$ computing Perm_n must have superpoly(n) size, assuming the generalized Riemann Hypothesis. This is restated in terms of classes VP and VNP (Valiant, 1979) – the arithmetic analogues of (nonuniform) P and NP respectively: $\mathsf{P} \neq \mathsf{NP}$ (nonuniformly) $\implies \mathsf{VP} \neq \mathsf{VNP}$ over $\mathbb{C}$ (under the generalized Riemann Hypothesis). This connection suggests that working first towards proving $\mathsf{VP} \neq \mathsf{VNP}$ is plausible, and motivates the goal of proving superpolynomial lower bounds against VNP-polynomials (i.e. against polynomial families in VNP).

The lower bound problem has an interesting connection with derandomizing PIT. Kabanets and Impagliazzo (Kabanets and Impagliazzo, 2004) showed that a superpolynomial (similarly, exponential) lower bound for arithmetic circuits implies subexponential (similarly, quasipolynomial) time PIT. In the other direction, Agrawal (Agrawal, 2005) showed that a polynomial time blackbox PIT algorithm implies a superpolynomial lower bound for circuits computing an explicit (PSPACE-computable) polynomial.

Some known formula lower bounds. While the conjectured lower bound for formulas computing

Det_n is superpoly(n), the best known lower bound for the same is $\Omega(n^3)$ (Kalorkoti, 1985). (A slightly better $\Omega(N^2)$ bound is known for formulas computing an N-variate VNP-polynomial).[1] This long-standing wide gap has prompted the community to consider restricted variants of formulas and prove better lower bounds for them. *Multilinear* formulas are one such variant: in a multilinear formula, the formal degree of every gate with respect to every variable is at most 1.[2] In other words, the formula is syntactically forced to compute a multilinear polynomial. A polynomial is said to be multilinear if its degree with respect to every variable is at most 1. The choice of multilinearity constraint is justified from the fact that important polynomials such as Det_n, Perm_n, $\mathrm{IMM}_{n,d}$ (which is the (1,1)-th entry of the iterated product of d symbolic matrices of size $n \times n$ each) are all multilinear.

A lower bound of $n^{\Omega(\log n)}$ on multilinear formulas computing Det_n (and Perm_n) was shown by Raz(Raz, 2009). Subsequently (Raz and Yehudayoff, 2008, 2009) showed a superpolynomial lower bound on multilinear circuits of constant depth computing Det_n.

Formulas with high formal degree

Keeping in mind the open problem of superpolynomial lower bound on general formulas, particularly with a multilinear polynomial (like Det_n) as the target polynomial to be computed, it is natural at this point to wonder how general formulas compare with multilinear formulas. The total formal degree of a multilinear formula is bounded by the number of variables N, whereas that of a general formula is virtually unbounded (rather bounded by size of formula which can be much larger than N). This makes it difficult to adapt many of the prevalent proof techniques to general formulas, as they seem to only work when the total formal degree is low. General formulas, having essentially a free hand on the maximum formal degree, can employ 'clever' cancellations of high degree monomials at intermediate gates and use this possibility to efficiently compute some otherwise-hard multilinear polynomials. For example, the best known circuit of depth three[3] computing Det_n (which is of degree n) has formal degree $n^{\Omega(\sqrt{n})}$(Gupta et al., 2013). This prompted Kayal and Saha (Kayal and Saha, 2015) to turn the attention to high formal degree models and define multi-r-ic formulas.

Multi-r-ic formulas. In a multi-r-ic formula the formal degree of every gate with respect to every variable is at most r. Clearly, multilinear formulas are the $r = 1$ case of multi-r-ic formulas. The circuit shown in Figure 1 is a multi-2-ic formula (albeit computing a non-multilinear polynomial). Multi-r-ic formulas, allowing the total formal degree as high as r times the number of variables, form an intermediate model between multilinear and general formulas.

Homogeneous formulas. Another direction of attack could be to first reduce general formulas to *homogeneous* formulas and then prove a prove a lower bound on homogeneous formulas. A formula is homogeneous if every gate in it computes a homogeneous polynomial. (It follows that the total formal degree of a homogeneous formula is 'low', in fact exactly the degree of the polynomial computed.) However, we do not know of any efficient[4] formula homogenizing algorithm (although such an algorithm is known for circuits (Strassen, 1973)), unless the degree of the polynomial computed is as low as $O(\frac{\log N}{\log\log N})$ (Raz, 2013). Nevertheless, a homogenous formula is an interesting model in its own right and proving superpolynomial lower bounds for it would be a great progress.

[1] The best known lower bound for circuits is $\Omega(N \log N)$, against a certain N-variate VNP-polynomial.

[2] Similarly, multilinear circuits are defined.

[3] over fields of characteristic zero

[4] costing only a poly-size blowup

Depth reduction. Yet another possible route to proving superpolynomial formula (in fact, circuit) lower bound goes via depth reduction. A series of works (Valiant et al., 1983; Agrawal and Vinay, 2008; Koiran, 2012; Gupta et al., 2013; Tavenas, 2013) imply that any arithmetic circuit of size s computing an N-variate degree-d polynomial can be transformed into a depth three circuit of size $2^{O(\sqrt{d\log(ds)\log N})}$ (provided the underlying field is of characteristics zero). Hence if one shows a sufficiently high superpolynomial lower bound of $N^{\omega(\sqrt{d})}$ on depth three circuits computing a VNP-polynomial, then a superpolynomial lower bound on general circuits immediately follows, proving $\mathsf{VP} \neq \mathsf{VNP}$. An important point here, relevant to the preceding discussion, is that the depth three circuit resulting from the depth reduction potentially has as high a formal degree as $2^{\Omega(\sqrt{d\log(ds)\log N})}$. We note that a similar depth reduction result also holds for homogenous depth four circuit, but there the formal degree is not high. In essence, these depth reduction results show that low depth circuits, particularly depth three and depth four circuits, serve as an interesting testbed for proving lower bounds.

Previous works on multi-r-ic formulas

Kayal and Saha (Kayal and Saha, 2015) proved a $2^{\Omega(N/2^{25r})}$ lower bound on multi-r-ic formulas of depth three, computing a certain (non-multilinear) polynomial. The choice of depth three is natural: it is the smallest depth at which we do not know of a superpolynomial circuit/formula lower bound[5]. As mentioned before, another important motivation for depth three (and four) comes from the depth reduction results.

Kayal, Saha and Tavenas (Kayal et al., 2016b) improved the dependence on r and showed a lower bound of $\left(\frac{n}{r}\right)^{\Omega(d)}$ for depth three multi-r-ic formulas computing $\mathrm{IMM}_{n,d}$. Further, they showed a lower bound of $\left(n/r^{1.1}\right)^{\Omega(\sqrt{d/r})}$ for multi-r-ic depth four formulas computing the same polynomial. They proved an improved lower bound of $2^{\Omega(N)}$ on depth three multi-r-ic circuits (computing a multi-r-ic VNP-polynomial). (Kayal et al., 2016b) also showed that a certain polynomial computed by a small multi-r-ic formula of depth three is 'hard' for multi-r-ic homogeneous formulas of arbitrary depth. The underlying hope is, techniques used to prove depth three and depth four multi-r-ic formula lower bounds will shed some light on general multi-r-ic fomulas just like in the multilinear ($r = 1$) case – for instance, the proof of multilinear formula lower bound using log-product formula (Raz and Yehudayoff, 2009), which is a kind of multilinear depth four formula.

II. Our results

While (Kayal et al., 2016b) show a nontrivial lower bound on depth four multi-r-ic circuits for $r < N^{1/3}$, we give a lower bound on the same model that remains superpolynomial for a wider range of r (see discussion after the theorem).

Theorem 1. *Let N, d, r be positive integers such that $0.51N \leq d \leq 0.9N$ and $r \leq (N\log N)^{0.9}$. There is an explicit N-variate degree-d multilinear polynomial in* VNP *such that any multi-r-ic depth four circuit computing it has size $2^{\Omega\left(\sqrt{\frac{N\log N}{r}}\right)}$.*

Comparison with previous results

1. Better range on r. In (Kayal et al., 2016b), a lower bound of $\left(\frac{N}{dr^2}\right)^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$ was shown for multi-r-ic depth four circuits computing $\mathrm{IMM}_{n,d}$

[5] In the context of superpolynomial lower bound and a constant depth like three or four, we use terms circuits and formulas interchangeably. This is because when the depth is a constant, the circuit-to-formula conversion only costs poly-size blowup.

where $N \approx n^2 d$. For the bound to remain superpolynomial, r can be at the most $\min(\sqrt{\frac{N}{d}}, d)$. The expression $\min(\sqrt{\frac{N}{d}}, d)$ is maximized at $d = N^{1/3}$, and r has to be less than $N^{1/3}$. We show a lower bound of $2^{\Omega\left(\sqrt{\frac{N\log N}{r}}\right)}$ for $d \in [0.51N, 0.9N]$ and $r \leq (N\log N)^{0.9}$ which remains superpolynomial in this range for r. Observe that a higher range for r essentially means we prove lower bound for newer classes of depth four circuits.

2. Improved lower bound. For any fixed function $r = r(N)$, (Kayal et al., 2016b)'s lower bound of $(\frac{N}{dr^2})^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$ is maximized (as a function of N and r) to $2^{\Omega\left(\sqrt{\frac{N}{r^3}}\right)}$ at $d = \Theta\left(\frac{N}{r^2}\right)$. In comparison, Theorem 1 shows a bound of $2^{\Omega(\sqrt{\frac{N\log N}{r}})}$, which is an asymptotically better function of N and r.

3. Extending the result of Raz and Yehudayoff (Raz and Yehudayoff, 2009). The best known lower bound for multilinear ($r = 1$) depth four circuits is $2^{\Omega(\sqrt{N\log N})}$ (Raz and Yehudayoff, 2009). Our result can be seen as an extension of this lower bound to multi-r-ic depth four circuits, although the proof techniques in (Raz and Yehudayoff, 2009) and in here are quite different. In particular, (Raz and Yehudayoff, 2009) used rank of a partial derivatives matrix as the measure whereas we use the dimension of shifted partial derivatives, denoted as SP (see below for more details).

Proof outline and comparison with previous proof techniques

The proof of Theorem 1 follows a template for depth four circuit lower bound that is already existing in the literature, particularly in (Kayal et al., 2016b) and in related prior works. We briefly describe the proof outline before listing the differences with (Kayal et al., 2016b). The proof has the following structure:

1. *Reduction to low-bottom-support depth four circuits (step* 1*)*: Consider a depth four multi-r-ic circuit of 'small' size computing a 'hard' polynomial H. At first, we show that there exists a restriction of the circuit (i.e. setting of some variables to field constants in the circuit) which converts it into a more structured circuit called a *low-bottom-support* depth four circuit computing a restriction of H (say, F). Section III has the precise definition of low-bottom-support depth four circuits, and the reduction to this kind of circuits is formally stated in Lemma 3 (Section IV).

2. *Lower bound for low-bottom-support circuits (step* 2*)*: In this step, we show that any low-bottom-support depth four circuit must have high size (in particular, high top fanin) in order to compute F from step 1. Lemma 4 (which is stated in Section IV and proved in Section V) has the formal statement of this lower bound. The bound is achieved by proving (in Lemma 5) that circuits of this kind having low top fanin have a low *shifted partials measure* (defined in Section III), and subsequently proving in step 3 below that F has a high measure. Here, a measure is a function that maps polynomials to integers.

3. *Constructing the hard polynomial H (step* 3*)*: Finally, a VNP-polynomial H having high measure is constructed in this step. For this, we pick a variant of the *Nisan-Wigderson polynomial*, which was defined in (Kayal et al., 2014, 2016a). The construction is inspired by the well known Nisa-Wigderson design (Nisan and Wigderson, 1994) and Reed-Solomon codes (Reed and Solomon, 1960). Basically, H is defined in such a way that its restriction F is a multilinear polynomial whose monomials are sufficiently 'far' from each other. In this sense, the monomials correspond to codewords of a good code. The

precise construction of F is given in Section VI, and that of H (using the construction of F) is given in Section IV. Lemma 6 (which is stated is Section V and proved in Section VI) shows that F has a high measure.

The above three steps together imply a high lower bound on the size of any depth four multi-r-ic circuits computing H. As mentioned before, much of the proof machinery is borrowed from earlier works. However, we opt to present the proof in detail not only because of self containment but also because our parameter settings are often different from that in prior works.

The difference between (Kayal et al., 2016b) and our proof is in the exact choice of the measure and the hard polynomial:

1. *The choice of the measure*: (Kayal et al., 2016b) introduced a measure called *shifted skewed partials*, a variant of an already existing measure (defined in (Kayal, 2012)) called *shifted partials* (SP). For our proof, SP suffices. (Kayal et al., 2016b)'s focus was to get the lower bound as a function of both N and d i.e. the number of underlying variables and the degree of the polynomial computed respectively. For low degree (and $\mathrm{IMM}_{n,d}$ as the target polynomial), (Kayal et al., 2016b) found that a certain 'skew' between two sets of variables, with suitable parameters, was crucial in obtaining a better lower bound. However, for high degree, it seems that the skew does not offer an added advantage. Instead, we use SP itself as the measure and prove an improved bound for a high degree range. The improvement also stems from the different hard polynomial we choose.
2. *The choice of the hard polynomial*: (Kayal et al., 2016b) used $\mathrm{IMM}_{n,d}$, a VP-polynomial, whereas our proof works with a VNP-polynomial, ensuring that the latter has a sufficiently high SP measure.

III. Preliminaries

We use a bold letter, like $\mathbf{x},\mathbf{y}$ etc., to denote a set of variables. Elements of $\mathbf{x}$ are denoted by $x_1, x_2, \ldots$ etc. and are called $\mathbf{x}$-variables. We denote with $\mathbf{x}^{\leq \ell}$ the set of monomials in $\mathbf{x}$-variables of degree at most ℓ. Let f be a polynomial. Then $\deg_x f$ denotes the degree of f with respect to variable x, and $\deg f$ denotes the total degree of f. Also, for sets S and $\tilde{S}$ of polynomials, expressions $f \cdot S$, S/f, and $S \cdot \tilde{S}$ naturally denote the sets $\{fg : g \in S\}$, $\{g/f : g \in S\}$ and $\{g\tilde{g} : g \in S, \tilde{g} \in \tilde{S}\}$ respectively. $[n]$ denotes the set $\{1,2,\ldots,n\}$ and $\mathbb{N}$ the set of natural numbers. For a set $\mathbf{x}$ and integers $a \leq b$, the set of all subsets of $\mathbf{x}$ of size between a and b (inclusive) is denoted by $\binom{\mathbf{x}}{[a,b]}$, and simply by $\binom{\mathbf{x}}{a}$ when $a = b$. 'log' and 'ln' denote logarithms to base 2 and base e respectively. Sometimes we use the term $\mathsf{poly}(n)$ to mean $n^{O(1)}$. We assume N, the number of variables, to be sufficiently large (so as to legitimize inequalities that hold asymptotically). Also, sometimes we omit floor ($\lfloor\ \rfloor$) and ceil ($\lceil\ \rceil$) notations for real-valued functions of N,d etc. for simplicity of presentation, without affecting any of the implications.

i. Some well-known bounds

For a real number x,

$$1 + x \leq e^x. \tag{1}$$

For integers $1 \leq k \leq n$,

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k. \tag{2}$$

Chernoff bound. Let X be the sum of several independent 0-1 random variables. Then for any constant $\varepsilon > 0$,

$$\Pr[X \geq (1+\varepsilon)E[X]] \leq e^{-\varepsilon^2 E[X]/3},$$
$$\Pr[X \leq (1-\varepsilon)E[X]] \leq e^{-\varepsilon^2 E[X]/3}.$$

ii. Arithmetic circuits

We specify some of the concepts, stated in Section I, in a bit more details. The reader familiar with these may skip this part. An *arithmetic circuit* is a directed acyclic graph in which every node with in-degree 0 (called *input node*) is labelled with a variable or a field element, and every node with positive in-degree is labelled with either '+' (in which case the node is a *addition gate*) or '×' (in which case the node is a *multiplication gate*). If there is an edge from a node u to a node v then u is called a *child* of v. With every node we associate a polynomial and say that the node *computes* the polynomial, as follows: An input node is said to compute what it is labelled with. A sum (respectively product) gate is said to compute the sum (respectively product) of the polynomials associated with its children. We consider circuits which have exactly one *root*, i.e. the node with out-degree 0, and a circuit is said to compute the polynomial its root computes. Also, we allow edges to be labelled with field constants. If an edge from node u to node v is labelled with a constant α and u is computing a polynomial f then v considers αf, rather than mere f, as the input coming from u.

The *size* of a circuit is the number of edges in it. The *depth* of a circuit is the length of the longest path from an input node to the root. An arithmetic circuit in which all nodes have out-degree at most one is called a *formula*.

Depth three and depth four circuits. By a depth three circuit (also called a $\Sigma\Pi\Sigma$ circuit) we mean a circuit that has a top addition gate followed by a layer of multiplication gates and finally a bottom layer of addition gates. Similarly a circuit with a addition gate on top, followed by a layer of multiplication gates, then a layer of addition gates again, and finally a bottom layer of multiplication gates corresponds to a depth four circuit (also called a $\Sigma\Pi\Sigma\Pi$ circuit). Further if the monomials computed at the bottom layer of multiplication gates of a depth four circuit are such that each of them has at most τ variables appearing in it, then we say that the depth four circuit has *τ-bottom-support*.

Formal degree. The *formal degree* of an input gate g with respect to a variable x is defined to be 1 if g is labelled with x, and 0 if g is labelled with a different variable or a field element. The formal degree of a sum (respectively product) gate g with respect to a variable x is defined to be the maximum (respectively sum) of the formal degrees of its children with respect to x.

Multi-r-ic formulas. Let r be a positive integer. A *multi-r-ic* formula is an arithmetic formula such that every gate in it has formal degree at most r with respect to every variable. If $r = 1$, a multi-r-ic formula is called a *multilinear* formula. A polynomial is said to be multilinear if the degree of every variable is at most one in every monomial of the polynomial. Clearly, multilinear formulas compute multilinear polynomials.

Arithmetic complexity classes. A family of polynomials $\{f_n\}$ over a field $\mathbb{F}$, indexed by $n \geq 1$, is in the class VP if there is a polynomial $p : \mathbb{N} \to \mathbb{N}$ such that for every n, f_n has at most $p(n)$ variables, has degree at most $p(n)$ and can be computed by a circuit of size at most $p(n)$. A family of polynomials $\{f_n\}$ over $\mathbb{F}$ is in VNP if there is a polynomial family $\{g_n\}$ in VP and polynomials $p, \tilde{p} : \mathbb{N} \to \mathbb{N}$ such that

$$f(x_1, \ldots, x_{p(n)}) = \sum_{\substack{(w_1,\ldots,w_{\tilde{p}(n)}) \\ \in\{0,1\}^{\tilde{p}(n)}}} g(x_1, \ldots, x_{p(n)}, w_1, \ldots, w_{\tilde{p}(n)}).$$

It is clear that $\mathsf{VP} \subseteq \mathsf{VNP}$. In a later section, to check whether a polynomial f_n is in VNP we use *Valiant's criterion*: If there is a poly(n)-time algorithm to output the coefficient of a given monomial in f_n then $f_n \in \mathsf{VNP}$ (Valiant, 1979).

iii. The shifted partials measure

Let $\mathbb{F}$ be a field. For integer parameters $k, \ell \geq 0$, the shifted partials dimension is a function $\mathsf{SP}_{k,\ell} : \mathbb{F}[\mathbf{x}] \to \mathbb{N}$ defined as follows. Let $f \in \mathbb{F}[\mathbf{x}]$. For any multilinear degree-k monomial $\mu = x_{i_1} x_{i_2} \ldots x_{i_k}$, we write $\partial_\mu f$ to denote $\frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_k}}$. Also, for a set S of polynomials we write $\partial_\mu S$ to denote the set $\{\partial_\mu f : f \in S\}$, which will be convenient in Section V. Let $\partial^{=k} f$ denote the set $\{\partial_\mu f : \mu$ is a multilinear monomial of degree $k\}$. We define

$$\mathsf{SP}_{k,\ell}(f) \stackrel{\text{def}}{=} \dim\left(\operatorname{span}_{\mathbb{F}}\left(\mathbf{x}^{\leq \ell} \cdot \partial^{=k} f\right)\right). \qquad (3)$$

The following property is easy to establish.

Proposition 2 (Subadditivity)**.** *Let $f, g \in \mathbb{F}[\mathbf{x}]$. Then $\mathsf{SP}_{k,\ell}(f+g) \leq \mathsf{SP}_{k,\ell}(f) + \mathsf{SP}_{k,\ell}(g)$.*

IV. Proving Theorem 1

In Section VI we describe a multilinear polynomial $F_{\tilde{d}}(\mathbf{y})$ where $\mathbf{y}$ is the set of underlying variables and $\tilde{d}$ is the degree. Polynomial $F_{\tilde{d}}$ has mainly two properties:

1. It has $\binom{|\mathbf{y}|/4001}{k}$ monomials where $k = \frac{11}{840000}\sqrt{\frac{\tilde{d}}{r \log \tilde{d}}}$, and all of them are of degree $\tilde{d}$, i.e., $F_{\tilde{d}}$ is homogeneous.
2. For any two multilinear monomials μ_1 and μ_2, $|\mu_1 \setminus \mu_2|$ is at least $0.006\tilde{d}$. Here $\mu_1 \setminus \mu_2$ refers to the set of variables appearing in μ_1 but not in μ_2. Note that $|\mu_1 \setminus \mu_2| = |\mu_2 \setminus \mu_1|$. We call it the *distance* between μ_1 and μ_2.

We use $F_{\tilde{d}}$ to define the polynomial H (mentioned in step 1 of the proof outline in Section II).

Polynomial H. Let $\mathbf{x}, \mathbf{u}, \mathbf{v}$ be sets of variables of size N_0, N_0, and $0.02N_0$ respectively, making a total of $2.02N_0 = N$ (say) variables. Also let ρ denote the range $[0.95N_0, 0.97N_0]$. Let d be any integer in $[0.51N, 0.9N]$. Set $\tilde{d} = d - 0.97N_0$, and so $\tilde{d} \in [0.06N_0, 0.85N_0]$. Polynomial H, which is N-variate and of degree $\tilde{d}$, is defined as below:

$$H(\mathbf{x}, \mathbf{u}, \mathbf{v}) \stackrel{\text{def}}{=}$$

$$\sum_{\mathbf{y} \in \binom{\mathbf{x}}{\rho}} F_{\tilde{d}}(\mathbf{y}) \cdot \prod_{i:\, x_i \in \mathbf{y}} u_i \cdot \prod_{j=1}^{0.97N_0 - |\mathbf{y}|} v_j. \qquad (4)$$

Polynomial H is homogeneous and multilinear.

Proof of Theorem 1. Let C be a multi-r-ic depth four circuit computing H. H defines a polynomial in VNP, as we will show at the end of section VI (after fully describing $F_{\tilde{d}}(\mathbf{y})$).

The *sparsity* of a depth four circuit is defined as the sum of the fanin of nodes at the bottom summation layer. If the sparsity of C is greater than $2^{\sqrt{\frac{N \log N}{100r}}}$ then so is the size of C and there is nothing to prove. Hence we assume from now on that C has sparsity at most $2^{\sqrt{\frac{N \log N}{100r}}}$.

A *restriction* of a circuit means a substitution of field constants to some variables in the circuit. We are now ready to precisely state the reduction in step 1 of the proof outline in Section II.

Lemma 3 (Reduction to low-bottom-support depth four circuits)**.** *There exists a restriction σ of circuit C that converts it into a depth four multi-r-ic circuit of τ-bottom-support computing $F_{\tilde{d}}(\mathbf{y})$, where $\tau = 20 \cdot \sqrt{\frac{\tilde{d} \log \tilde{d}}{r}}$, and $\mathbf{y}$ is an element of $\binom{\mathbf{x}}{\rho}$.*

Proof of the above lemma is given at the end of the section. Let $\sigma(C)$ denote the circuit resulting from applying σ on C. $\sigma(C)$ has τ-bottom-support (due to the lemma above). Also, $r \leq (N \log N)^{0.9} = o(\frac{\tilde{d}}{\log \tilde{d}})$ (as $\tilde{d} = \Theta(N)$), and $\tilde{d} = d - 0.97N_0 \leq 0.9 \cdot |\mathbf{y}|$. Hence the lemma below, which formalizes step 2 of the proof outline in Section II, is applicable on $\sigma(C)$.

Lemma 4 (Lower bound for low-bottom-support depth four circuits)**.** *Let* $\mathbf{y}$ *be a set of variables and let* $\tilde{d} \le 0.9 \cdot |\mathbf{y}|$ *and* $r \le \frac{\tilde{d}}{10^{10}\log\tilde{d}}$ *be positive integers. Then every depth four multi-r-ic circuit having* τ*-bottom-support and computing* $F_{\tilde{d}}(\mathbf{y})$*, where* $\tau = 20 \cdot \sqrt{\frac{\tilde{d}\log\tilde{d}}{r}}$*, must have top fanin at least* $\left(\frac{\tau^{20}\tilde{d}}{|\mathbf{y}|\cdot r}\right)^{\frac{\tilde{d}}{10^5\tau r}}$. [6]

Proof of the lemma is given in the next section. Lemma 4 implies that $\sigma(C)$ has top fanin at least

$$\left(\frac{\tau^{20}\tilde{d}}{|\mathbf{y}|\cdot r}\right)^{(\frac{\tilde{d}}{10^5\tau r})}$$

$$= \left(20^{20} \cdot \left(\frac{\tilde{d}\log\tilde{d}}{r}\right)^{10} \cdot \frac{\tilde{d}}{|\mathbf{y}|\cdot r}\right)^{(\frac{1}{20\cdot 10^5}\cdot\sqrt{\frac{\tilde{d}}{r\log\tilde{d}}})}$$

$$\ge \left(0.02^{10} \cdot (N\log N)^{0.1}\right)^{(\frac{1}{20\cdot 10^5}\cdot\sqrt{\frac{\tilde{d}}{r\log\tilde{d}}})} \qquad (5)$$

$$= 2^{\Omega(\sqrt{\frac{N\log N}{r}})},$$

where Equation (5) follows from the fact that $\tilde{d} \ge 0.02N$ (since $d \ge 0.51N$), $r \le (N\log N)^{0.9}$, and $|\mathbf{y}| \le \frac{0.97N}{2.02}$. Thus, C too must have top fanin (and hence size) $2^{\Omega(\sqrt{\frac{N\log N}{r}})}$. □

Proof of Lemma 3. The proof uses the probabilistic method. We begin by describing the sample space of restrictions.

Restriction σ_R. Given a subset $R \subseteq \mathbf{x}$, let σ_R denote the following restriction (substitution) on some variables in $\mathbf{x} \uplus \mathbf{u} \uplus \mathbf{v}$. If $|R| \in \rho$ then

1. assign 0 to variables in $\mathbf{x} \setminus R$,
2. assign 0 to u_i's where $x_i \notin R$ and assign 1 to the other u_i's, and
3. assign 0 to v_j's where $j > 0.97N_0 - |R|$ and 1 to the other v_j's.

Otherwise, assign all variables 0. We note that if (and only if) $|R| \in \rho$ then $\sigma_R(H) = F_{\tilde{d}}(R)$. To elaborate, after Step 2 above, terms in H corresponding to $F_{\tilde{d}}(\mathbf{y})$ vanish for every proper superset $\mathbf{y} \supsetneq R$. Similarly, after Step 3, terms in H corresponding to $F_{\tilde{d}}(\mathbf{y})$ vanish for every proper subset $\mathbf{y} \subsetneq R$.

Random restriction of C. Recall that C computes $H(\mathbf{x},\mathbf{u},\mathbf{v})$. Consider forming the set $R \subseteq \mathbf{x}$ randomly as follows: Independently, with probability 0.96 pick every $\mathbf{x}$-variable and include it in R. Now, to prove Lemma 3 it suffices to show that

$$\Pr_R\Big[\sigma_R(C) \text{ has } \tau\text{-bottom-support and}$$
$$\text{computes } F_{\tilde{d}}(\mathbf{y}) \text{ for some } \mathbf{y} \in \binom{\mathbf{x}}{\rho}\ \Big] > 0.$$

Equivalently, by union bound, it suffices to show that $\Pr_R[E_1] + \Pr_R[E_2] < 1$, where E_1 is the event that $\sigma_R(C)$ has bottom support greater than τ and E_2 is the event that for every $\mathbf{y} \in \binom{\mathbf{x}}{p}$, $\sigma_R(C)$ does *not* compute $F_{\tilde{d}}(\mathbf{y})$.

Let $\langle C\rangle$ denote the set of monomials computed at the bottom multiplication gates of C. (Thus $|\langle C\rangle|$ is at most the sparsity of C.) For a monomial μ, let $\boldsymbol{\mu}$ denote the set of variables appearing in μ. Then

$$\begin{aligned}
&\Pr_R[E_1] \\
&\le \Pr_R[\exists\sigma_R(\mu) \in \langle\sigma_R(C)\rangle \text{ s.t. } |\boldsymbol{\mu}| > \tau] \\
&\le \Pr_R[\exists\mu \in \langle C\rangle \text{ s.t. } |\boldsymbol{\mu} \cap \mathbf{x}| > \tau \text{ and} \\
&\qquad \sigma_R(\mu) \ne 0] \\
&\le |\langle C\rangle| \cdot 0.96^{\tau} \qquad \text{(from union bound)} \\
&\le 2^{\sqrt{\frac{N\log N}{100r}}} \cdot 0.96^{20\cdot\sqrt{\frac{\tilde{d}\log\tilde{d}}{r}}} \\
&\le 2^{-0.01\sqrt{\frac{N\log N}{r}}} \qquad (\text{as } \tilde{d} \ge 0.029N).
\end{aligned}$$

To upper bound $\Pr_R[E_2]$, we note that E_2 is equivalent to the event $|R| \notin \rho$. Hence

$$\Pr_R[E_2] = \Pr_R[|R| \notin \rho]$$

[6] The same lower bound holds for a range of τ and r satisfying $1000\log|\mathbf{y}| \le \tau r \le \tilde{d}/5000$, provided the parameter k used in the construction of $F_{\tilde{d}}(\mathbf{y})$ is adjusted suitably.

$$\leq 2e^{-\frac{1}{3}\cdot\left(\frac{0.01}{0.96}\right)^2\cdot\frac{0.96N}{2.02}},$$

by noting that $\mathbf{E}[|R|] = 0.96N_0 = \frac{0.96N}{2.02}$ and applying Chernoff bound. Clearly, $\Pr_R[E_1] + \Pr_R[E_2] < 1$, as required. □

V. Proving Lemma 4

A depth four multi-r-ic circuit Γ with τ-bottom-support is of the following form:

$$\begin{aligned} \Gamma &= T_1 + T_2 + \cdots + T_s, \\ T_i &= Q_{i1} \cdot Q_{i2} \ldots Q_{im_i} \quad \forall i \in [s], \end{aligned} \tag{6}$$

where, for every $i \in [s]$ and every $j \in [m_i]$, $Q_{ij} \in \mathbb{F}[\mathbf{y}]$ is a polynomial such that

1. every monomial in it contains at most τ variables (due to τ-bottom-support), and
2. for every $x \in \mathbf{y}$, $\sum_{j=1}^{m_i} \deg_x Q_{ij} \leq r$ (due to multi-r-icity).

Proof of Lemma 4. Suppose that Γ computes $F_{\tilde{d}}(\mathbf{y})$. Then our task is to show that the top fanin s is high.

Suppose that we estimate an upper bound $U = U(k,\ell)$ on $\mathsf{SP}_{k,\ell}(T_i)$, for every $i \in [s]$. Then Proposition 2 implies that

$$\mathsf{SP}_{k,\ell}(\Gamma) \leq sU.$$

Suppose also that we find a lower bound $L = L(k,\ell)$ on $\mathsf{SP}_{k,\ell}(F_{\tilde{d}})$, perhaps by fixing parameters k,ℓ. Then, since Γ computes $F_{\tilde{d}}$, it follows that

$$\begin{aligned} L \leq \mathsf{SP}_{k,\ell}(F_{\tilde{d}}(\mathbf{y})) = \mathsf{SP}_{k,\ell}(\Gamma) \leq sU \\ \Rightarrow s \geq L/U. \end{aligned}$$

To estimate U, we make use of the lemma below.

Lemma 5 ('Low' SP measure for circuits). *For any $i \in [s]$ and positive integers k,ℓ where $k \leq 2|\mathbf{y}|/\tau + 1$,*

$$\mathsf{SP}_{k,\ell}(T_i) \leq \binom{3|\mathbf{y}|/\tau}{k} \cdot \binom{|\mathbf{y}| + k\tau r + \ell}{|\mathbf{y}|}.$$

Proof of the lemma is at the end of this section. Let $\varepsilon = 0.0055$. We fix

$$k = \frac{\varepsilon\tilde{d}}{21\tau r} = \frac{11}{840000}\sqrt{\frac{\tilde{d}}{r\log\tilde{d}}} \geq 1 \quad \text{and} \tag{7}$$

$$\ell = \frac{0.006\tilde{d}\cdot|\mathbf{y}|}{\ln\binom{|\mathbf{y}|/4001}{k}} - |\mathbf{y}|. \tag{8}$$

For such ℓ, it can be shown that $\ell > 400 \cdot |\mathbf{y}|$, from which follows an inequality we require shortly:

$$\frac{|\mathbf{y}| + \ell}{1+\ell} \leq \frac{12}{11}. \tag{9}$$

To estimate L we use the following lemma:

Lemma 6 ('High' SP measure for $F_{\tilde{d}}$). *For integers k,ℓ fixed as above,*

$$\mathsf{SP}_{k,\ell}(F_{\tilde{d}}(\mathbf{y})) \geq \frac{1}{2} \cdot \binom{|\mathbf{y}|/4001}{k} \cdot \binom{|\mathbf{y}| + \ell}{|\mathbf{y}|}.$$

In the next section we give the description of $F_{\tilde{d}}(\mathbf{y})$ and then prove Lemma 6.

From Lemmas 5 and 6,

$$\begin{aligned} s &\geq \frac{\frac{1}{2}\cdot\binom{|\mathbf{y}|/4001}{k}\cdot\binom{|\mathbf{y}|+\ell}{|\mathbf{y}|}}{\binom{3\cdot|\mathbf{y}|/\tau}{k}\cdot\binom{|\mathbf{y}|+k\tau r+\ell}{|\mathbf{y}|}} \\ &\geq \frac{\frac{1}{2}\cdot\binom{|\mathbf{y}|/4001}{k}\cdot\frac{(|\mathbf{y}|+\ell)\ldots(1+\ell)}{|\mathbf{y}|!}}{\binom{3\cdot|\mathbf{y}|/\tau}{k}\cdot\frac{(|\mathbf{y}|+k\tau r+\ell)\ldots(1+k\tau r+\ell)}{|\mathbf{y}|!}} \\ &\geq \frac{\frac{1}{2}\cdot\binom{|\mathbf{y}|/4001}{k}}{\binom{3\cdot|\mathbf{y}|/\tau}{k}\cdot\left(1+\frac{k\tau r}{1+\ell}\right)^{|\mathbf{y}|}} \\ &\geq \frac{\frac{1}{2}\cdot\binom{|\mathbf{y}|/4001}{k}}{\binom{3\cdot|\mathbf{y}|/\tau}{k}\cdot e^{\frac{k\tau r}{1+\ell}\cdot|\mathbf{y}|}} \qquad \text{(from (1))} \\ &= \frac{\frac{1}{2}\cdot\binom{|\mathbf{y}|/4001}{k}}{\binom{3\cdot|\mathbf{y}|/\tau}{k}\cdot e^{\frac{0.006\tilde{d}\cdot|\mathbf{y}|}{|\mathbf{y}|+\ell}\cdot\frac{k\tau r}{0.006\tilde{d}}\cdot\frac{|\mathbf{y}|+\ell}{1+\ell}}} \\ &= \frac{\frac{1}{2}\cdot\binom{|\mathbf{y}|/4001}{k}}{\binom{3\cdot|\mathbf{y}|/\tau}{k}\cdot\binom{|\mathbf{y}|/4001}{k}^{\frac{k\tau r}{0.006\tilde{d}}\cdot\frac{|\mathbf{y}|+\ell}{1+\ell}}} \qquad \text{(from (8))} \\ &\geq \frac{\frac{1}{2}\cdot\binom{|\mathbf{y}|/4001}{k}}{\binom{3\cdot|\mathbf{y}|/\tau}{k}\cdot\binom{|\mathbf{y}|/4001}{k}^{\frac{k\tau r}{0.0055\tilde{d}}}} \qquad \text{(from (9))} \end{aligned}$$

$$\geq \frac{1}{2}\cdot\left(\frac{k\tau}{3e\cdot|\mathbf{y}|}\cdot\left(\frac{|\mathbf{y}|}{4001k}\right)^{(1-\frac{k\tau r}{\varepsilon\tilde{d}})}\right)^{k} \quad \text{(from (2))}$$

$$\geq \frac{1}{2}\cdot\left(\frac{\varepsilon\tilde{d}}{3e\cdot|\mathbf{y}|\cdot 21r}\cdot\left(\frac{|\mathbf{y}|\cdot 21\tau r}{4001\varepsilon\tilde{d}}\right)^{1-\frac{1}{21}}\right)^{\frac{\varepsilon\tilde{d}}{21\cdot\tau r}}$$

$$= \frac{1}{2}\cdot\left(\frac{1}{3e}\cdot\left(\frac{\varepsilon\tilde{d}}{21r\cdot|\mathbf{y}|}\right)^{\frac{1}{21}}\cdot\left(\frac{\tau}{4001}\right)^{\frac{20}{21}}\right)^{\frac{\varepsilon\tilde{d}}{21\tau r}}$$

$$= \frac{1}{2}\cdot\left(\left(\frac{1}{3e}\right)^{21}\cdot\frac{\varepsilon\tilde{d}}{21r\cdot|\mathbf{y}|}\cdot\left(\frac{\tau}{4001}\right)^{20}\right)^{\frac{\varepsilon\tilde{d}}{21\cdot 21\tau r}}$$

$$\geq \left(\frac{\tau^{20}\tilde{d}}{|\mathbf{y}|\cdot r}\right)^{\frac{\tilde{d}}{10^5\tau r}}.$$

□

In the rest of this section we prove Lemma 5.

Proof of Lemma 5. For brevity we drop the subscript i and rewrite Equation 6 as

$$T = Q_1\cdots Q_m.$$

We begin by observing that $\deg T \leq |\mathbf{y}|\cdot r$, and that $\deg Q_j \leq \tau r$ for every $j\in[m]$. Now, by grouping Q_j's that have degree less than $\tau r/2$ and multiplying out, it is possible to ensure that every grouping has degree between $\tau r/2$ and τr (except possibly one last grouping with degree less than $\tau r/2$). This grouping operation does not cost us as the lower bound in Lemma 4 is on the top fanin. Therefore, we assume without loss of generality that for every $j\in[m-1]$,

$$\begin{aligned}
 & \deg Q_j \geq && \tau r/2.\\
\Rightarrow\quad & \deg T \geq && (m-1)\tau r/2\\
\Rightarrow\quad & |\mathbf{y}|\cdot r \geq && (m-1)\tau r/2\\
\Rightarrow\quad & m \leq && 2|\mathbf{y}|/\tau + 1 \leq\ 3|\mathbf{y}|/\tau.
\end{aligned}$$

For the case $m\leq k$, we note that the elements of $\partial^{=k}(Q_1\dots Q_m)\cdot\mathbf{y}^{\leq\ell}$ are of degree at most $\deg(Q_1\dots Q_m)+\ell \leq m\tau r+\ell \leq k\tau r+\ell$. Hence $\mathsf{SP}_{k,\ell}(Q_1\dots Q_m)\leq\binom{|\mathbf{y}|+k\tau r+\ell}{|\mathbf{y}|}$, trivially proving the bound. For the case $k<m$, we use the claim below.

Claim 7. *If $k<m$ then*

$$\partial^{=k}\Big(\prod_{j\in[m]}Q_j\Big) \subseteq \operatorname{span}_{\mathbb{F}}\left(\bigcup_{A\in\binom{[m]}{m-k}}\Big(\mathbf{y}^{\leq k\tau r}\cdot\prod_{j\in A}Q_j\Big)\right).$$

Proof. We induct on k. The case $k=0$ is trivial. Suppose that the claim is true for $\tilde{k}=k-1$. To prove the case for k, we consider the element $b=\partial_{y_1y_2\dots y_k}(\prod_{j\in[m]}Q_j)\in\partial^{=k}(\prod_{j\in[m]}Q_j)$:

$$\begin{aligned}
b &= \partial_{y_1}\partial_{y_2y_3\dots y_k}\Big(\prod_{j\in[m]}Q_j\Big)\\
&\in \operatorname{span}_{\mathbb{F}}\left(\bigcup_{A\in\binom{[m]}{m-\tilde{k}}}\partial_{y_1}\Big(\mathbf{y}^{\leq\tilde{k}\tau r}\cdot\prod_{j\in A}Q_j\Big)\right),
\end{aligned}$$

from the inductive hypothesis. Let $\tilde{Q}\in\mathbf{y}^{\leq\tilde{k}\tau r}$ be a polynomial. Then from the product rule,

$$\begin{aligned}
&\partial_{y_1}\Big(\tilde{Q}\prod_{j\in A}Q_j\Big)\\
&= (\partial_{y_1}\tilde{Q})\prod_{j\in A}Q_j + \tilde{Q}\cdot\sum_{j\in A}(\partial_{y_1}Q_j)\cdot\prod_{\substack{i\in A\\ i\neq j}}Q_i\\
&\in \operatorname{span}_{\mathbb{F}}\bigcup_{B\in\binom{A}{|A|-1}}\mathbf{y}^{\leq k\tau r}\cdot\prod_{i\in B}Q_i,
\end{aligned}$$

as $\deg Q_j\leq\tau r$ for every j. Hence

$$\begin{aligned}
b &\in \operatorname{span}_{\mathbb{F}}\left(\bigcup_{A\in\binom{[m]}{m-\tilde{k}}}\bigcup_{B\in\binom{A}{|A|-1}}\mathbf{y}^{\leq k\tau r}\cdot\prod_{i\in B}Q_i\right)\\
&= \operatorname{span}_{\mathbb{F}}\left(\bigcup_{A\in\binom{[m]}{m-k}}\mathbf{y}^{\leq k\tau r}\prod_{j\in A}Q_j\right).
\end{aligned}$$

□

From the claim above it follows that

$$\begin{aligned}
&\partial^{=k}\Big(\prod_{j\in[m]} Q_j\Big)\cdot \mathbf{y}^{\le \ell} \\
&\subseteq \operatorname{span}_{\mathbb{F}}\left(\bigcup_{A\in\binom{[m]}{m-k}} \Big(\mathbf{y}^{\le k\tau r+\ell}\cdot \prod_{j\in A} Q_j\Big)\right) \\
&\Rightarrow \mathsf{SP}_{k,\ell}\Big(\prod_{j\in[m]} Q_j\Big) \\
&\le \binom{m}{k}\cdot |\mathbf{y}^{\le k\tau r+\ell}| \\
&\le \binom{3|\mathbf{y}|/\tau}{k}\cdot\binom{|\mathbf{y}|+k\tau r+\ell}{|\mathbf{y}|},
\end{aligned}$$

as $m \le 3|\mathbf{y}|/\tau$. □

VI. CONSTRUCTING $F_{\tilde{d}}(\mathbf{y})$ AND PROVING LEMMA 6

This section is devoted to constructing the hard polynomial $F = F_{\tilde{d}}$ mentioned in step 3 of the proof outline in Section II and showing that it has a high SP measure. In Section IV we mentioned two properties $F_{\tilde{d}}(\mathbf{y})$ would have. The claim below (which is essentially taken from (Chillara and Mukhopadhyay, 2014) with suitable adjustments) makes them precise and shows how they ensure a high SP measure for F, something that Lemma 6 claims. Let $D = \binom{|\mathbf{y}|/4001}{k}$.

Claim 8. *Suppose $\partial^{=k}F_{\tilde{d}}(\mathbf{y})$ contains at least $\binom{|\mathbf{y}|/4001}{k}$ monomials (as individual elements) such that they all are of the same degree and have pairwise distance at least $\delta = 0.006\tilde{d}$. Then $\mathsf{SP}_{k,\ell}(F_{\tilde{d}}) \ge \frac{1}{2}\binom{|\mathbf{y}|/4001}{k}\binom{|\mathbf{y}|+\ell}{|\mathbf{y}|}$.*

Proof. Let $\mu_1,\dots,\mu_D$ be the monomials present in $\partial^{=k}F_{\tilde{d}}(\mathbf{y})$, of degree d_0 (say) each, and pairwise distance at least δ. Then from the inclusion-exclusion principle

$$\begin{aligned}
&|\mathbf{y}^{\le \ell}\cdot\{\mu_a\}_{a\in[D]}| \\
&\ge \sum_{a=1}^{D} |\mathbf{y}^{\le \ell}\cdot \mu_a| - \sum_{1\le a<b\le D} |(\mathbf{y}^{\le \ell}\cdot\mu_a)\cap(\mathbf{y}^{\le \ell}\cdot\mu_b)|.
\end{aligned} \tag{10}$$

Clearly $|\mathbf{y}^{\le \ell}\cdot\mu_a| = |\mathbf{y}^{\le \ell}| = \binom{|\mathbf{y}|+\ell}{|\mathbf{y}|}$. Next, let us estimate an upper bound on the size of the set $(\mathbf{y}^{\le \ell}\cdot\mu_a)\cap(\mathbf{y}^{\le \ell}\cdot\mu_b) = I_{a,b}$ (say). It is given that the elements of $I_{a,b}$ are of degree at most $d_0+\ell$ and that the LCM(μ_a,μ_b) is of degree at least $d_0+\delta$. Hence

$$\begin{aligned}
|I_{a,b}| &= |I_{a,b}/\mathrm{LCM}(\mu_a,\mu_b)| \\
&\le |\mathbf{y}^{\le d_0+\ell-(d_0+\delta)}| \\
&= \binom{|\mathbf{y}|+\ell-\delta}{|\mathbf{y}|}.
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \sum_{1\le a<b\le D} |(\mathbf{y}^{\le \ell}\cdot\mu_a)\cap(\mathbf{y}^{\le \ell}\cdot\mu_b)| \\
&= \sum_{1\le a<b\le D} |I_{a,b}| \\
&\le \frac{D^2}{2}\binom{|\mathbf{y}|+\ell-\delta}{|\mathbf{y}|} \\
&= \frac{D^2}{2}\cdot\frac{(|\mathbf{y}|+\ell)\dots(1+\ell)}{|\mathbf{y}|!} \\
&\quad\cdot\frac{(|\mathbf{y}|+\ell-\delta)\dots(1+\ell-\delta)}{(|\mathbf{y}|+\ell)\dots(1+\ell)} \\
&\le \frac{D^2}{2}\cdot\binom{|\mathbf{y}|+\ell}{|\mathbf{y}|}\cdot\Big(1-\frac{\delta}{|\mathbf{y}|+\ell}\Big)^{|\mathbf{y}|} \\
&\le \frac{D}{2}\cdot\binom{|\mathbf{y}|+\ell}{|\mathbf{y}|}\cdot D\cdot e^{-\frac{\delta\cdot|\mathbf{y}|}{|\mathbf{y}|+\ell}} \quad \text{(from (1))} \\
&= \frac{D}{2}\cdot\binom{|\mathbf{y}|+\ell}{|\mathbf{y}|} \quad \text{(from (8))}.
\end{aligned}$$

Plugging the bounds in Equation (10) we get

$$\begin{aligned}
&|\mathbf{y}^{\le \ell}\cdot\{\mu_a\}_{a\in[D]}| \\
&\ge D\cdot\binom{|\mathbf{y}|+\ell}{|\mathbf{y}|} - \frac{D}{2}\cdot\binom{|\mathbf{y}|+\ell}{|\mathbf{y}|} \\
&= \frac{1}{2}\binom{|\mathbf{y}|/4001}{k}\binom{|\mathbf{y}|+\ell}{|\mathbf{y}|},
\end{aligned}$$

by plugging the value for D. Now $\mathbf{y}^{\le \ell}\cdot\{\mu_a\}_{a\in[D]}$, being a set of monomials, is linearly independent. Hence $\mathsf{SP}_{k,\ell}(F_{\tilde{d}}) \ge \frac{1}{2}\binom{|\mathbf{y}|/4001}{k}\binom{|\mathbf{y}|+\ell}{|\mathbf{y}|}$. □

i. Description of $F_{\tilde{d}}(\mathbf{y})$

We show an explicit construction of same degree monomials $\mu_1, \ldots, \mu_D$, with large pairwise distance, using the $\mathbf{y}$-variables. Let $\mathbf{z}$ be a subset of $\mathbf{y}$, of size $n = \frac{\tilde{d}-k}{0.9} \cdot \frac{4000}{4001} < \frac{4000}{4001} \cdot |\mathbf{y}|$, as $\tilde{d} \leq 0.9 \cdot |\mathbf{y}|$. Note that $n = \Theta(N)$. We partition $\mathbf{z}$ into $\frac{n}{c \log n} = n_0$ (say) disjoint subsets of size $c \log n$ each and call them $Z^{(i)}$, $i \in [n_0]$. Here c is a constant in $[1000, 2000]$, chosen in such a way that n_0 is a prime number. Now, we apply the following claim, whose proof is essentially a well known probabilistic argument (with an associated greedy algorithm) for existence of codes with good distance (akin to the Gilbert-Varshamov bound (Gilbert, 1952; Varshamov, 1957)). The proof is given in the next subsection.

Claim 9. *For every $i \in [n_0]$, there is a set $M^{(i)}$ of n multilinear monomials (in $Z^{(i)}$ variables) each of degree $0.9 \cdot \frac{4001}{4000} \cdot c \log n$ and pairwise distance at least $0.007 c \log n$. Furthermore, $M^{(i)}$ can be generated in* $\mathsf{poly}(n)$ *time.*

Thus $M^{(i)}$ has at least $n \geq n_0$ monomials. Let us identify $\eta^{(i)}_1, \ldots, \eta^{(i)}_{n_0}$ with the (lexicographically first n_0 many) monomials of $M^{(i)}$. Let $\mathbb{K}$ be a prime field of size n_0. Elements of $\mathbb{K}$ will be denoted with $1, 2, 3, \ldots, n_0$. Finally, we define μ_b, where $b \in [D]$, as the b-th element of the following set that is ordered according to lexicographic ordering of the coefficient vectors of the defining univariate polynomials.

$$L \stackrel{\text{def}}{=} \left\{ \prod_{i \in [n_0]} \eta^{(i)}_{h(i)} \right\}_{\substack{h \in \mathbb{K}[t], \\ \deg h = 0.1 n_0, \\ h \text{ is monic}}} . \qquad (11)$$

For example, the first element of L is the one corresponding to the monic, degree-$0.1n_0$ univariate polynomial $h \in \mathbb{K}[t]$ whose coefficient vector is lexicographically the smallest. At the end of this section we show that indeed $|L| \geq D$ (so the definition above, which is inspired by Reed-Solomon codes, makes sense). Observe, μ_i's are multilinear and of degree $\tilde{d} - k$.

Defining $F_{\tilde{d}}(\mathbf{y})$. The construction uses the idea of 'code composition' that ensures $F_{\tilde{d}}(\mathbf{y})$ is a VNP-polynomial (see Subsection iii). From $\mathbf{y} \setminus \mathbf{z}$ one can form $\binom{|\mathbf{y} \setminus \mathbf{z}|}{k} = D$ many multilinear monomials of degree k, as $|\mathbf{y} \setminus \mathbf{z}| \geq |\mathbf{y}|/4001$. Let us call these monomials $\nu_1 \prec \nu_2 \prec \ldots \prec \nu_D$, under lexicographic ordering. Then we define $F_{\tilde{d}}(\mathbf{y})$ as follows:

$$F_{\tilde{d}}(\mathbf{y}) \stackrel{\text{def}}{=} \sum_{b=1}^{D} \mu_b \nu_b. \qquad (12)$$

Clearly $F_{\tilde{d}}(\mathbf{y})$ is multilinear and all its monomials are of degree $\tilde{d}$. Since $\partial_{\nu_b}(\mu_b \nu_b) = \mu_b$, $\partial^{=k}(F_{\tilde{d}}(\mathbf{y}))$ contains μ_b's as required by Claim 8. The other requirement, namely that $F_{\tilde{d}}$-monomials have a minimum pairwise distance of $0.006\tilde{d} = \delta$, is also satisfied: Consider two monomials $\mu_b \nu_b$ and $\mu_a \nu_a$, where $b \neq a$. It suffices to show that $|\mu_b \setminus \mu_a| \geq \delta$. Indeed, we have $\mu_b = \prod_{i \in [n_0]} \eta^{(i)}_{h(i)}$ and $\mu_a = \prod_{i \in [n_0]} \eta^{(i)}_{g(i)}$, for two different monic univariate polynomials $h, g \in \mathbb{K}[t]$ of degree $0.1n_0$. If R is the set of at most $0.1n_0$ roots of $h - g$ in $[n_0]$ then clearly $h(i) \neq g(i)$ for $i \in [n_0] \setminus R$. Hence from Claim 9 we have $|\eta^{(i)}_{h(i)} \setminus \eta^{(i)}_{g(i)}| \geq 0.007 c \log n$, for $i \in [n_0] \setminus R$. As $M^{(i)}$ and $M^{(j)}$ are variable-disjoint for $i \neq j$, we have

$$\begin{aligned}
& |\mu_b \setminus \mu_a| \\
&= \sum_{i \in [n_0]} |\eta^{(i)}_{h(i)} \setminus \eta^{(i)}_{g(i)}| \\
&\geq \sum_{i \in [n_0] \setminus R} |\eta^{(i)}_{h(i)} \setminus \eta^{(i)}_{g(i)}| \\
&\geq (n_0 - 0.1n_0) \cdot 0.007 c \log n \\
&> 0.006n \\
&> 0.006\tilde{d},
\end{aligned}$$

where the last step follows from the expression for n and noting therein that $k = o(\tilde{d})$ (from (7)).

Verifying that $|L| \geq D$. The nonzero pairwise distance implies that $|L| = |\{h : \ h \in \mathbb{K}[t], \deg h = 0.1n_0, \ h \text{ is monic}\}|$, which is at least $|\mathbb{K}|^{0.1n_0} = n_0^{0.1n_0}$. Hence $\log|L| \geq 0.1n_0 \log n_0 > 0.1 \cdot \frac{n}{2c} = \Theta(\tilde{d})$ (for large enough n). On the other hand, $\log D = \log\binom{|\mathbf{y}|/4001}{k} \leq \log(\frac{e \cdot |\mathbf{y}|}{4001k})^k = k\log\frac{e\cdot|\mathbf{y}|}{4001k}$, from Bound (2). But from Equation (7), $k = O(\sqrt{\frac{\tilde{d}}{r\log\tilde{d}}})$, thus $\log D = O(\sqrt{\tilde{d}})$ as both $|\mathbf{y}|$ and $\tilde{d}$ are $\Theta(N)$, proving $|L| \geq D$.

Proof of Lemma 6. $F_{\tilde{d}}(\mathbf{y})$ is in VNP (see Subsection iii) and meets the conditions required by Claim 8, which implies the result. □

ii. A greedy algorithm

Proof of Claim 9. For brevity, let $\tilde{\varepsilon} = 0.9 \cdot \frac{4001}{4000} < 0.91$. In Algorithm 1 we outline a greedy way to construct the required monomials. Clearly, Algorithm 1 runs in $\mathsf{poly}(n)$ time, and the output monomials have the required degree and distance. It remains to show that as long as $|M^{(i)}| < n$, there is some $\tilde{j} > j$ such that $\alpha_{\tilde{j}}$ can be included in $M^{(i)}$. We use the probabilistic method for this purpose, as below.

Algorithm 1: A greedy algorithm to generate distant monomials

```
Input   : The variables $Z^{(i)}$
Output : The set of monomials $M^{(i)}$
Let $\alpha_1, \alpha_2, ..., \alpha_t$, where $t = \binom{c\log n}{\tilde{\varepsilon}c\log n}$, be
  multilinear monomials of degree $\tilde{\varepsilon}c\log n$,
  in lexicographical order.
$M^{(i)} := \emptyset$
$j := 1$
while $|M^{(i)}| < n$ and $j \leq t$ do
    if $|\alpha_j \setminus \eta| \geq 0.007c\log n$ for all
      $\eta \in M^{(i)}$ then
        $M^{(i)} := M^{(i)} \cup \{\alpha_j\}$
    end
    $j := j + 1$
end
return $M^{(i)}$
```

Consider picking every variable independently with probability $\frac{\tilde{\varepsilon}}{0.99} < 1$ and multiplying the picked variables to form a monomial μ (say). Then $E[\deg \ \mu] = \frac{\tilde{\varepsilon}}{0.99} \cdot c\log n$. From Chernoff bound,

$$\Pr\left[\deg \ \mu < 0.99 \cdot \frac{\tilde{\varepsilon}}{0.99} \cdot c\log n\right]$$

$$\leq e^{-\frac{0.01^2}{3} \cdot \frac{\tilde{\varepsilon}}{0.99} \cdot c\log n}$$

$$< e^{-0.00003c\log n}$$

$$= e_1 \quad \text{(say)}.$$

Let η be some fixed monomial from $M^{(i)}$. Then

$$E[|\eta \setminus \mu|]$$

$$= \sum_{i=1}^{\tilde{\varepsilon}c\log n} (1 - \frac{\tilde{\varepsilon}}{0.99})$$

$$= \tilde{\varepsilon} \cdot (1 - \frac{\tilde{\varepsilon}}{0.99}) \cdot c\log n.$$

Thus

$$\Pr[|\eta \setminus \mu| < \ 0.1 \cdot (\tilde{\varepsilon} \cdot (1 - \frac{\tilde{\varepsilon}}{0.99}) \cdot c\log n)]$$

$$\leq e^{-\frac{0.9^2}{3} \cdot \tilde{\varepsilon} \cdot (1 - \frac{\tilde{\varepsilon}}{0.99}) \cdot c\log n}.$$

$$\Rightarrow \Pr[|\eta \setminus \mu| < \ 0.007c\log n]$$

$$\leq e^{-0.022c\log n},$$

from Chernoff bound. From union bound, the probability that there is a monomial $\nu \in M^{(i)}$ with $|\Delta(\nu, \mu)| < \ 0.007c\log n$ is at most

$$|M^{(i)}| \cdot e^{-0.022 \cdot c\log n}$$

$$\leq ne^{-0.022c\log n}$$

$$\leq e^{0.001c\log n} \cdot e^{-0.022c\log n}$$

$$= e^{-0.021c\log n}$$

$$= e_2 \quad \text{(say)}.$$

Thus, μ has degree at least $\tilde{\varepsilon}c\log n$ and distance $|\nu \setminus \mu|$ at least $0.007c\log n$ for all $\nu \in M^{(i)}$ with probability at least $1 - e_1 - e_2 =$

$1 - e^{-0.00003\cdot c\log n} - e^{-0.021c\log n} \gg 0$ (for n large enough). In other words, there exists a multilinear monomial μ with distance (from monomials of $M^{(i)}$) at least $0.007\log n$ and degree at least $\tilde{\varepsilon}\cdot c\log n$. However we want the degree to be exactly $\tilde{\varepsilon}c\log n$. We can chop off a few variables from μ to ensure that. Such a chopping results in $|\mu \setminus \nu| = |\nu \setminus \mu| \geq 0.007c\log n$, as desired. $\square$

iii. VNP membership of $F_{\tilde{d}}$ and H

Proof of $F_{\tilde{d}}(\mathbf{y}) \in \mathsf{VNP}$. We recall Equation (12). According to Valiant's criterion, it suffices to give a $\mathsf{poly}(|\mathbf{y}|)$-time procedure that checks if a given monomial equals $\mu_b\nu_b$ for some $b \in D$. (The coefficient is 1 if it does and 0 otherwise.) The procedure is as follows. We call the $\mathbf{z}$-part of the input monomial as μ_b, where b is unknown. We determine b by writing μ_b in the form $\prod_{i\in[n_0]} \eta^{(i)}_{h(i)}$ (as per Equation (11)) and determining h first, using polynomial interpolation. From h, the index b can be computed efficiently as the ordering of the set L (in Equation (11)) is quite explicit. Finally, from b we can efficiently compute ν_b following lexicographic ordering and check if the non-$\mathbf{z}$-part of the input monomial is ν_b as well. All the steps above can be done in $\mathsf{poly}(|\mathbf{y}|)$ time.

Proof of $H_{\tilde{d}}(\mathbf{x},\mathbf{u},\mathbf{v}) \in \mathsf{VNP}$. We recall Equation (4). H is constructed in such a way that for every $\mathbf{y} \in \binom{\mathbf{x}}{\rho}$, a unique monomial in $\mathbf{u}$ and $\mathbf{v}$ variables is attached to the monomials of $F_{\tilde{d}}(\mathbf{y})$. Thus, given a monomial, we can easily find which $F_{\tilde{d}}$ its $\mathbf{y}$-part potentially 'belongs to' and then run the procedure described above that checks membership in $F_{\tilde{d}}(\mathbf{y})$.

VII. Acknowledgement

We thank Neeraj Kayal for helpful discussions. Thanks also to the anonymous reviewers for their useful comments and suggestions that have improved the presentation of this paper.

References

Agrawal, M. (2005). Proving lower bounds via pseudo-random generators. In *Proceedings of the 25th Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, pages 92–105.

Agrawal, M. and Vinay, V. (2008). Arithmetic circuits: A chasm at depth four. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 67–75.

Bürgisser, P. (2000). Cook's versus Valiant's hypothesis. *Theoretical Computer Science - Selected papers in honor of Manuel Blum*, 235:71–78.

Chillara, S. and Mukhopadhyay, P. (2014). Depth-4 Lower Bounds, Determinantal Complexity : A Unified Approach. In *32nd Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 239–250.

Gilbert, E. N. (1952). A comparison of signalling alphabets. *Bell System Technical Journal*, 31:504–522.

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

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

Kalorkoti, K. (1985). A lower bound for the formula size of rational functions. *SIAM journal of Computing*, 14(3):678–687.

Kayal, N. (2012). An exponential lower bound for the sum of powers of bounded degree polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:81.

Kayal, N. and Saha, C. (2015). Multi-k-ic depth three circuit lower bound. In *Proceedings of the 32nd Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 30, pages 527–539.

Kayal, N., Saha, C., and Saptharishi, R. (2014). A super-polynomial lower bound for regular arithmetic formulas. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 146–153.

Kayal, N., Saha, C., and Tavenas, S. (2016a). An almost cubic lower bound for depth three arithmetic circuits. In *43rd International Colloquium on Automata, Languages and Programming (ICALP)*, pages 33:1–33:15.

Kayal, N., Saha, C., and Tavenas, S. (2016b). On the size of homogeneous and of depth four formulas with low individual degree. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing (STOC)*, pages 626–632.

Koiran, P. (2012). Arithmetic circuits: The chasm at depth four gets wider. *Theoretical Computer Science*, 448:56–65.

Nisan, N. and Wigderson, A. (1994). Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167.

Raz, R. (2009). Multi-linear formulas for permanent and determinant are of super-polynomial size. *Journal of the Association for Computing Machinery*, 56(2):8:1–8:17.

Raz, R. (2013). Tensor-rank and lower bounds for arithmetic formulas. *J. ACM*, 60(6):40:1–40:15.

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

Raz, R. and Yehudayoff, A. (2009). Lower Bounds and Separations for Constant Depth Multilinear Circuits. *Computational Complexity*, 18(2):171–207.

Reed, I. S. and Solomon, G. (1960). Polynomial codes over certain finite fields. *Journal of the Society for Industrial and Applied Mathematics (SIAM)*, 8(2):300–304.

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

Strassen, V. (1973). Vermeidung von divisionen. *Journal für die reine und angewandte Mathematik*, 264:184–202.

Tavenas, S. (2013). Improved bounds for reduction to depth 4 and depth 3. *Information and Computation*, pages 240:2–11.

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

Valiant, L. G. (1979). Completeness Classes in Algebra. In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 249–261, New York, NY, USA. ACM Press.

Varshamov, R. R. (1957). Estimate of the number of signals in error correcting codes. *Dokl. Acad. Nauk SSSR*, 117:739–741.