Bootstrapping variables in algebraic circuits

Manindra Agrawal * Sumanta Ghosh † Nitin Saxena ‡

Abstract

We show that for the blackbox polynomial identity testing (PIT) problem it suffices to study circuits that depend only on the first extremely few variables. One only need to consider size-s degree-s circuits that depend on the first $\log^{\circ c} s$ variables (where c is a constant and we are composing c logarithms). Thus, hitting-set generator (hsg) manifests a *bootstrapping* behavior— a partial hsg against very few variables can be efficiently grown to a complete hsg. A boolean analog, or a pseudorandom generator property of this type, is unheard-of. Our idea is to use the partial hsg and its annihilator polynomial to efficiently bootstrap the hsg exponentially wrt variables. This is repeated c times in an efficient way.

Pushing the envelope further we show that: **(1)** a quadratic-time blackbox PIT for 6913-variate degree-s size-s polynomials, will lead to a "near"-complete derandomization of PIT, and **(2)** a blackbox PIT for n-variate degree-s size-s circuits in s^{n^δ}-time, for $\delta < 1/2$, will lead to a "near"-complete derandomization of PIT (in contrast, s^n-time is trivial).

Our second idea is to study depth-4 circuits that depend on constantly many variables. We show that a polynomial-time computable, $O(s^{1.49})$-degree hsg for *trivariate* depth-4 circuits bootstraps to a quasipolynomial time hsg for general poly-degree circuits, and implies a lower bound that is a bit stronger than Kabanets-Impagliazzo (STOC 2003).

2012 ACM CCS concept: Theory of computation– Algebraic complexity theory, Fixed parameter tractability, Pseudorandomness and derandomization; Computing methodologies– Algebraic algorithms; Mathematics of computing– Combinatoric problems.
Keywords: hitting-set, tiny, depth-3, depth-4, derandomization, identity testing, lower bound, VP, VNP, E, #P/poly, SUBEXP, NW design, circuit factoring, approximative, tetration.

1 Introduction

Polynomial identity testing (PIT) problem is to decide whether a multivariate polynomial is zero, where the input is given as an algebraic circuit. An algebraic circuit over a field $\mathbb{F}$ is a layered acyclic directed graph with one sink node called output node; source nodes are called input nodes and are labeled by variables or field constants; non-input nodes are labeled by $\times$ (multiplication gate) and $+$ (addition gate) in alternate layers. Sometimes edges may be labeled by field constants. The computation is defined in a natural way. The complexity parameters of a circuit are: 1) *size*- number of edges and vertices (including the *variables*), 2) *depth*- number of layers, and 3) *degree*- maximum degree among all polynomials computed at each node. Note– The degree of the computed polynomial may be much smaller than the degree of its circuit.

The polynomial computed by a circuit may have, in the worst-case, an exponential number of monomials compared to its size. So, by computing the explicit polynomial from input circuit, we cannot solve PIT problem in polynomial time. However, evaluation of the polynomial at a point can be done, in time polynomial in the circuit size, by assigning the values at input

*Department of Computer Science & Engineering, IIT Kanpur, India, manindra@cse.iitk.ac.in
†CSE, IITK, sumghosh@cse.iitk.ac.in
‡CSE, IITK, nitin@cse.iitk.ac.in

nodes. This helps us to get a polynomial time randomized algorithm for PIT by evaluating the circuit at a random point, since any nonzero polynomial evaluated at a random point gives a nonzero value with high probability [DL78, Zip79, Sch80]. However, finding a deterministic polynomial time algorithm for PIT is a long-standing open question in algebraic complexity theory. It naturally appears in the algebraic-geometry approaches to the P$\neq$NP question, eg. [Mul17, GMQ16, Gro15, Mul12b, Muk16]. The famous algebraic analog is the VP$\neq$VNP question [Val79]. The PIT problem has applications both in proving circuit lower bounds [HS80, KI03, Agr05] and in algorithm design [MVV87, AKS04, KSS14, DdOS14]. For more details on PIT, see the surveys [Sax09, Sax13, SY10] or review articles [Wig17, Mul12a].

PIT algorithms are of two kinds: 1) *whitebox*- use the internal structure of the circuit, and 2) *blackbox*- only evaluation of the circuit is allowed at points in a 'small' extension $\mathbb{K} \supseteq \mathbb{F}$. Blackbox PIT for a set of polynomials $\mathcal{P} \subset \mathbb{K}[\mathbf{x}]$ is equivalent to efficiently finding points $\mathcal{H} \subset \mathbb{K}^n$, called a *hitting-set*, such that for any nonzero $P \in \mathcal{P}$, the set $\mathcal{H}$ contains a point at which $P \neq 0$. For us a more functional approach would be convenient. We think in terms of an n-tuple of univariates $\mathbf{f}(y) = (f_1(y), \ldots, f_n(y))$, in $\mathbb{K}[y]$, whose set of evaluations contain an $\mathcal{H}$. Such an $\mathbf{f}(y)$ can be efficiently obtained from a given $\mathcal{H}$ (using interpolation) and vice-versa. Clearly, if $\mathcal{H}$ is a hitting-set for $\mathcal{P}$ then $P(\mathbf{f}(y)) \neq 0$, for any nonzero $P \in \mathcal{P}$. This tuple of univariates is called a hitting-set generator (*hsg*) and its *degree* is $\max_{i\in[n]} \deg(f_i)$, which is $\leq |\mathcal{H}|$.

Our work. We study the phenomenon of *bootstrapping*: converting an hsg for size-s degree-s n-variate circuits to hsg for size-s degree-s $L(n)$-variate circuits with $L(n) > n$. In the boolean settings, this phenomenon is well understood. The analog of hsg is *pseudo-random generator (prg)* that *stretches a seed* by several bits, or, the *s-extender* that stretches n by a *single* bit. By [NW94, Sec.2-3] it is known that an extender for size-s $(\log s)$-variate boolean circuits can be converted to an optimal prg for size-s circuits with $L(n) = 2^n$. No further "reduction" in number of variables is possible since the size of a $(\epsilon \log s)$-variate circuit can be reduced to $< s$ if $\epsilon < 1$.

The situation is less clear in algebraic settings. On one hand, n-variate polynomials requiring circuits of size s exist for every n and s (due to the fact that polynomials can have arbitrarily large degrees unlike boolean settings where every function is multilinear). On the other hand, bootstrapping from $O(\log s)$ variables to s variables is not studied explicitly in the literature.

We close this gap in knowledge by showing that an hsg for size-s degree-s $(\log^{\circ c} s)$-variate circuit can be efficiently converted to an hsg for size-s degree-s s-variate circuit; where $\log^{\circ c} s := \log \cdots (c\,times) \cdots \log s$. Furthermore, at the cost of making the final hsg slightly superpolynomial $= s^{\exp\circ\exp(O(\log^\star s))}$, we show that bootstrapping can be done from even a *constant* number of variables! Our results can also be viewed as a powerful *amplification* of derandomization: a "slight" derandomization ($= s^{n^\delta}$ time hsg for size-s degree-s n-variate circuits, for a constant $\delta < 1/2$) implies "nearly" complete derandomization ($= s^{\exp\circ\exp(O(\log^\star s))}$ time hsg for size-s degree-s s-variate circuits). Compare the required s^{n^δ}-time PIT with the trivial s^n-time PIT.

We prove an additional result for shallow circuits: poly(s)-time computable and $O\left(s^{n/2}/\log^2 s\right)$ degree hsg for size-s n-variate depth four circuits (for some constant $n \geq 3$) implies quasipolynomial time blackbox PIT for size-s degree-s s-variate circuits (& strong exponential lower bounds). See Theorems 1–4 for more formal statements.

We see our results as a positive development; since, they reduce PIT to cases that are special in an unprecedented way. Such special-case PIT algorithms are waiting to be discovered.

Existing deterministic algorithms solving PIT for restricted classes have been developed by leveraging insight into their weaknesses. For example, deterministic PIT algorithms are known for subclasses of depth-3 circuits [KS07, Sax08, SS12], subclasses of depth-4 circuits [ASSS12, BMS13, SSS13, For15, KS16a, KS16b, PSS16], read-once algebraic branching programs (ROABP) and related models [FS12, ASS13, FSS14, AGKS15, GKST16, GKS17, MV17], certain symbolic determinants [FGT16, GT16, ST17, GTV17], as well as non-commutative models

[LMP16, GGOW16, LLS17]. An equally large number of special models have been used to prove lower bounds, see for example the ongoing online survey of Saptharishi [Sap16]. Also, blackbox PIT relates to conjectures that bar certain algebraic circuit lower bound methods [FSV17].

Our notation. $[n]$ refers to $\{1, 2, \ldots, n\}$. Logarithms are wrt base 2. Iterated logarithm $\log^\star s$ is the least number of iterated applications of log that gives a result ≤ 1. When we say that a circuit is of size-s (resp. depth-Δ, or degree-d) we use the parameters as an *upper* bound.

Field: To appreciate the most important aspects of this work keep in mind the "practical" fields $\mathbb{F} = \mathbb{Q}$ or $\mathbb{F}_q$. Interestingly, our main theorems (Thms. 1–4) hold for *any* field. However, the other theorems require field characteristic to be zero or large. Common examples are: complex $\mathbb{C}$, reals $\mathbb{R}$, algebraic numbers $\overline{\mathbb{Q}}$, local fields $\mathbb{Q}_p$ or their extensions, or finite fields $\mathbb{F}_q$ of characteristic $p >$ degree of the input.

Finally, one can generalize our work to the field $K = \mathbb{F}(\epsilon)$ with $\epsilon \to 0$ in a certain way. This leads to *approximative* complexity $\overline{\text{size}}$ of polynomials in $\mathbb{F}[\mathbf{x}]$ [Bür01, Defn.3.1]. Efficient hitting-sets wrt $\overline{\text{size}}$ are equivalent to explicit system of parameters (*esop*) of the invariant ring of a related variety $\Delta[\det(X), s]$ with a given group action [Mul17, Thm.4.9]. Our work (Theorem 4) will imply that to prove the existence of such a (quasi-)esop it suffices to study esop wrt X that depend on 'constantly few' variables (also see the reduction of derandomized Noether Normalization problem NNL to blackbox PIT in [Mul17, Sec.4.3]).

A basic algebraic algorithm used in our results is circuit factoring, that relies on field properties. A classic result is [Kal89] that constructs small circuits for factors that have multiplicity coprime to the characteristic (see [DSS17] for recent factoring results and the related rich background).

Hitting-set generator (hsg): Let $\mathcal{P}$ be a set of n-variate polynomials. We call an n-tuple of univariates $\mathbf{f}(y) = (f_1(y), \ldots, f_n(y))$ a (t, d)-*hsg* for $\mathcal{P}$ if: (1) for any nonzero $P \in \mathcal{P}$, $P(\mathbf{f}(y)) \neq 0$, and (2) $\mathbf{f}$ has time-complexity t and the degree of each f_i is less than d. By t-*time hsg* or t-*time hitting-set* or t-*time blackbox PIT*, we always mean a (t, t)-hsg.

The computational problem of designing and verifying an hsg for size-s circuits family is in PSPACE; however, that for $\overline{\text{size}}$-s circuits family is in EXPSPACE (recently brought down to PSPACE [GSS18, FS17]). The major open question is to bring this complexity down to P; this is christened 'GCT Chasm' in [Mul17, Sec.11] and has since then become a fundamental difficulty common to geometry and complexity theories. It means that we have to discover algebraic properties that are specific to only those polynomials that have *small* circuit representation. We will investigate such properties closely in this work.

Variables: A polynomial P computed by a size-s algebraic circuit C can have at most $\{x_1, \ldots, x_s\}$ variables. For $k < s$, if we say that C *depends only on the first k variables*, then it is meant that the computed polynomial $P \in \mathbb{F}[x_1, x_2, \ldots, x_k]$.

Multi-δ-ic: A polynomial family $\{f_n(x_1, \ldots, x_n)\}_{n\geq 1}$ over a field $\mathbb{F}$ is called *multi-δ-ic*, if degree of each variable in f_n is less than δ. For eg. when $\delta = 2$, $\{f_n\}_{n\geq 1}$ is multilinear.

E-computable polynomial family: For constant δ, a multi-δ-ic polynomial family $\{f_n\}_n$ with integer coefficients is called *E-computable* if: there exists a $2^{O(n)}$-time algorithm that on input $\mathbf{e}$, outputs the coefficient of $\mathbf{x}^{\mathbf{e}}$ in f_n in binary; say the leading bit will denote the *sign* of the coefficient, with 0 implying a positive coefficient and 1 implying negative. This makes $\text{coeff}(\cdot)(f_n)$ a boolean function ($\{0,1\}^* \to \{0,1\}^*$) whose bits are E-computable as well.

1.1 Our motivation and main results

Pseudorandom generator (*prg*) is a well studied object in boolean circuit complexity theory and cryptography [Yao82] & [AB09, Chap.10]. One of the main motivations of studying prg is to efficiently derandomize all randomized algorithms. Indeed one can show that if we have an *optimal prg* against BPP, then BPP=P. By optimal prg, we mean a prg which stretches an

n-length string to 2^n-length and is computable in $2^{O(n)}$ time. Interestingly, an optimal prg is closely related to strong circuit lower bound. It is a celebrated result that designing optimal prg against P/poly is equivalent to finding an E-computable boolean function which has boolean circuit complexity $2^{\Omega(n)}$ [NW94, Secs.2.5 & 3.1].

Naturally, an algebraic analog of the latter property would be to identify an E-computable polynomial family which has *algebraic* circuit complexity $2^{\Omega(n)}$. By Valiant's criterion [Bür13, Prop.2.20] if one replaces E by #P/poly then we are directly talking about a *strong* version of VNP≠VP. As a first challenge, we can pose the following reasonable complexity conjecture.

Conjecture 1. There is an E-computable polynomial which has algebraic complexity $2^{\Omega(n)}$. Thus, either $E \nsubseteq$ #P/poly or VNP has a polynomial family of algebraic circuit complexity $2^{\Omega(n)}$.

In the world of algebraic circuits, hitting-set generator (hsg) is in direct analogy with prg. So one can naturally ask about the relation between hsg and algebraic circuit lower bound. Heintz and Schnorr [HS80, Thm.4.5] introduced the concept of an efficient annihilator of the hsg. They showed that if we can efficiently compute an hsg for a set of polynomials $\mathcal{P}$, then we can also efficiently compute a polynomial (namely, annihilator) which does not belong to $\mathcal{P}$. This technique can be easily extended to get the following circuit lower bound result. Like boolean world, our hard polynomial is also E-computable but has algebraic circuit complexity $2^{\Omega(n)}$.

Theorem 0 (Connection). If we have poly(s)-time blackbox PIT for size-s degree-s circuits $\mathcal{P}_s$, then Conjecture 1 holds. (Proof sketched in Section A.)

A weak converse of the above theorem, i.e. hardness to hsg, is well-known due to [KI04, Thm.7.7]. We state a revised version of it as Lemma 9. If we have an exponentially hard but E-computable polynomial family, then by using Lemma 9 we can efficiently reduce the number of variables in any circuit, from s to $O(\log s)$, preserving the nonzeroness. Next, one applies a "trivial" hitting-set on the $O(\log s)$ variables, which gives a quasipolynomial time hsg for $\mathcal{P}_s$ [CDGK91]. This suggests that the 'hardness vs randomness' connection here is less satisfactory than the boolean world. Nonetheless, one wonders whether the conclusion in Theorem 0 can be strengthened in a different way, so that we get a perfect equivalence. In this work, we answer this question by introducing the concept of partial hsg. Indeed, we give infinitely many different-looking statements that are all equivalent to the hypothesis in Theorem 0.

Partial hsg. For all $s \in \mathbb{N}$, let $\mathbf{g}_s = (g_{s,1}(y), \ldots, g_{s,s}(y))$ be an hsg of $\mathcal{P}_s$. Suppose we can efficiently compute only the first "few" polynomials of the hsg. Can we *bootstrap* it, i.e. recover the whole hsg efficiently? Formally, we can describe this as follows. For any $m \in [s-1]$, the partial hsg $\mathbf{g}_{s,m}$ is defined as $(g_{s,1}, \ldots, g_{s,m})$. The partial hsg $\mathbf{g}_{s,m}$ can be seen as the hsg of those polynomials in $\mathcal{P}_s$ which depend only on the first m variables. Suppose that for $m \ll s$, we can compute $\mathbf{g}_{s,m}$ in poly(s)-time. Then, using this partial hsg, can we also design a complete hsg for $\mathcal{P}_s$ in poly(s)-time?

If $m = s^{1/c}$ for some $c \in \mathbb{N}$, then the answer is 'Yes' and it follows from the definition. The set $\mathcal{P}_s$ can be thought of as a subset of those polynomials in $\mathcal{P}_{s^c}$ which depend only on the first s variables. So $\mathbf{g}_{s^c,s} = (g_{s^c,1}, \ldots, g_{s^c,s})$ is a hsg for $\mathcal{P}_s$. Clearly, $\mathbf{g}_{s^c,s}$ can be computed in poly(s)-time. However, for $m \le s^{o(1)}$, we cannot use the same argument for the following reason. To compute the hsg of $\mathcal{P}_s$, we have to compute the partial hsg for $\mathcal{P}_{s^{\omega(1)}}$, which may not be computable in poly(s)-time. Naively speaking, there is no reason why a partial hsg $\mathbf{g}_{s,s^{o(1)}}$ could be bootstrapped efficiently to $\mathbf{g}_s$. The former is a property of the polynomial ring $\mathbb{F}[x_1, \ldots, x_{s^{o(1)}}]$ compared to that of the latter "much larger" polynomial ring $\mathbb{F}[x_1, \ldots, x_s]$; so in the underlying algebraic-geometry concepts a terrible blow up is warranted.

For any $c \in \mathbb{N}$, let $\log^{\circ c}$ be defined as c-times application of the base-2 logarithm function (eg. $\log^{\circ 3} s = \log\log\log s$). Somewhat surprisingly, we give a positive answer for m as small as $\log^{\circ c} s$, for any $c \in \mathbb{N}$. For smaller values of m (eg. $m = \log^\star s$), we leave it as an open question.

Theorem 1 (Bootstrap hsg)**.** *Suppose, for some $c \in \mathbb{N}$, we have a poly(s)-time blackbox PIT for size-s degree-s circuits that depend only on the first $\lceil \log^{\circ c} s \rceil$ variables. Then, we have a poly(sd)-time blackbox PIT for size-s degree-d circuits and Conjecture 1 holds.*

Remark– 1) In the boolean world, there is no extender that can stretch $0.99 \log s$ bits and "fool" size-s circuits. Because boolean functions on that many bits have circuit-size $< s$.

2) We also study the case when our partial hsg can be computed in subexponential time, which is far worse than polynomial time. In this case, our result is not as strong as Theorem 1. However, in the hypothesis we still deal with an $m = s^{o(1)}$ and manage to bootstrap that partial hsg in subexponential time. Also, an E-computable super-polynomially hard polynomial family is implied (say, *weak Conjecture 1*). For details see Theorem 13.

The bootstrapping idea brings forth pleasant surprises if we are willing to content ourselves with a "slightly super"-polynomial time blackbox PIT in the conclusion. Though we do not get an equivalence result now, we do however weaken the hypothesis very significantly.

Theorem 2. *Suppose, for constants $e \geq 2$ and $1 > \epsilon \geq (3 + 6\log(128e^2))/(128e^2)$, we have an $O(s^e)$-time blackbox PIT for degree-s polynomials computed by size-s circuits that depend only on the first $n := \lceil \max\{192e^2 \log(128e^2)^{1/\epsilon}, (64e^2)^{1/\epsilon}\} \rceil$ variables. Then, we have an $s^{\exp \circ \exp(O(\log^\star s))}$-time blackbox PIT for size-s degree-s circuits and Conjecture 1 holds.*

Remark: If we fix $e = 2$ and $\epsilon = 6912/6913$, then the hypothesis required is: Quadratic-time (i.e. $O(s^2)$) blackbox PIT for 6913-variate degree-s size-s polynomials.

In the above theorem, the exponent e in the complexity of PIT is a constant just below $\sqrt{n}/8$, where n is the (constant) number of variables. This can be achieved from a "poor" quality blackbox PIT algorithm (varying both s and n as independent parameters):

Theorem 3. *Suppose, for constant $\delta < 1/2$, we have an s^{n^δ}-time blackbox PIT for size-s degree-s circuits that depend only on the first n variables. Then, we have an $s^{\exp \circ \exp(O(\log^\star s))}$-time blackbox PIT for size-s degree-s circuits and Conjecture 1 holds.*

Note that in an n-variate degree-s polynomial, there are at most $1 + s^n$ monomials. So, the above hypothesis is unexpectedly weak. Additionally, the lower bound result that it will give is truly exponential. Next, we show that bootstrapping can be done even at shallow depths.

Theorem 4 (Depth-4 tiny variables)**.** *Suppose, for constant $n \geq 3$, we have a $\big(poly(s^n), O\big(\frac{s^{n/2}}{\log^2 s}\big)\big)$-hsg for size-$s$ depth-4 circuits that depend only on the first n variables. Then, we have a quasipoly(sd)-time blackbox PIT for size-s, degree-d circuits and Conjecture 1 holds.*

Remarks– 1) If we fix $n = 3$, then the hypothesis required is: $\big(\text{poly}(s), O(s^{1.5}/\log^2 s)\big)$-hsg for trivariate size-s depth-4 circuits. While $\big(\tilde{O}(s^3), (s+1)^3\big)$-hsg is trivial to design.

2) Depth-4 circuit is denoted as $\Sigma\Pi\Sigma\Pi$ to specify the alternating layers starting with the top addition gate. In older works it had been fruitful to restrict one of the product layer to mere *powering* gates [Sax08, GKKS13]. Indeed, we can prove stronger versions of Theorem 4: for $\Sigma \wedge \Sigma\Pi$ (Theorem 21) resp. $\Sigma\Pi\Sigma\wedge$ (Theorem 23) circuits in the hypothesis. But, these results (unlike Theorems 1–4) require the field characteristic to be zero or large.

3) Our conclusion is as strong as those obtained via the well-known 'constant-depth reduction' results in [AV08, GKKS13]. But our hypothesis needs an hsg only slightly better than the trivial; this cannot be done, not even guessed, using the old methods.

Finally, we want to change the viewpoint and see blackbox PIT for depth-3 circuits through the lens of *fixed parameter tractability* (fpt). This is discussed in Section 5.1. Bootstrapping of variables from *log*-variate width-2 ABP is done in Section 5.2.

1.2 Proof idea and our techniques

Proof idea of Theorem 1: We have to prove two results; one related to PIT and the other one related to lower bound. The latter will follow from Theorem 0, so we only describe the proof idea of PIT part. Suppose that for all $s, d, i \in \mathbb{N}$, $\mathcal{P}_{s,d,i}$ is the set of degree-d polynomials computed by size-s circuits that depend only on the first $f_i(sd)$ variables, where $f_i(s)$ is intended to be $\omega(\log^{\circ i} s)$. For all $0 \le i \le c+1$, $f_i(s) := (\log^{\circ i} s)^2$. Using reverse induction, we show that for $0 \le i \le c+1$, we have a poly(sd) time hsg for $\mathcal{P}_{s,d,i}$. First, we design a poly(sd) time hsg for $\mathcal{P}_{s,d,c+1}$ using the hypothesis mentioned in the theorem. Next, for all $i \in [c+1]$, we use the poly$(s'd')$ time hsg of $\mathcal{P}_{s',d',i}$ to design a poly(sd) time hitting-set of $\mathcal{P}_{s,d,i-1}$.

Our induction step can be broken into three smaller steps.

1) Hsg of $\mathcal{P}_{s',d',i}$ to hard polynomial family: For all $s \in \mathbb{N}$, let $\mathcal{T}_{s,i}$ be the s-degree polynomials computed by size-s circuit that depends only on the first $2c_1\lceil \log^{\circ i} s \rceil$ variables, where c_1 is some constant. Using poly$(s'd')$ hsg of $\mathcal{P}_{s',d',i}$, we can design a poly(s) time hsg for $\mathcal{T}_{s,i}$. Applying Lemma 5, we consider an annihilator, of the hsg, and get a family of hard polynomials which satisfies the properties mentioned in Lemma 12 (that we need in the next step).

2) Hard polynomial to variable reduction map: Lemma 12 designs an efficient variable reduction map using a hard polynomial family with certain properties. Thus, we perform a variable reduction on the polynomials in $\mathcal{P}_{s,d,i-1}$; significantly reducing variables from f_{i-1} to f_i.

3) The map to poly(sd) time hsg for $\mathcal{P}_{s,d,i-1}$: The above variable reduction converts every nonzero polynomial in $\mathcal{P}_{s,d,i-1}$ to a nonzero one in $\mathcal{P}_{s',d',i}$, where $s', d' = \text{poly}(sd)$. Thus, on applying the polynomial time hsg for $\mathcal{P}_{s',d',i}$, we get a polynomial time hsg for $\mathcal{P}_{s,d,i-1}$.

The crucial technical step is provided by Lemma 12, which is a strict generalization of Lemma 9. As mentioned earlier, the latter itself is a revised version of [KI04, Thm.7.7] as it can handle hard *non*-multilinear polynomials. It designs an efficient variable reduction using an exponentially hard but E-computable polynomial family. If we have a poly(s) time hsg for $\mathcal{T}_{s,1}$, then using Lemma 5, one can get such a polynomial family (as in the proof of Theorem 0 but now the hard polynomial will be non-multilinear). In Step 1 above, we are working with poly(s) time hsg for $\mathcal{T}_{s,i}$, where $i > 1$. In such an extremely low variate regime, Lemma 5 cannot give us a polynomial family with constant individual degree. So, we cannot use Lemma 9 ideas if we desire polynomial time computation.

There are several technical challenges faced in choosing parameters that should lead to a contradiction in the proof. Since the individual degree of the hard polynomial depends on the time-complexity s^e of the hsg of $\mathcal{T}_{s,i}$, the factor circuits will have a blown up size after using Kaltofen factoring. Care is needed to counterbalance the size of the Nisan-Wigderson (NW) design and the hardness of the polynomial family with the circuit complexity of the factors. The more sophisticated statement of Lemma 12 takes care of all those issues. Why is this lemma invoked multiple times? The answer lies in the way Kaltofen factoring yields a contradiction: using the fact that the third-parameter (i.e. set-intersection size) in the NW design is much smaller than the second-parameter (i.e. set size). This gives a smaller factor of the composite circuit after fixing certain variables. So, we need to apply NW design for *each* exponential stretch of variables; we do not know how to directly get a *hyper*-exponential stretch and save on time.

Proof idea of Theorem 2: Theorem 1 assumes an s^e-time hsg, where e is a constant, for $\log^{\circ c} s$-variate degree-s size-s polynomials. On the other hand, Theorem 2 assumes an s^e-time hsg for n-variate degree-s size-s polynomials, where $n := \lceil \max\{192e^2 \log(128e^2)^{1/\epsilon}, (64e^2)^{1/\epsilon}\} \rceil$ and $1 > \epsilon \ge (3 + 6\log(128e^2))/(128e^2)$ are constants. In both the cases, our hypotheses demand improved hsgs over the trivial ones (namely, $s^{\log^{\circ c} s}$ and s^n time respectively). This is the common strength of both the hypotheses which is exploited in the proofs.

Broadly, the proof of Theorem 2 is similar to the previous one. However, in Theorem 2 we desire, for a given e, to find the minimum number of *constant* variables for which we can reach the conclusion. This imposes more technical challenges and in many steps of the proof we have to work with much finer parameters. For example, our calculation suggests that for $e = 2$, the number of variables that we need is $n = 6913$ (or, for $e = 3$, $n = 17574$ suffices).

Like Theorem 1, in each inductive step, we stretch the number of variables exponentially. However, here we finally stretch n variables to s variables, where n is a constant. So, we need around $\log^\star s$ steps, which is non-constant wrt s. We show that if we have an s^{f_i}-time hsg, in the i-th induction step, then in the next step we get an $s^{f_{i+1}}$-time hsg, where $f_{i+1} := 16f_i^2$. So, after $\log^\star s$ steps, we get an hsg of our desired complexity (=slightly super-polynomial).

Like Lemma 12, here Lemma 19 combines all the crucial tools needed in the inductive step of Theorem 2. Our key ingredients here are again Nisan-Wigderson design and Kaltofen's factoring. However, we use them in a more optimized way. It will help us to improve the constants that underlie. This theorem and the next are very sensitive to these technicalities.

Thus, we show that a significant improvement of blackbox PIT within the polynomial-time domain itself (from s^n to s^e) would near-completely solve PIT for VP. This reminds us of other famous algebraic problems in computing where improvements in the time-exponent have been widely studied (& still open)— integer multiplication [Für09] and matrix multiplication [LG14].

Proof idea of Theorem 3: Suppose we have, for constant $\delta < 1/2$, an s^{n^δ}-time hsg for size-s degree-s circuits that depend only on the first n variables. Then, there exists an $\epsilon \in [2\delta, 1)$ and a large enough *constant* e such that: there is an s^e-time hsg for size-s degree-s circuits that depend only on the first $n := \lceil (64e^2)^{1/\epsilon} \rceil \geq 192e^2 \log(128e^2)^{1/\epsilon}$ variables. Note that $e \geq (n-1)^{\epsilon/2}/8 > n^\delta$ can be easily ensured, thus, s^e-time is more than s^{n^δ}-time. Now we simply invoke Theorem 2.

In fact, this proof needs the hypothesis only for: *infinitely* many n and large enough s.

Proof idea of Theorem 4: We argue using two intermediate models. For all $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of polynomials computed by size-s $\Sigma \wedge^a \Sigma\Pi$ circuits, $a(s)$ is an arbitrarily slow growing function, that depend only on the first n variables. Let $\mathcal{T}_s$ be the set of polynomials computed by size-s $\Sigma\Pi\Sigma\wedge$ circuits that depend only on the first n variables.

To prove Theorem 4, first we show that $(\mathrm{poly}(s), O(s^{n/2}/\log^2 s))$-hsg for $\mathcal{P}_s$ resp. $\mathcal{T}_s$ gives an efficient variable reduction and Conjecture 1 (see Theorems 21 resp. 23). This variable reduction converts a d-degree *nonzero* polynomial computed by a size-s circuit to a $O(\log(sd))$-variate $\mathrm{poly}(sd)$-degree *nonzero* polynomial. For $O(\log(sd))$-variate and $\mathrm{poly}(sd)$-degree polynomials, we have a $(sd)^{O(\log(sd))}$ time hitting-set. This completes the proof of PIT part. Next, we give the proof sketch of the variable reduction part.

First, we discuss the variable reduction part assuming the $O(s^{n/2}/\log^2 s)$-degree hsg of $\mathcal{P}_s$. We do it via an intermediate *multilinear* model. For all $s \in \mathbb{N}$, let $\mathcal{P}'_s$ be the set of $\frac{n}{2}\log s$ degree multilinear polynomials computed by size-s $\Sigma \wedge^a \Sigma\Pi$ circuits, that depend only on the first $n \log s$ variables. Next we describe how to get a hard polynomial family from an $(\mathrm{poly}(s), O(s^n/\log^2 s))$-hsg of $\mathcal{P}'_s$.

Since the number of $\frac{n}{2}\log s$ degree multilinear monomials over $m := n\log s$ variables is $\binom{m}{m/2} \geq 2^m/\sqrt{2m} = s^n/\sqrt{2m} > O(s^n/\log^2 s)\cdot m$ (for large enough s), we get an m-variate and $(m/2)$-degree multilinear homogeneous polynomial (annihilator) $q_m \notin \mathcal{P}'_s$ and computable in $\mathrm{poly}(s)$ time. The linear algebra is similar to Lemma 5; only difference being that Lemma 5 does not ensure q_m multilinear. However, the parameters of $\mathcal{P}'_s$ ensure the latter. Since q_m is m-variate $(m/2)$-degree multilinear polynomial and is not in $\mathcal{P}'_s$, q_m is not computed by size-s $\Sigma \wedge^a \Sigma\Pi$ circuits. Using depth reduction of [AV08], one can also ensure that q_m has circuit complexity $> s \geq 2^{\Omega(m)}$. This in turn gives the variable reduction using Lemma 9.

Now we show that an efficient $O(s'^{n/2}/\log^2 s')$-degree hsg of $\mathcal{P}_{s'}$ gives an efficient $O(s^n/\log^2 s)$-degree hsg for $\mathcal{P}'_s$, where s and s' are polynomially related. In $\mathcal{P}'_s$, divide the $n\log s$ variables into n blocks with each block of length $\log s$. Now take fresh variables $y_1, \ldots, y_n$, one for each block, and apply Kronecker map ($x_{u(j)+i} \mapsto y_j^{2^i}$, $i \in [\log s]$) within the j-th block $\{x_{u(j)+i} | i \in [\log s]\}$. Since polynomials in $\mathcal{P}'_s$ are multilinear, the above map preserves nonzeroness. This converts a nonzero polynomial in $\mathcal{P}'_s$ to a nonzero polynomial in $\mathcal{P}_{s'}$, where $s' = O(s^2)$. Now use the $O(s'^{n/2}/\log^2 s')$-degree hsg of $\mathcal{P}_{s'}$ to get one for $\mathcal{P}'_s$. For details see the proof of Theorem 21.

Second, we discuss the variable reduction part assuming an efficient $O(s^{n/2}/\log^2 s)$-degree hsg of $\mathcal{T}_s$. Proof idea is similar to the previous one; only difference is in the intermediate model. Here we consider the following model: for all $s \in \mathbb{N}$, let $\mathcal{T}'_s$ be the set of *multilinear* polynomials computed by size-s $\Sigma\Pi\Sigma$ circuits that depend only on the first $n \log s$ variables. Again, we show that an efficient $O(s'^{n/2}/\log^2 s')$-degree hsg of $\mathcal{T}_{s'}$ gives an $O(s^n/\log^2 s)$-degree hsg for $\mathcal{T}'_s$, which in turn gives the variable reduction as above coupled with [GKKS13]. For details see Theorems 22 & 23.

2 Brushing-up relevant techniques

In this section we will revisit the techniques that have appeared in some form in [HS80, NW94, KI03, Agr05, AV08, GKKS13].

From a hitting-set generator $\mathbf{f}(y)$ of a set of polynomials $\mathcal{P}$, we get an explicit polynomial outside $\mathcal{P}$ simply by looking at an annihilating polynomial of $\mathbf{f}(y)$. Previously, this approach was discussed in [HS80, Theorem 4.5] and [Agr05, Theorem 51]. In the next lemma, we prove a revised version. Later, it will be used to get hard polynomial from hitting-set generator.

Lemma 5 (Hitting-set to hardness). *Let $\mathbf{f}(y) = (f_1(y), \ldots, f_n(y))$ be a (t, d)-hsg for a set of n-variate polynomial $\mathcal{P}$. Then, there exists an n-variate polynomial $g(\mathbf{x})$ that is not in $\mathcal{P}$, is computable in* $\text{poly}(tdn)$*-time, has individual degree less than $\delta := \lceil d^{3/n} \rceil$, and is homogeneous of degree $(\delta - 1)n/2$.* (See Appendix B)

Corollary 6 (E-computable). *In the proof of Lemma 5, if $td = 2^{O(n)}$ then the polynomial family $g_n := g$, indexed by the variables, is E-computable.* (See Appendix B)

Towards a converse of the above lemma, a crucial ingredient is the Nisan-Wigderson design [NW94]. To describe it simply, the design *stretches a seed* from ℓ to $m \geq 2^{\frac{d}{10}}$ as follows,

Definition 7. *Let $\ell > n > d$. A family of subsets $\mathcal{D} = \{I_1, \ldots, I_m\}$ on $[\ell]$ is called an* (ℓ, n, d)-design, *if $|I_i| = n$ and for all $i \neq j \in [m]$, $|I_i \cap I_j| \leq d$.*

Lemma 8 (Nisan-Wigderson design, Chap.16 [AB09]). *There exists an algorithm which takes (ℓ, n, d) and a base set S of size $\ell > 10n^2/d$ as input, and outputs an (ℓ, n, d)-design $\mathcal{D}$ having $\geq 2^{d/10}$ subsets, in time $2^{O(\ell)}$.* (Lemma 15 improves this.)

Our next lemma is a revised version of the counterpositive of [KI04, Lemma 7.6]. If we have an exponentially hard but E-computable polynomial family, then we can efficiently reduce the variables from n to $O(\log(sd))$, for n-variate d-degree polynomials computed by size-s circuits, preserving nonzeroness.

Lemma 9 (Hardness to variable reduction). *For some constant δ, let $\{q_m\}_{m \geq 1}$ be a multi-δ-ic polynomial family computable in $\delta^{O(m)}$ time, but it has no $\delta^{o(m)}$-size algebraic circuit.*

Then, for n-variate d-degree polynomials computed by size-s circuits we have a $\delta^{O(\log(sd))}$-time variable-reducing polynomial map, from n to $O(\log(sd))$, that preserves nonzeroness. Furthermore, after variable reduction, the degree of the new polynomial is poly(sd). (See Appendix B)

Next lemma shows that the existence of an exponentially hard but E-computable polynomial family has an interesting complexity consequence. It is based on Valiant's criterion.

Lemma 10 (Valiant class separation). *If we have an E-computable polynomial family $\{f_n\}_{n\geq 1}$ of algebraic circuit complexity $2^{\Omega(n)}$, then either E$\not\subseteq$#P/poly or VNP has polynomials of algebraic circuit complexity $2^{\Omega(n)}$.* (See Appendix B)

Next lemma converts a monomial into a sum of powers. It is called Fischer's trick in [GKKS13]. It requires char $\mathbb{F} = 0$ or large.

Lemma 11 (Fischer's trick [Fis94]). *Over a field $\mathbb{F}$ of char$(\mathbb{F}) = 0$ or $> r$, any expression of the form $g = \sum_{i\in[k]} \prod_{j\in[r]} g_{ij}$ with $\deg(g_{ij}) \leq \delta$, can be rewritten as $g = \sum_{i\in[k']} c_i g_i^r$ where $k' := k2^r$, $\deg(g_i) \leq \delta$ and $c_i \in \mathbb{F}$. In fact, each g_i is a linear combination of $\{g_{i'j} | j\}$ for some i'.*

3 Many-fold composition of NW design– Proof of Theorem 1

Lemma 9 gave an efficient variable reduction from an exponentially hard but E-computable polynomial family. However, while bootstrapping in Theorem 1, we work with a case where number of variables can be as low as $\log^{\circ c} s$ compared to s, size of the circuit. In this extremely low variate regime, we have to deal with hard polynomial family of *non*-constant individual degree. There are also technical challenges faced in choosing parameters that should lead to a contradiction in the proof. So, we cannot use Lemma 9 directly. In Lemma 12, we take care of those issues. Overall proof strategy will be again to use Nisan-Wigderson combinatorial design and Kaltofen's algebraic circuit factoring algorithm. This is done repeatedly in Theorem 1.

Lemma 12 (Tiny variable reduction). *Let $c_3 \geq 1$ be the exponent in Kaltofen's factoring algorithm [Bür13, Thm.2.21]. For a constant $e \geq 1$ define, $c_0 := \lceil 9\sqrt{e} + 3\rceil c_3$, $c_1 := \lceil 30e + 10\sqrt{e+1}\rceil c_3$ and $c_2 := 1 + c_1^2$. Let ε be a tiny function say $\varepsilon(s) := 2\lceil \log^{\circ k} s\rceil$ for $k \geq 1$. Suppose we have a family $\{q_{m,s} \mid s \in \mathbb{N},\ m = c_1\varepsilon(s)\}$ of multi-$\delta_{m,s}$-ic m-variate polynomials that can be computed in $s^{O(1)}$ time, but has no size-s algebraic circuit, where $\delta_{m,s} := \lceil s^{3e/m}\rceil$.*

Then, there is a poly(sd)-time variable reduction map, reducing $n \leq 2^{\varepsilon((sd)^{c_0})}$ to $c_2\varepsilon((sd)^{c_0})$ and preserving nonzeroness, for degree-d n-variate polynomials computed by size-s circuits. Furthermore, after variable reduction, the degree of the new polynomial will be poly(sd).

Proof. Let $s' := sd$. Let $\mathcal{P}$ be the set of degree-d polynomials computed by size-s circuits that depend only on the first n-variables. We intend to stretch $c_2\varepsilon(s'^{c_0})$ variables to n. Define $m' := c_1\varepsilon((sd)^{c_0})$. Note that $q := q_{m',s'^{c_0}}$ has no algebraic circuit of size s'^{c_0}. Its individual-degree is $\leq \delta := \lceil s'^{3ec_0/m'}\rceil = s'^{o(1)}$.

Let $\mathcal{D} = \{S_1, \ldots, S_n\}$ be a $(c_2\varepsilon(s'^{c_0}), m', 10\varepsilon(s'^{c_0}))$-design on the variable set $Z = \{z_1, \ldots, z_{c_2\varepsilon(s'^{c_0})}\}$. Constants $c_2 > c_1 > 10$ will ensure the existence of the design by Lemma 8. Our hitting-set generator for $\mathcal{P}$ is defined as: for all $i \in [n]$, $x_i \mapsto q(S_i) =: p_i$ with S_i as variables. Then, we show that for any nonzero polynomial $P(\mathbf{x}) \in \mathcal{P}$, $P(p_1, \ldots, p_n)$ is also nonzero.

For the sake of contradiction, assume that $P(p_1, \ldots, p_n)$ is zero. Since $P(\mathbf{x})$ is nonzero, we can find the smallest $j \in [n]$ such that $P(p_1, \ldots, p_{j-1}, x_j, \ldots, x_n) =: P_1$ is nonzero, but $P_1\big|_{x_j=p_j}$ is zero. Thus, $(x_j - p_j)$ divides P_1. Let $\mathbf{a}$ be a constant assignment on all the variables in P_1, except x_j and the variables S_j in p_j, with the property: P_1 at $\mathbf{a}$ is nonzero. Since P_1 is nonzero, we can find such an assignment [Sch80]. Now our new polynomial P_2 on the variables S_j and x_j is of the form $P_2(S_j, x_j) = P(p'_1, \ldots, p'_{j-1}, x_j, a_{j+1}, \ldots, a_n)$, where for each $i \in [j-1]$, p'_i is the polynomial on the variables $S_i \cap S_j$, and a_i's are field constants decided by our assignment $\mathbf{a}$. By the design, for each $i \in [j-1]$, $|S_i \cap S_j| \leq 10\varepsilon(s'^{c_0})$. Since p'_i are polynomials on variables $S_i \cap S_j$ of individual degree$\leq \delta$, each p'_i has a circuit (of trivial form $\Sigma\Pi$) of size at most $m'\delta \cdot \delta^{10\varepsilon(s'^{c_0})}$ $= m'\delta \cdot \delta^{10m'/c_1}$.

Thus, we have a circuit for P_2 of size at most $s_1 := s + nm'\delta \cdot \delta^{10m'/c_1}$, and degree of the computed polynomial is at most $d_1 := dm'\delta$. Since $(x_j - p_j)$ divides P_2, we can invoke Kaltofen's factorization algorithm [Kal89] (see [Bür13, Theorem 2.21] for the algebraic circuit complexity of factors) and get an algebraic circuit for p_j of size $(s_1 d_1)^{c_3}$

$$\leq (snm'\delta \cdot \delta^{10m'/c_1} \cdot dm'\delta)^{c_3} = \left(s'nm'^2\delta^{2+\frac{10m'}{c_1}}\right)^{c_3} < (s'^{2+o(1)} \cdot \delta^{10m'/c_1})^{c_3}$$

$< s'^{(3+\,30ec_0/c_1)c_3}$. This exponent $= \left(\frac{3}{\lceil(9\sqrt{e}+3)\rceil} + \frac{30e}{\lceil 30e+10\sqrt{e}+1\rceil}\right) c_0 \leq \left(\frac{1}{(3\sqrt{e}+1)} + \frac{3\sqrt{e}}{3\sqrt{e}+\sqrt{1+1/e}}\right) c_0$ $< c_0$. So, $p_j = q(S_j)$ has circuit of size smaller than s'^{c_0}, which contradicts the hardness of q. Thus, $C(p_1, \ldots, p_n)$ is nonzero.

The time for computing $(p_1, \ldots, p_n)$ depends on: **(1)** computing the design (i.e. poly$(2^{m'})$-time), and **(2)** computing q (i.e. poly(sd)-time). Thus, the variable reduction map is computable in $\delta^{O(m')} = \text{poly}(sd)$-time. After variable reduction, the degree of the new polynomial is $< nd \cdot \deg(q) = \text{poly}(sd)$. $\square$

Remark. In the case of a finite field $\mathbb{F} = \mathbb{F}_{r^t}$ of prime characteristic r, we have to be careful while invoking Kaltofen's factoring. As, the latter outputs a small circuit for $p_j^{r^{t'}}$ where $r^{t'}$ is the highest power dividing the multiplicity of $x_j - p_j$ in P_2. However, when we raise the output by $r^{t-t'}$ we get a circuit that is small and agrees with p_j on $\mathbb{F}$-points. This is used, like in [KI04, Rmk.7.5], to redefine algebraic complexity of q over $\mathbb{F}_{r^t}$ suitably and the above lemma works.

Proof of Theorem 1. Consider the following two statements. **S1:** we have a poly(s)-time hsg for size-s degree-s circuits that depend only on the first $\lceil \log^{\circ c} s\rceil$ variables, and **S2:** we have a poly(s)-time hsg for degree-s polynomials computed by size-s circuits that depend only on the first $\lceil \log^{\circ c} s\rceil$ variables. **S1** is our given hypothesis. However, in this proof, we work with **S2** which is stronger than **S1**, as in the former case circuits may have degree *larger* than s. So we first argue that they are equivalent up to polynomial overhead. **S2** trivially implies **S1**. For the other direction, we invoke (the proof of) the 'log-depth reduction' result for arithmetic circuits. For any size-s circuit C computing a degree-s polynomial, we have an s^{e_0}-size s-degree circuit C' computing the same polynomial, for some constant e_0 (see [Sap16, Thm.5.15]). Now apply **S1** for s^{e_0}-size s-degree and get poly(s)-hsg for C. Next, we focus on designing poly(sd)-hsg for degree-d polynomials computed by size-s circuits, using our stronger hypothesis **S2**.

Suppose that for all $s, d, i \in \mathbb{N}$, $\mathcal{P}_{s,d,i}$ is the set of degree-d polynomials computed by size-s circuits that depend only on the first $f_i(sd)$ variables, where $f_i(s) := (\log^{\circ i} s)^2$. We prove that for all $0 \leq i \leq c+1$, we have a polynomial time hitting set for $\mathcal{P}_{s,d,i}$. We will use reverse induction on i. Define function $\varepsilon_i(s) := 2\lceil \log^{\circ i} s\rceil$.

Base case– Poly(sd)-hsg for $\mathcal{P}_{s,d,c+1}$: Let $t := \max\{s, d\}$. Then $\mathcal{P}_{s,d,c+1}$ is a subset of $\mathcal{P}_{t,t,c+1}$. For all $s \in \mathbb{N}$, let $\mathcal{T}_s$ be the set of degree-s polynomials computed by s-size circuits that depend only on the first $\lceil \log^{\circ c} s\rceil$ variables. Using the hypothesis **S2**, we have a poly(s) time hsg for $\mathcal{T}_s$. Since $f_{c+1}(t) \leq \lceil \log^{\circ c} t\rceil$ for large t, $\mathcal{P}_{t,t,c+1}$ is a subset of $\mathcal{T}_t$. So $\mathcal{P}_{t,t,c+1}$ also has a poly(t)-time hsg. This gives a poly(sd)-time hsg for $\mathcal{P}_{s,d,c+1}$.

Induction step– From poly$(s'd')$-hsg of $\mathcal{P}_{s',d',i}$ to poly(sd)-hsg of $\mathcal{P}_{s,d,i-1}$: We divide this step into three smaller steps, for $i \in [c+1]$.

1) Hsg of $\mathcal{P}_{s',d',i}$ to hard polynomial family: For some constant e, we have $((s'd')^{e/2}, (s'd')^{e/2})$-hsg for $\mathcal{P}_{s',d',i}$. Let for all $s, i \in \mathbb{N}$, $\mathcal{T}_{s,i}$ be the set of degree-s polynomials computed by size-s circuits that depend only on the first $c_1\varepsilon_i(s)$ variables, where c_1 is a constant as defined in Lemma 12 using the e. Note that $m := c_1\varepsilon_i(s)$ is smaller than $f_i(s^2)$ for large enough s. So, polynomial time hsg for $\mathcal{P}_{s',d',i}$ gives a (s^e, s^e)-hsg for $\mathcal{T}_{s,i}$. Then using Lemma 5, we get an m-variate polynomial $q_{m,s}$ such that **1)** individual degree is less than $\delta_{m,s} = \lceil s^{3e/m} \rceil$, **2)** $q_{m,s} \notin \mathcal{T}_{s,i}$, and **3)** computable in $s^{O(1)}$-time.

Suppose $q_{m,s}$ has a circuit C of size less than s. Since the degree $(m \cdot \delta_{m,s})$ is also less than s, the polynomial $q_{m,s}$ is in $\mathcal{T}_{s,i}$, which is a contradiction. So using (s^e, s^e)-hsg for $\mathcal{T}_{s,i}$, for all $s \in \mathbb{N}$, we get a polynomial family $\{q_{m,s} \mid \ s \in \mathbb{N},\ m = c_1 \varepsilon_i(s)\}$ of multi-$\delta_{m,s}$-ic that can be computed in $s^{O(1)}$ time, but has no size-s algebraic circuit.

2) Hard polynomial to variable reduction map: Note that $f_{i-1}(sd) \leq 2^{\varepsilon_i((sd)^{c_0})}$, where c_0 is a constant defined in Lemma 12 using the e. Using the lemma (for $\varepsilon = \varepsilon_i$), any nonzero polynomial $P \in \mathcal{P}_{s,d,i-1}$ can be converted, in poly(sd)-time, to another poly(sd)-degree nonzero polynomial P' computed by poly(sd)-size circuit which depends only on the first $c_2 \varepsilon_i((sd)^{c_0})$ variables.

3) The map to poly(sd) time hsg for $\mathcal{P}_{s,d,i-1}$: Since, in P', the number of variables $c_2 \varepsilon_i((sd)^{c_0})$ is less than $f_i(sd)$, using poly-time hsg of $\mathcal{P}_{s',d',i}$ we get a poly-time hsg for $P \in \mathcal{P}_{s,d,i-1}$.

Repetition– After applying the induction step $c+1$ times, we have a poly(sd)-time hsg for $\mathcal{P}_{s,d,0}$. In other words, we have a poly(sd)-time hsg for size-s degree-d circuits.

Now we show that Conjecture 1 holds. We just obtained a poly(s)-time hsg for $\mathcal{T}_{s,1}$. Let $m = \lceil \log s \rceil$. Then applying Lemma 5, we get a family of polynomials $\{q_m\}_{m \geq 1}$ such that **1)** it is multi-δ-ic, for some constant δ, and **2)** computable in $\delta^{O(m)}$-time, but has no $\delta^{o(m)}$-size algebraic circuit. Now, applying Lemma 10, we get Conjecture 1. □

Remark. In the case of a finite field $\mathbb{F} = \mathbb{F}_{r^t}$ of prime characteristic r, we have to redefine the hardness of the polynomial $q_{m,s}$ in Step (1) of the induction step above. As remarked before, we can define $\mathcal{T}_{s,i}$ to be the set of polynomials $f(x_1, \ldots, x_{c_1 \varepsilon_i(s)})$, such that for some e, f^{r^e} agrees *on all* $\mathbb{F}$*-points* with some nonzero degree-s polynomial computed by a size-s circuit. It can be seen that an hsg for $\mathcal{T}_{s,i}$ gives a hard $q_{m,s}$ (via the annihilator approach of Lemma 5) that can be used in Step (2).

Next, we relax the hypothesis of Theorem 1 by allowing a subexponential time hitting-set. In the following discussion, we use a constant e_0. It is the exponent of 'log-depth reduction' algorithm ([Sap16, Thm.5.15]), i.e. for every size-s circuit computing a degree-d n-variate polynomial, we also have an $(sdn)^{e_0}$-size d-degree circuit computing the same polynomial. We recall the following standard definition.

subexp: A function $f(s)$ is in *subexp* if $f(s) = \exp(s^{o(1)})$. Eg. $2^{\sqrt{s}} \notin$ subexp, but $\exp(2^{\sqrt{\log s}}) \in$ subexp. One can recall the standard complexity class, SUBEXP $:= \cap_{\epsilon > 0}$DTIME$(\exp(n^{\epsilon}))$. Basically, these are decision problems whose time-complexity is a subexp function.

Theorem 13 (Subexp bootstrap)**.** *Let f be a function in subexp. Suppose that we have a poly($f(s)$) time blackbox PIT for size-s degree-s circuits that depend only on the first* $10\lceil \log f(s) \rceil$ *variables. Then, we have blackbox PIT for size-s degree-d circuits in subexponential,* $\exp((sd)^{o(1)})$, *time. Furthermore, either E $\not\subseteq$ #P/poly or VNP$\neq$VP.*

Remark. As an fpt-algorithm the hypothesis requires a blackbox PIT, for size-s degree-s n-variate circuits, of time complexity potentially as large as $\exp(s^{o(1)} + O(n))$.

Proof. Our proof is divided into three parts. First, we show how to construct a hard polynomial family using subexponential time partial hsg. Next, we show a nontrivial variable reduction for circuits using the hard polynomial family. Finally, we apply the hsg due to trivial derandomization of low-variate PIT [Sch80] and get a subexponential hsg. For lower bound part, we show that our hard polynomial family satisfies the required conditions.

Define the function $\varepsilon(s) := 10\lceil \log f(s) \rceil$. For all $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of polynomials computed by size-s degree-s circuits that depend only on the first $\varepsilon(s)$ variables. We have, for some constant e, an $(f(s)^e, f(s)^e)$-hsg for $\mathcal{P}_s$ from the hypothesis. Using Lemma 5, we get an

m-variate polynomial $q_{m,s}$, where $m := \varepsilon(s)$, such that: **1)** it is multi-δ-ic, for some constant δ, **2)** computable in poly$(f(s))$ time, and **3)** $q_{m,s} \notin \mathcal{P}_s$. Now we prove that $q_{m,s}$ is not computable by circuits of size less than $s^{1/2e_0}$.

For the sake of contradiction assume that $q_{m,s}$ has a circuit of size $s_1 < s^{1/2e_0}$. Since $f \in$ subexp, the number of variables $m = \varepsilon(s) = s^{o(1)}$. Similarly, the degree bound $d := m\delta$ of $q_{m,s}$ is also $s^{o(1)}$. Now applying 'log-depth reduction' algorithm (see [Sap16, Theorem 5.15]), we get a d-degree circuit C of size $(s_1 dm)^{e_0}$ for $q_{m,s}$. Since $m, d = s^{o(1)}$, the size of C is $< s$. This implies that $q_{m,s} \in \mathcal{P}_s$, which is a contradiction. So $q_{m,s}$ is not computed by circuits of size less than $s^{1/2e_0}$. This gives us a family of hard polynomial $\mathcal{F} := \{q_{m,s} \mid s \in \mathbb{N}, m = \epsilon(s)\}$ such that it is: **1)** m-variate and multi-δ-ic for some constant δ, and **2)** computable in poly$(f(s))$ time but no circuits of size less than $s^{1/2e_0}$ can compute it.

In the following claim, we describe how to reduce variables nontrivially using $\mathcal{F}$'s hardness.

Claim 14 (Subexp var.reduction)**.** *Using $\mathcal{F}$, for some constant c, we have an exp$(\varepsilon((sd)^c)^2/\log s)$-time computable variable reduction map, from n to $\lceil \varepsilon((sd)^c)^2/\log s \rceil$, that preserves nonzeroness for degree-d n-variate polynomials computed by size-s circuits. Furthermore, after variable reduction, the degree of the new polynomial will be poly(sd).* (See Appendix C)

Define $\varepsilon' = \varepsilon'(s,d) := \lceil \varepsilon((sd)^c)^2/\log s \rceil$. Using the above claim, any degree-d nonzero polynomial P computed by a size-s circuit can be converted to a ε'-variate nonzero polynomial P' of degree $(sd)^{O(1)}$. P' has $(sd)^{O(\varepsilon')}$ time hsg. Total time taken (variable reduction + hsg complexity) is $\exp(O(\varepsilon')) + \exp(O(\varepsilon')\log(sd))$. Since $f \in$ subexp, $\varepsilon' = (sd)^{o(1)}$. So the total time is also in subexp. In terms of f, our time complexity is $\exp(\log(sd) \cdot \log^2 f(s')/\log s)$, where $s' := (sd)^c$.

Now we discuss the hardness of $\{q_{m,s} \mid s \in \mathbb{N},\ m = \varepsilon(s)\}$ wrt m, the number of variables of $q_{m,s}$. We know that $q_{m,s}$ requires circuit size $\geq s^{1/2e_0}$. Since $m = \varepsilon(s) = s^{o(1)}$, the circuit size is $m^{\omega(1)}$. On the other hand, $q_{m,s}$ is poly$(f(s)) = 2^{O(m)}$ time computable and is a multi-δ-ic polynomial, for some constant δ. So $\mathcal{F}$ is an E-computable polynomial family and like Lemma 10, we get our lower bound result. The lower bound for $q_{m,s}$ directly in terms of f can also be calculated: Since $m = \varepsilon(s) = 10\lceil \log f(s) \rceil$ and f is an increasing function, so $s = f^{-1}(2^{\Omega(m)})$. This implies that $q_{m,s}$ requires circuit size $(f^{-1}(2^{\Omega(m)}))^{\Omega(1)}$. $\square$

4 Bootstrap constant-variate PIT– Proof of Theorems 2 & 3

The overall strategy is similar to the last section. However, the details would now change drastically. Some of the technical proofs of this section have been moved to Appendix D.

First, we describe an optimized version of the Nisan-Wigderson (NW) design, where the parameters are different from that in Lemma 8. Later, it will help us improve the constants.

Lemma 15 (NW design)**.** *There exists an algorithm which takes (ℓ, n, d), with $\ell \geq 100$ and $d \geq 13$, and a base set S of size $\ell := \lceil 4n^2/d \rceil$ as input, and outputs an (ℓ, n, d)-design $\mathcal{D}$ having $\geq 2^{d/4}$ subsets, in time $O((4\ell/n)^n)$.*

Exponent vs variables. In this section, to describe the complexity parameters of the circuits and the hsg, we use two families of functions $\{f_i\}_{i\geq 0}$ ("exponent of time") and $\{m_i\}_{i\geq 0}$ ("number of variables"). They are defined as follows: $f_0 \geq 2$ and $m_0 \geq 1024$ are constants and for all $i \geq 1$,

$$f_i := 16 f_{i-1}^2 \quad \text{and} \quad m_i := 2^{m_{i-1}/(64 f_{i-1}^2)}\,.$$

Our strategy is to use an NW $(m_i, \frac{m_i}{8f_i}, \frac{m_i}{16 f_i^2})$-design to stretch m_i variables to m_{i+1}. We want to show that m_i grows much faster in contrast to f_i. In particular, we need m_i to be a *tetration* in i (i.e. iterated exponentiation), while f_i is "merely" a double-exponentiation in i. Seeing this needs some effort and we will do this in the next two propositions.

From now on we will assume that ϵ is a constant fraction satisfying $1 > \epsilon \geq (3 + 6\log(128f_i^2))/(128f_i^2)$, for $i = 0$. Since f_i increases with i, the fraction $(3+6\log(128f_i^2))/(128f_i^2)$ decreases. Thus, the constant ϵ remains larger than the latter, for all $i \geq 0$.

Proposition 16. *If, for some $i \geq 0$, $m_i \geq 192f_i^2 \cdot \frac{1}{\epsilon}\log(128f_i^2)$, then the same relation holds between m_{i+1} and f_{i+1}.*

Proposition 17 (m_i is a tetration)**.** *Suppose that $m_0 \geq \max\{(8f_0)^{\frac{2}{\epsilon}},\ 192f_0^2 \cdot \frac{1}{\epsilon}\log(128f_0^2)\}$. Then for all $i \geq 0$: 1) $m_{i+1} \geq 2^{m_i^{1-\epsilon}}$ and 2) $m_{i+1} \geq 2m_i > 3456f_i^2$.*

Once we know that m_i grows extremely rapidly, we want to estimate the number of iterations before which it reaches s.

Proposition 18 (Iteration count)**.** *The least i, for which $m_i \geq s$, is $\leq \frac{3}{1-\epsilon}\log\left(\frac{3}{1-\epsilon}\right) + 2\log^\star s$.*

Now we describe the i-th step of bootstrapping.

Lemma 19 (Induction step)**.** *Let s be the input size parameter, $i \geq 0$, $m_i = s^{o(1)}$ and $m' := \min\{m_{i+1}, s\}$. Suppose that we have an s^{f_i}-time hsg for m_i-variate degree-s polynomials computed by size-s circuits. Then, we have an $s^{f_{i+1}}$-time hsg for m'-variate degree-s polynomials computed by size-s circuits.*

Proof. Although i might grow (extremely) slowly wrt s, it helps to think of i and s as two independent parameters. Suppose that for all $s \in \mathbb{N}$, $\mathcal{P}_{s,i}$ is the set of m_i-variate degree-s polynomials computed by size-s circuits, and $\mathcal{P}_{s,i+1}$ is the set of m'-variate degree-s polynomials computed by size-s circuits. Our proof can be broken into three main *steps*. First, using the hsg of $\mathcal{P}_{s,i}$ we construct a hard polynomial family. Next, using that hard polynomial family we do variable reduction on the polynomials in $\mathcal{P}_{s,i+1}$. This variable reduction is relatively "low"-cost and it reduces a *nonzero* polynomial in $\mathcal{P}_{s,i+1}$ to some *nonzero* polynomial in $\mathcal{P}_{s^{9f_i},i}$, for sufficiently large value of s. Finally, we apply the hsg of $\mathcal{P}_{s^{9f_i},i}$ to get the desired hsg for $\mathcal{P}_{s,i+1}$. The challenge is to analyze this; which we do now in detail. Keep in mind the properties of the functions m_i, f_i that we proved before.

Hard polynomial family construction: We describe the construction of a hard polynomial family from the hsg of $\mathcal{P}_{s,i}$. Let $d_i(s) := s^{f_i}$ and for all $s \in \mathbb{N}$, let $\mathcal{T}_s$ be the set of $\frac{m_i}{8f_i}$-variate degree-s polynomials computed by size-s circuits. The $d_i(s)$-time hsg of $\mathcal{P}_{s,i}$ also gives an hsg for $\mathcal{T}_s$ with same time complexity. Like Lemma 5, the annihilator of the hsg of $\mathcal{T}_s$ gives a polynomial q_s such that: 1) $q_s \notin \mathcal{T}_s$, 2) it is computable in d_i^4-time by linear algebra, and 3) it is multi-δ_s-ic, where $\delta_s := 1 + d_i(s)^{\frac{8f_i+1}{m_i}} = 1 + s^{f_i(8f_i+1)/m_i}$. Here, the main difference is that the individual degree bound δ_s is smaller than what Lemma 5 ensures. It will help us reduce the initial constants in our calculations. We give a brief sketch of how we get an annihilator with this individual degree.

The number of monomials on $\frac{m_i}{8f_i}$ variables with individual degree $< \delta_s$ is at least $m := d_i^{1+\frac{1}{8f_i}} = s^{f_i+\frac{1}{8}}$. After evaluating an $\frac{m_i}{8f_i}$-variate multi-δ_s-ic polynomial on the hsg of $\mathcal{T}_s$, we get a univariate polynomial of degree at most $d := \frac{m_i}{8f_i} \cdot \left(d_i + d_i^{1+\frac{8f_i+1}{m_i}}\right) \leq \frac{m_i}{8f_i} \cdot 2s^{f_i+\frac{8f_i^2+f_i}{m_i}}$. To make the linear algebra argument of Lemma 5 work, we need $m > d$. This holds as $m_i = s^{o(1)}$ and as by Proposition 17 we have $m_i \geq 1728f_i^2$.

Now we argue that q_s has no circuit of size$\leq s$. For the sake of contradiction, assume that q_s has a circuit of size$\leq s$. The degree of q_s is at most $\frac{m_i}{8f_i} \cdot 2d_i^{\frac{8f_i+1}{m_i}} \leq \frac{m_i}{8f_i} \cdot 2s^{\frac{f_i(8f_i+1)}{m_i}}$. Applying $m_i = s^{o(1)}$ and $m_i \geq 1728 f_i^2$, we get that q_s has degree $< s$. This implies that $q_s \in \mathcal{T}_s$, which is a contradiction. Thus, q_s has no circuit of size$\leq s$. So we have a multi-δ_s-ic polynomial family $\{q_s \mid s \in \mathbb{N}\}$ such that, 1) q_s is computable in $d_i^4 = s^{4f_i}$ time but has no circuit of size$\leq s$, 2) it has individual degree $\delta_s = 1 + s^{f_i(8f_i+1)/m_i}$ and number of variables $\frac{m_i}{8f_i}$.

Variable reduction map: Now we convert every non-zero polynomial in $\mathcal{P}_{s,i+1}$ to a non-zero polynomial in $\mathcal{P}_{s^{12f_i},i}$. Consider a slightly larger size parameter $s_0 := s^7$. Let $\{S_1, \ldots, S_{m'}\}$ be an NW $(m_i, \frac{m_i}{8f_i}, \frac{m_i}{16f_i^2})$-design on the variable set $\{z_1, \ldots, z_{m_i}\}$. The growth properties of m_i, togetherwith Lemma 15, ensures that such a design exists. Define for all $j \in [m'], p_j := q_{s_0}(S_j)$. Next, we show that for any non-zero $P \in \mathcal{P}_{s,i+1}$, $P(p_1, \ldots, p_{m'})$ is also non-zero.

For the sake of contradiction, assume that $P(p_1, \ldots, p_{m'})$ is zero. Since $P(\mathbf{x})$ is nonzero, we can find the smallest $j \in [m']$ such that $P(p_1, \ldots, p_{j-1}, x_j, \ldots, x_{m'}) =: P_1$ is nonzero, but $P_1\big|_{x_j=p_j}$ is zero. Thus, $(x_j - p_j)$ divides P_1. Let $\mathbf{a}$ be a constant assignment on all the variables in P_1, except x_j and the variables S_j in p_j, with the property: P_1 at $\mathbf{a}$ is nonzero. Since P_1 is nonzero, we can find such an assignment [Sch80]. Now our new polynomial P_2, on the variables S_j and x_j, is of the form $P_2(S_j, x_j) := P(p'_1, \ldots, p'_{j-1}, x_j, a_{j+1}, \ldots, a_{m'})$, where for each $i \in [j-1]$, p'_i is the polynomial on the variables $S_i \cap S_j$, and a_i's are field constants decided by our assignment $\mathbf{a}$. By the design, for each $i \in [j-1]$, $|S_i \cap S_j| \leq \frac{m_i}{16f_i^2}$. Since p_is are polynomials on variables S_i of individual degree$\leq \delta_{s_0}$, each p'_i has a circuit (of trivial form $\Sigma\Pi$) of size at most

$$\frac{m_i}{16f_i^2}\delta_{s_0} \cdot \delta_{s_0}^{\frac{m_i}{16f_i^2}}.$$

Thus, we have a circuit for P_2 of size at most s_1 and the degree of P_2 is at most d_1, where

$$s_1 := s + \frac{m' m_i \delta_{s_0}}{16 f_i^2} \cdot \delta_{s_0}^{\frac{m_i}{16f_i^2}} \quad \text{and} \quad d_1 := s \cdot \frac{m_i \delta_{s_0}}{16f_i^2}.$$

Since $(x_j - p_j)$ divides P_2, we can invoke Kaltofen's factorization algorithm [Kal89] (see [Bür13, Thm.2.21] for the improved complexity of factors) and get an algebraic circuit for p_j of size $s'_0 := s_1 \tilde{O}(d_1^2)$. Now we prove that $s'_0 < s_0$, for large enough s. This implies that q_{s_0} has a circuit of size$\leq s_0$ which contradicts the hardness of q_{s_0}.

Recall that $\delta_{s_0} = 1 + s_0^{f_i(8f_i+1)/m_i}$. Let us upper bound $s'_0 =$

$$\begin{aligned}
s_1\tilde{O}(d_1^2) &\leq \left(s + \frac{m' m_i}{16f_i^2} \cdot \delta_{s_0}^{1+\frac{m_i}{16f_i^2}}\right) \cdot \tilde{O}\left(\frac{s m_i \delta_{s_0}}{16 f_i^2}\right)^2 \\
&\leq \frac{s^{3+o(1)}\delta_{s_0}^2}{f_i^2} + \frac{s^{3+o(1)}\delta_{s_0}^{3+\frac{m_i}{16f_i^2}}}{f_i^4} \qquad \left(\because m_i = s^{o(1)}, m' \leq s\right) \\
&\leq s^{3+o(1)} s_0^{\frac{2f_i(8f_i+1)}{m_i}} + s^{3+o(1)} s_0^{\left(3+\frac{m_i}{16f_i^2}\right)\frac{f_i(8f_i+1)}{m_i}} \\
&\leq s^{3+o(1)} s^{\frac{14f_i(8f_i+1)}{m_i}} + s^{3+o(1)} s^{\left(3+\frac{m_i}{16f_i^2}\right)\frac{7f_i(8f_i+1)}{m_i}} \\
&\leq s^{3+o(1)} s^{\frac{112f_i^2+14f_i}{m_i}} + s^{3+o(1)} s^{\frac{21f_i(8f_i+1)}{m_i}+\frac{7(8f_i+1)}{16f_i}}
\end{aligned}$$

$$
\begin{aligned}
&\leq\ s^{3+o(1)} s^{\frac{112 f_i+14}{1728 f_i}} \ +\ s^{3+o(1)} s^{\frac{21(8f_i+1)}{1728 f_i}+\frac{7(8f_i+1)}{16 f_i}} \qquad (\because m_i > 1728 f_i^2) \\
&\leq\ s^{3+o(1)+\frac{112}{1728}+\frac{7}{1728}} \ +\ s^{3+o(1)+\frac{168}{1728}+\frac{21}{3456}+\frac{56}{16}+\frac{7}{32}} \qquad (\because f_i \geq 2) \\
&\leq\ s^{3.1+o(1)} \ +\ s^{6.83+o(1)} \\
&<\ s^7 = s_0\,.
\end{aligned}
$$

This gives a contradiction for sufficiently large s. So $P' := P(p_1,\ldots,p_{m'})$ is non-zero.

Using the given hsg: The above variable reduction converts P to a m_i-variate degree-d' non-zero polynomial P' computable by s'-size circuit, where

$$
d' := \frac{s m_i}{8 f_i} \cdot \delta_{s_0} \quad \text{and} \quad s' := s + \frac{m' m_i \delta_{s_0}}{8 f_i} \cdot \delta_{s_0}^{\frac{m_i}{8 f_i}}\,.
$$

Now we give an upper bound of s':

$$
\begin{aligned}
s' \ &=\ s\ +\ \frac{m' m_i \delta_{s_0}}{8 f_i} \cdot \delta_{s_0}^{\frac{m_i}{8 f_i}} \\
&=\ s\ +\ \frac{m' m_i}{8 f_i} \cdot \left(1 + s_0^{\frac{f_i(8f_i+1)}{m_i}}\right)^{(\frac{m_i}{8f_i}+1)} \\
&\leq\ s\ +\ \frac{m' m_i}{8 f_i} \cdot (1+s)^{\frac{7 f_i(8f_i+1)}{m_i}(\frac{m_i}{8f_i}+1)} \\
&\leq\ s\ +\ s^{1+o(1)+7(f_i+\frac{1}{8})(1+\frac{8f_i}{m_i})} \\
&<\ s\ +\ s^{1+o(1)+7(f_i+\frac{1}{8})(1+\frac{8}{3456})} \qquad (\because m_i > 1728 f_i^2, f_i \geq 2) \\
&<\ s^{9 f_i}\,.
\end{aligned}
$$

Since $d', s' < s^{9f_i} =: s_1$, P' is m_i-variate degree-s_1 polynomial that is computable by size-s_1 circuit. So P' has an hsg of time complexity $s_1^{f_i} = s^{9f_i^2}$.

Final time complexity: First, let us review our overall algorithm: It takes $(1^s, 1^{i+1})$ as input, and in $s^{f_{i+1}}$-time, outputs an $s^{9f_i^2}$-time hsg of $\mathcal{P}_{s,i+1}$, under the assumption that for all $t \geq s$, there is a t^{f_i}-time hsg for $\mathcal{P}_{t,i}$.

a. $s_0 \leftarrow s^7$.

b. By linear algebra, compute an annihilator q_{s_0} (in dense representation) of the given hsg of $\frac{m_i}{8f_i}$-variate degree-s_0 size-s_0 polynomials.

c. Compute NW design (by the greedy algorithm sketched in Lemma 15) $\{S_1,\ldots,S_{m'}\}$ on the variable set $\{z_1,\ldots,z_{m_i}\}$.

d. Compute an m_i-input and m'-output circuit C (in the form $\Sigma\Pi$) on the variables $\{z_1,\ldots,z_{m_i}\}$ such that: for all $j \in [m']$, the j-th output is $p_j = q_{s_0}(S_j)$.

e. Compute the hsg $\mathbf{a} = (a_1,\ldots,a_{m_i})$ of $\mathcal{P}_{s^{9f_i},i}$. Then, the above proof shows that an hsg for $\mathcal{P}_{s,i+1}$ is $C(\mathbf{a})$.

The total time complexity of hsg for P has four components:

1. Computing q_{s_0} (step b): It takes $(s_0^{f_i})^4 = s^{7\times f_i \times 4} = s^{28 f_i} \leq s^{14 f_i^2}$.

2. Nisan Wigderson design from Lemma 15 (step c): It takes time $O\left(4m_i/(m_i/8f_i)\right)^{m_i/8f_i} = O(32f_i)^{m_i/8f_i}$. If $m_i > 64f_i^2 \log s$ then we will run the i-th induction step only for (relabelled) $m_i := 64f_i^2 \log s$, as the stretch obtained will already be to $2^{m_i/64f_i^2} = s$ variables. Note that at that point, i would be non-constant and hence $f_i > 4$. In this regime, $(32f_i)^{m_i/8f_i} = (32f_i)^{8f_i \log s} = s^{8f_i \log(32f_i)} < s^{12f_i^2}$.

3. Computing C (step d): Essentially, compute m' copies of q_{s_0} (in dense representation). As seen before, the total time-complexity is s^{9f_i}.

4. Computing hsg of $\mathcal{P}_{s^{9f_i},i}$. Then, computing hsg of $\mathcal{P}_{s,i+1}$ by composition (step e): It takes $s^{9f_i^2} + s^{9f_i^2} \cdot s^{9f_i} < s^{14f_i^2}$ time.

So, the total time is smaller than $s^{16f_i^2} = s^{f_{i+1}}$ and we have an hsg for m'-variate P. $\square$

Proof of Theorem 2. In the hypothesis of the theorem statement we are given constants $e \geq 2$ and $n \geq 1024$. Let us define the m_i, f_i polynomial family with the initialization $f_0 := e$ and $m_0 := n$. The idea is to simply use the induction step (Lemma 19) several times to boost m_0 variables to an arbitrary amount.

Let P be a degree-s polynomial computed by size-s circuit. Then, it can have at most s variables. Let k be the smallest integer such that $m_k \geq s$ (k is an extremely slow growing function in s as described in Proposition 18). By Proposition 17, we have that $m_{k-1} \leq s^{o(1)}$.

For $i \in \mathbb{N}_{\geq 0}$ and large enough parameters $t > t' > s$, let $\mathcal{P}_{t,i}$ denote the set of m_i-variate degree-t polynomials computed by size-t circuits. From the hypothesis, we have a t^{f_0}-time hsg for $\mathcal{P}_{t,0}$. Now for each $i < k$, we apply Lemma 19, to get the $t'^{f_{i+1}}$-hsg for $\mathcal{P}_{t',i+1}$. After k such applications of Lemma 19, we get an s^{f_k}-time hsg for *s-variate* degree-s polynomials computed by size-s circuits.

Note that $f_k = (16f_0)^{2^k}/16 = 2^{O(2^k)} = 2^{2^{O(\log^\star s)}}$. Thus, we have an $s^{\exp\circ\exp(O(\log^\star s))}$-time blackbox PIT for VP circuits.

Since $f_0 < m_0/2$ one can see that the hypothesis of Theorem 4 is easily satisfied. This gives us an E-computable polynomial family $\{q_m\}_{m\geq 1}$ with hardness $2^{\Omega(m)}$. $\square$

Proof of Theorem 3. Suppose we have, for constant $\delta < 1/2$, an s^{n^δ}-time hsg for size-s degree-s circuits that depend only on the first n variables. Wlog (using depth-reduction proofs), we can assume that we have an s^{n^δ}-time hsg for degree-s polynomials computed by size-s circuits that depend only on the first n variables.

Then, there exists an $\epsilon \in [2\delta, 1)$ and a large enough *constant* e such that: there is an s^e-time hsg for degree-s polynomials computed by size-s circuits that depend only on the first $n := \lceil (64e^2)^{1/\epsilon} \rceil \geq 192e^2 \log(128e^2)^{1/\epsilon}$ variables. Note that $e \geq (n-1)^{\epsilon/2}/8 > n^\delta$ can be easily ensured, thus, s^e-time is more than s^{n^δ}-time. Now we simply invoke Theorem 2. $\square$

Remarks– 1) The NW (ℓ, n, d)-design that we are using, in the i-th iteration (Lemma 19), has its respective parameters in the "ratio" $f_i^2 : f_i : 1$ (roughly). This seems to be the reason why we need second-exponent δ slightly less than 1/2. We leave it as an open question to improve this.
2) We can give a more refined analysis in the above proofs by "decoupling" the time-complexity from the degree of the hsg. For example, we can begin with a much weaker hypothesis— for constant $\delta < 1/2$ and an arbitrarily large function $\mu(\cdot)$, an $\left(s^{\mu(n)}, s^{n^\delta}\right)$-hsg for size-$s$ degree-s circuits that depend only on the first n variables —and still get the same conclusion as in Theorem 3. This will require analysing the bit complexity (i.e. time) and the algebraic complexity (i.e. degree of the hsg) separately in the proof of Lemma 19. We skip the details for now.

5 Shallow depths, tiny variables– Proof of Theorem 4

Shallow circuits. *Diagonal depth*-4 circuits compute polynomials of the form $\sum_{i\in[k]} c_i f_i^{a_i}$ where f_i's are sparse polynomials in $\mathbb{F}[x_1,\ldots,x_n]$ of degree $\le b$, $a_i \le a$ and c_i's in $\mathbb{F}$. A standard notation to denote this class is $\Sigma\wedge^a\Sigma\Pi^b(n)$. This is a special case of the *depth*-4 $\Sigma^k\Pi^a\Sigma\Pi^b(n)$ model that computes polynomials of the form $\sum_{i\in[k]}\prod_{j\in[a]} f_{i,j}$ where $f_{i,j}$'s are sparse polynomials in $\mathbb{F}[x_1,\ldots,x_n]$ of degree $\le b$. The superscripts k,a,b on the gates denote an upper bound on the respective fanin (whenever it needs to be emphasized).

We denote $\Sigma\Pi\Sigma\Pi^1$ circuits by $\Sigma\Pi\Sigma$ and call them *depth*-3. We also study a model quite close to it– $\Sigma\Pi\Sigma\wedge^b$ –we call it *preprocessed depth*-3 because, in this work, this model will appear on simply substituting univariate monomials in the variables of a depth-3 circuit. It degenerates to depth-3 again if $b = 1$.

We prove Theorem 4 in two different ways. First, by assuming an efficient $O(s^{n/2}/\log^2 s)$-degree hsg for polynomials computed by size-s $\Sigma\wedge^a\Sigma\Pi$ circuits that depend only on the first n variables ($a(s)$ is an arbitrarily slow growing function), we get to the conclusion of Theorem 4. Second, by assuming an efficient $O(s^{n/2}/\log^2 s)$-degree hsg for polynomials computed by size-s $\Sigma\Pi\Sigma\wedge$ circuits that depend only on the first n variables, we get to the conclusion of Theorem 4. Both the models seem weaker than general depth-4 circuits. So one would expect that solving PIT for these models would be easier.

Our proofs will go via a plethora of intermediate models. Theorems 20 & 21 together give the proof of our first approach. Theorems 22 & 23 together give the proof of the second approach. One can notice that in all these theorems we prove the existence of an efficient variable reduction map for circuits that preserves nonzeroness. It is stronger than proving quasipolynomial hsg for size-s degree-d circuits. However, after the variable reduction, if we apply hsg of the trivial PIT derandomization [Sch80], we get an $(sd)^{O(\log(sd))}$ time hsg.

Theorem 20 ($\Sigma\wedge^a\Sigma\Pi$ computing multilinear)**.** *Suppose that for some constant $n \ge 2$ and some arbitrarily slow growing function $a(s)$, we have a $\big(poly(s), O(s^n/\log^2 s)\big)$-hsg for* multilinear *polynomials computed by size-s $\Sigma\wedge^a\Sigma\Pi$ circuits that depend only on the first $n\log s$ variables.*

Then, for N-variate d-degree size-s circuits, we have a poly(sd)-time nonzeroness preserving variable reducing polynomial map ($N \mapsto O(\log(sd))$) and Conjecture 1 holds. Furthermore, after variable reduction, the degree of the new polynomial will be poly(sd).

Proof sketch. The proof is along the lines of [AV08, Thm.3.2] and is described in Appendix E.

For all $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of multilinear polynomials computed by size-s $\Sigma\wedge^a\Sigma\Pi$ circuits that depend only on the first $n\log s$ variables. First, using the $O(s^n/\log^2 s)$-degree hsg we can construct a family of multilinear polynomials $\{q_m\}_m$ which is E-computable (Lemma 5) but not computable by $2^{o(m)}$-size circuits (by 'depth-4 chasm').

Using this hard polynomial family we get both the variable reduction and Conclusion 1. Invoking Lemma 9, in poly(sd) time, we can convert a nonzero d-degree N-variate polynomial computed by a size-s circuit to a nonzero $O(\log(sd))$-variate poly(sd)-degree polynomial. Conjecture 1 follows from Lemma 10. $\square$

Remarks– 1) Note that a $\big(\tilde{O}(s^n), s^n\big)$-hsg for *multilinear* $n\log s$ variate polynomials is trivial. As one can simply use $\{0,1\}^{n\log s}$ as the hitting-set.

2) An efficient $s^n/\omega(\log s)$ degree hsg in the hypothesis would also suffice.

3) Can we get a conclusion as strong as in Theorem 1? In the proof above we get a variable reduction map to log-variate; but this map when applied on a general circuit results in a *non*-multilinear polynomial. So, we cannot use the hsg provided in the hypothesis and have to do poly(s)-time PIT on the log-variate $\Sigma\wedge^a\Sigma\Pi$ circuit by some other means (currently unknown).

Theorem 21 (Tiny variate $\Sigma \wedge^a \Sigma\Pi$)**.** *Suppose that for some constant $n \geq 3$ and some arbitrarily slow growing function a, we have a $\big(poly(s), O(s^{n/2}/\log^2 s)\big)$-hsg for size-$s$ $\Sigma \wedge^a \Sigma\Pi$ circuits that depend only on the first n variables. Then, we get all the conclusions of Theorem 20.*

Proof. For all $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of multilinear polynomials computed by size-s $\Sigma \wedge^a \Sigma\Pi$ circuits that depend only on the first $n \log s$ variables. For all $s \in \mathbb{N}$, let $\mathcal{T}_s$ be the set of polynomials computed by size-s $\Sigma \wedge^a \Sigma\Pi$ circuits that depend only on the first n variables. By the hypothesis, we have an efficient $O(s^{n/2}/\log^2 s)$-degree hsg for $\mathcal{T}_s$. Next, we convert every nonzero polynomial in $\mathcal{P}_s$ to a nonzero polynomial in $\mathcal{T}_{O(s^2)}$ in poly(s) time. Now applying the given hsg for $\mathcal{T}_{O(s^2)}$, we get an efficient $O(s^n/\log^2 s)$-degree hsg for $\mathcal{P}_s$. Next invoking Theorem 20, we get our conclusion.

We describe the reduction from $\mathcal{P}_s$ to $\mathcal{T}_{O(s^2)}$. Let P be a nonzero polynomial in $\mathcal{P}_s$. Let $m := n \log s$. Partition the variable set $\{x_1, \ldots, x_m\}$ into n blocks $B_j, j \in [n]$, each of size $\log s$. Let $B_j := \{x_{u(j)+1}, x_{u(j)+2}, \ldots, x_{u(j)+\log s}\}$, for all $j \in [n]$ and $u(j) := (j-1)\log s$. Consider the variable-reducing "local Kronecker" map $\varphi : x_{u(j)+i} \mapsto y_j^{2^i}$. Note that $\varphi(P) \in \mathbb{F}[y_1, \ldots, y_n]$, and its individual-degree is at most $2s$. It is easy to see that $\varphi(P) \neq 0$ (basically, use the fact that P computes a nonzero multilinear polynomial and φ keeps the multilinear monomials distinct). Finally, $\varphi(P)$ becomes an n-variate $\Sigma \wedge^a \Sigma\Pi$ circuit of size at most $s + s \cdot 2^{\log s} = O(s^2)$. Thus, $\big(\text{poly}(s), O(s^n/\log^2 s)\big)$-hsg for $\mathcal{T}_{O(s^2)}$ gives a $\big(\text{poly}(s), O(s^n/\log^2 s)\big)$-hsg for P. $\square$

In next two lemmas, we describe our second approach to prove Theorem 4.

Theorem 22 (Depth-3 computing multilinear)**.** *Suppose that for some constant $n \geq 2$, we have a $\big(poly(s), O(s^n/\log^2 s)\big)$-hsg for* multilinear *polynomials computed by size-s depth-3 circuits that depend only on the first $n \log s$ variables. Then, we get all the conclusions of Theorem 20.*

Proof. Proof will be similar to proof of Theorem 20. Main difference is that there we were dealing with depth-4 circuits, but here we have depth-3 circuits. So we need 'depth-3-reduction' result [GKKS13] with 'depth-4-reduction' result [AV08]. We only sketch the main points here.

First, we construct a hard polynomial family from the hsg. According to the hypothesis, for $n \log s$-variate multilinear polynomials computed by size-s depth-3 circuits we have an $O(s^n/\log^2 s)$-degree hsg. For all $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of $n \log s$-variate multilinear polynomials computed by size-s depth-3 circuits. Let $m := n \log s$. Let $\mathbf{f}(y)$ be the $\big(\text{poly}(s), O(s^n/\log^2 s)\big)$-hsg of $\mathcal{P}_s$. Now we consider the annihilator of $\mathbf{f}(y)$ to get a hard polynomial. Let k be the number of m-variate $m/2$-degree multilinear monomials. Then $k = \binom{m}{m/2} \geq 2^m/\sqrt{2m} = s^n/\sqrt{2m} > O(s^n/\log^2 s) \cdot m$ (for large enough s). Thus, by linear algebra similar to Lemma 5, we get an m-variate $m/2$-degree multilinear homogeneous annihilating polynomial $q_m \notin \mathcal{P}_s$ and computable in poly(s)-time. Importantly $q_m \notin \mathcal{P}_s$, thus, no depth-3 circuit of size $< s = 2^{\Theta(m)}$ can compute it. Next we show that it is also not computable by any $2^{o(m)}$-size algebraic circuit.

For the sake of contradiction, assume that q_m has a $2^{o(m)}$-size circuit. Repeat the depth-reduction arguments, as in the proof of Theorem 20, after cutting at some depth $t = \omega(1)$. Let $a := 5^t$ and $b := m/2^{t+1}$. Here, we can also ensure $a, b = o(m) = o(\log s)$, $a = \omega(1)$, and we have a $2^{o(m)}$-size shallow circuit for q_m of the form $\Sigma\Pi^a\Sigma\Pi^b$.

It was shown in [GKKS13] that any size-s' n-variate $\Sigma\Pi^a\Sigma\Pi^b$ circuit can be transformed to a poly$(s'2^{a+b})$-size n-variate $\Sigma\Pi\Sigma^b$ circuit. Applying it here, we get a depth-3 circuit C', computing q_m, of the form $\Sigma\Pi\Sigma$ and size $2^{o(m)} \cdot 2^{a+b} = 2^{o(m)}$. This gives a contradiction, since no depth-3 circuit of size $< s = 2^{\Theta(m)}$ can compute it.

Thus, we have an E-computable family of multilinear polynomials $\{q_m\}_{m \geq 1}$ that has no circuit of size $2^{o(m)}$. Using this hard polynomial family we get both the variable reduction and Conjecture 1 as before. $\square$

Theorem 23 (Tiny variate $\Sigma\Pi\Sigma\wedge$)**.** *Suppose that for some constant $n \geq 3$, we have a $\big(poly(s), O(s^{n/2}/\log^2 s)\big)$-hsg for polynomials computed by size-s $\Sigma\Pi\Sigma\wedge$ circuits that depend only on the first n variables. Then, we get all the conclusions of Theorem 20.*

Proof. The proof is similar to that of Theorem 21. For all $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of multilinear polynomials computed by size-s depth-3 circuits that depend only on the first $n \log s$ variables. For all $s \in \mathbb{N}$, let $\mathcal{T}_s$ be the set of polynomials computed by size-s $\Sigma\Pi\Sigma\wedge$ circuits that depend only on the first n variables. According to the hypothesis, we have an $O(s^{n/2}/\log^2 s)$-degree hsg for $\mathcal{T}_s$. Next, we convert every nonzero polynomial in $\mathcal{P}_s$ to a nonzero polynomial in $\mathcal{T}_{O(s^2)}$ in poly(s) time. Now applying $O(s^n/\log^2 s)$-degree hsg for $\mathcal{T}_{O(s^2)}$, we get an efficient $O(s^n/\log^2 s)$-degree hsg for $\mathcal{P}_s$. Next invoking Theorem 22, we get our conclusion.

Now we describe the reduction from $\mathcal{P}_s$ to $\mathcal{T}_{O(s^2)}$. Let P be a nonzero polynomial in $\mathcal{P}_s$. Let $m := n \log s$. Partition the variable set $\{x_1, \ldots, x_m\}$ into n blocks $B_j, j \in [n]$, each of size $\log s$. Let $B_j := \{x_{u(j)+1}, x_{u(j)+2}, \ldots, x_{u(j)+\log s}\}$, for all $j \in [n']$ and $u(j) := (j-1)\log s$. Consider the variable-reducing "local Kronecker" map $\varphi : x_{u(j)+i} \mapsto y_j^{2^i}$. Note that $\varphi(P) \in \mathbb{F}[y_1, \ldots, y_n]$, and its individual-degree is at most $2s$. It is easy to see that $\varphi(P) \neq 0$ (basically, use the fact that P computes a nonzero multilinear polynomial and φ keeps the multilinear monomials distinct). Finally, $\varphi(P)$ becomes an n-variate $\Sigma\Pi\Sigma\wedge$ circuit of size at most $s + s \cdot 2^{\log s} = O(s^2)$. Thus, using the $O(s^n/\log^2 s)$-degree hsg for $\mathcal{T}_{O(s^2)}$, we get a $\big(\text{poly}(s), O(s^n/\log^2 s)\big)$-hsg for P. $\square$

Remark. Can we get a result like the above with depth-3 circuits in the hypothesis? At this point it is not clear how to get to arbitrarily tiny variate $\Sigma\Pi\Sigma$ because: **1)** the above trick of applying local-Kronecker map, to reduce variables from $n \log s$ to n, increases the circuit depth to 4. Moreover any such map has to be *non*-linear, otherwise the resulting monomials are too few, and **2)** in the tiny variate regime we need degree $\geq \Omega(s)$ so that the hsg of the model can be used to get a 'hard' polynomial. With such a high degree we cannot apply [GKKS13] to transform depth-4 (say in Theorem 21) to depth-3 in polynomial-time.

Proof of Theorem 4. Let a be an arbitrarily slow growing function. For all $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of polynomials computed by size-s $\Sigma \wedge^a \Sigma\Pi$ circuits that depend only on the first n variables. For all $s \in \mathbb{N}$, let $\mathcal{T}_s$ be the set of polynomials computed by size-s $\Sigma\Pi\Sigma\wedge$ circuits that depend only on the first n variables. We show that $\big(\text{poly}(s), O(s^{n/2}/\log^2 s)\big)$-hsg for $\mathcal{P}_s$ or $\mathcal{T}_s$ gives the conclusion of Theorem 4.

Using the hsg for $\mathcal{P}_s$, Theorem 21 gives an efficient variable reduction and Conjecture 1.

Using the hsg for $\mathcal{T}_s$, Theorem 23 gives an efficient variable reduction and Conjecture 1.

After the variable reduction, if we apply hsg of the trivial PIT derandomization [Sch80], we get an $(sd)^{O(\log(sd))}$ time hsg.

To see that the original statement could be proved for *any* field $\mathbb{F}$: Observe that 'depth-4 reduction' [AV08, Thm.3.2] works for any field. Similarly, we get versions of Theorems 20 & 21 using $\Sigma\Pi^a\Sigma\Pi$ in the respective hypothesis. Also, see the remarks after the proofs of Lemma 12 and Theorem 1. $\square$

5.1 Depth-3 fpt-blackbox PIT

In this section, we show that we merely need an fpt-algorithm (wrt parameters n, d) for polynomials computed by depth-3 circuits. In *fpt-algorithm*, one provides input with multiple parameters, with the intention that the running time will be polynomial in input size but possibly exponential (or worse) in other parameters [DF13]. We show that to get the same conclusion as Theorem 4, we merely need a fpt-blackbox PIT for depth-3 circuits computing *multilinear* polynomials. These polynomials have three important complexity parameters: **1)** s, the size of the depth-3 circuit computing the polynomial, **2)** m, the number of variables, and **3)** d, the degree of the polynomial which is upper bounded by m. Here, circuit is the input. So, the running time of the fpt-blackbox PIT must depend polynomially on s. We consider m and d as the extra parameters of the fpt-blackbox PIT. Next, we describe the desired dependence on them.

Theorem 24 (Depth-3 tiny variables & degree)**.** *Suppose that we have a poly($2^{m+d}, s$)-time computable $\left(2^m + 4^d + s^2\right)/\log^2 s$ degree hsg for m-variate, degree-d* multilinear *polynomials computed by size-s depth-3 circuits.*

Then, we get all the conclusions of Theorem 20.

Remark– 1) Since exponential dependence on m, d is allowed, one can hope that designing such an hsg would be easier than the numerous unsolved PIT cases that the community has attempted till date.

2) Another width-2 ABP version is stated in Theorem 25 (with a worse dependence on m).

3) The number of monomials is 2^m. Thus, the hsg design challenge in the hypothesis of Theorem 24 is barely "above" triviality!

Proof. The proof strategy is identical to that of Theorem 22. So, we only sketch the main points here. Pick a constant $n \geq 2$. First, we construct a hard polynomial family from the hsg. According to the hypothesis, for $n \log s$-variate and $\frac{n}{2} \log s$-degree multilinear polynomials computed by size-s depth-3 circuits, we have a $\left(\text{poly}(s), O(s^n/\log^2 s)\right)$-hsg.

From this point onwards the proof of Theorem 22 is identical and we are done.

In poly(sd)-time, the variable reduction map reduces any d-degree nonzero polynomial computed by a size-s circuit to a poly(sd)-degree and $O(\log sd)$-variate nonzero polynomial. Now, trivial PIT derandomization [Sch80] gives an $(sd)^{O(\log(sd))}$-time hsg easily. □

5.2 Log-variate width-2 ABP or depth-3 circuit

A polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ has a size-s *width-2 algebraic branching program* (ABP) if it is the $(1,1)$-th entry in the product of s 2×2 matrices (having entries in $\mathbb{F} \cup \{x_i | i\}$).

Theorem 25 (Log-variate width-2 ABP)**.** *Suppose that for some constant $e \geq 1$, we have a $\left(poly(s), O(s^e)\right)$-hsg for polynomials (resp. $2^{e+1} \log s$-degree polynomials) computed by size-s width-2 upper-triangular ABP (resp. depth-3 circuit) that depend only on the first $\log s$ variables. Then, we get all the conclusions of Theorem 20.*

Proof. In [SSS09, Thm.3] an efficient transformation was given that rewrites a size-s depth-3 circuit, times a special product of linear polynomials, as a poly(s)-size width-2 upper-triangular ABP. Thus, an hsg for the latter model gives a similar hsg for the former. So, from the hypothesis for some constant e, we have a $\left(\text{poly}(s), O(s^e)\right)$-hsg for $2^{e+1} \log s$-degree polynomials computed by size-s depth-3 circuits that depend only on the first $\log s$ variables.

For all $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of $\log s$-variate, $2^{e+1} \log s$-degree polynomials computed by size-s depth-3 circuits. Let $d := 2^{e+1} \log s$. Let $\mathbf{f}(y)$ be the $(\text{poly}(s), O(s^e))$-hsg of $\mathcal{P}_s$. Now we consider the annihilator of $\mathbf{f}(y)$ to get a hard polynomial. Let k be the number of $m := \log s$-variate d-degree monomials. Then $k = \binom{m+d-1}{m} > 2^{(e+1)\log s} = s^{e+1}$. Since $k > O(s^e) \cdot d$, we get an m-variate d-degree homogeneous annihilating polynomial $q_m \notin \mathcal{P}_s$ and computable in $s^{O(1)}$ time. The analysis is similar to Lemma 5. Importantly $q_m \notin \mathcal{P}_s$, thus no depth-3 circuit of size $< s = 2^{\Theta(m)}$ can compute it.

From this point onwards the proof of Theorem 22 is identical and we are done. □

6 Conclusion

We discover the phenomenon of 'efficient bootstrapping' a partial hitting-set generator to a complete one for poly-degree circuits. This inspires a plethora of circuit models. In particular, we introduce the tiny variable diagonal depth-4 (resp. tiny variants of depth-3, width-2 ABP and preprocessed depth-3) model with the motivation that its poly-time hitting-set would: **(1)** solve VP PIT (in quasipoly-time) via a poly-time variable-reducing polynomial map ($n \mapsto \log sd$), and **(2)** prove that either E$\not\subseteq$#P/poly or VNP has polynomials of algebraic complexity $2^{\Omega(n)}$.

Since now we could focus solely on the PIT of VP circuits that depend only on the first *sub*-log (or even constant!) many variables, we need to initiate a study of properties that are useful in that regime. Furthermore, we only need to optimize the size of the hitting-set (& not its time). This work throws up a host of tantalizing models and poses several interesting questions:

- Could the bootstrapping property in Theorem 1 be improved (say, to the function $\log^\star s$)?
- Could the constant parameters in Theorems 2 & 3 be improved? In particular, does $s^{o(n)}$-time blackbox PIT suffice in the latter hypothesis?
- Could we show that the g in Corollary 6 is in VNP and not merely E-computable? This would tie blackbox PIT tightly with the question VNP$\neq$VP (& we can drop 'E$\not\subseteq$# P/poly'). This will require starting with a more structured hsg, so that its annihilator g is a polynomial whose coefficient bits are (#P/poly)-computable. Numerous examples of such polynomials, arising from basic hitting-set designs, can be found in [Agr11, KP11] and [Koi11, Sec.4].
- Could we solve *whitebox* PIT for $\log^\star s$ variate (or degree) models? Could it be bootstrapped?
- Could we prove nontrivial lower bounds against the tiny variable (or degree) models?
- Could we solve PIT for n-variate degree-s size-s circuits in $s^{O(\sqrt{n})}$-time?
- Is there a poly(s)-time computable, $O(s^3)$-size hitting-set for 6-variate size-s $\Sigma\Pi\Sigma\wedge$ polynomials?
- An $s^{\exp(n)}$-time computable, $O(s^{n/2})$-size hitting-set for size-s $\Sigma\Pi\Sigma(n)$?
- Could we do blackbox PIT for tiny variable ROABP? For instance, given oracle $C = \sum_{i\in[k]} \prod_{j\in[n]} f_{i,j}(x_j)$ of size$\leq s$, we want a poly$(s, \mu(n))$-hsg, for some μ. It is known that diagonal depth-3 blackbox PIT reduces to this problem if we demand $\mu(n) \leq 2^{O(n)}$ [FSS14].

  Note that for n-variate size-s ROABPs, $s^{O(\log n)}$-time hsg is already known [AGKS15]. But, we can ask the following open questions:
- Efficient blackbox PIT for size-s, $\log s$-variate, individual-degree-($\log^\star s$) ROABPs?
- Blackbox PIT for size-s, ($\log^\star s$) $\log s$-variate, multilinear ROABPs?
- Blackbox PIT for size-s, ($\log^\star s$) $\log s$-variate, $\log s$-degree, diagonal depth-3 circuits?

Acknowledgements

We thank Michael Forbes, Ramprasad Saptharishi, Pranjal Dutta, Thomas Thierauf and Nikhil Balaji for many useful discussions. N.S. thanks the organizers of algebraic complexity workshops in 2014 (MPI Saarbrücken & TIFR Mumbai) that initiated the early discussions with Michael Forbes related to log-variate diagonal depth-3 PIT. S.G. & N.S. thank the organizers, and participants, of the IMSc Chennai workshop in 2017 for discussing the preliminary results. Thanks are due to the anonymous reviewers for giving numerous suggestions to improve an earlier draft. N.S. thanks the funding support from DST (DST/SJF/MSA-01/2013-14).

References

[AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, NY, USA, 1st edition, 2009. 3, 8, 30, 31

[AGKS15] Manindra Agrawal, Rohit Gurjar, Arpita Korwar, and Nitin Saxena. Hitting-sets for ROABP and sum of set-multilinear circuits. *SIAM Journal on Computing*, 44(3):669–697, 2015. 2, 21

[Agr05] Manindra Agrawal. Proving lower bounds via pseudo-random generators. In *FSTTCS 2005: Foundations of Software Technology and Theoretical Computer Science, 25th International Conference, Hyderabad, India, December 15-18, 2005, Proceedings*, pages 92–105, 2005. 2, 8, 27

[Agr11] Manindra Agrawal. On the arithmetic complexity of Euler function. In *Proceedings of the 6th International Computer Science Symposium in Russia, LNCS 6651*, pages 43–49. 2011. 21

[AKS04] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. PRIMES is in P. *Annals of mathematics*, pages 781–793, 2004. 2

[ASS13] Manindra Agrawal, Chandan Saha, and Nitin Saxena. Quasi-polynomial hitting-set for set-depth-Δ formulas. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 321–330, 2013. 2

[ASSS12] Manindra Agrawal, Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. Jacobian hits circuits: hitting-sets, lower bounds for depth-d occur-k formulas & depth-3 transcendence degree-k circuits. In *STOC*, pages 599–614, 2012. 2

[AV08] Manindra Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 67–75, 2008. 5, 7, 8, 17, 18, 19, 32

[BMS13] M. Beecken, J. Mittmann, and N. Saxena. Algebraic Independence and Blackbox Identity Testing. *Inf. Comput.*, 222:2–19, 2013. (Conference version in ICALP 2011). 2

[Bür01] Peter Bürgisser. The complexity of factors of multivariate polynomials. In *In Proc. 42th IEEE Symp. on Foundations of Comp. Science*, 2001. 3

[Bür13] Peter Bürgisser. *Completeness and reduction in algebraic complexity theory*, volume 7. Springer Science & Business Media, 2013. 4, 9, 10, 14, 27, 28, 29, 30

[CDGK91] Michael Clausen, Andreas Dress, Johannes Grabmeier, and Marek Karpinski. On zero-testing and interpolation of k-sparse multivariate polynomials over finite fields. *Theoretical Computer Science*, 84(2):151–164, 1991. 4

[DdOS14] Zeev Dvir, Rafael Mendes de Oliveira, and Amir Shpilka. Testing Equivalence of Polynomials under Shifts. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming, Part I*, volume 8572 of *Lecture Notes in Computer Science*, pages 417–428. Springer International Publishing, 2014. 2

[DF13] Rodney G Downey and Michael R Fellows. *Fundamentals of parameterized complexity*, volume 4. Springer, 2013. 19

[DL78] Richard A. Demillo and Richard J. Lipton. A probabilistic remark on algebraic program testing. *Information Processing Letters*, 7(4):193 – 195, 1978. 2

[DSS17] Pranjal Dutta, Nitin Saxena, and Amit Sinhababu. Discovering the roots: Uniform closure results for algebraic classes under factoring. Technical report, https://www.cse.iitk.ac.in/users/nitin/research.html, 2017. (To appear in 50th ACM Symposium on Theory of Computing (STOC), 2018). 3

[FGT16] Stephen A. Fenner, Rohit Gurjar, and Thomas Thierauf. Bipartite perfect matching is in quasi-NC. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 754–763, 2016. 2

[Fis94] Ismor Fischer. Sums of like powers of multivariate linear forms. *Mathematics Magazine*, 67(1):59–61, 1994. 9

[For15] Michael A Forbes. Deterministic divisibility testing via shifted partial derivatives. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 451–465. IEEE, 2015. 2

[FS12] Michael A. Forbes and Amir Shpilka. On identity testing of tensors, low-rank recovery and compressed sensing. In *STOC*, pages 163–172, 2012. 2

[FS17] Michael A. Forbes and Amir Shpilka. A PSPACE construction of a hitting set for the closure of small algebraic circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:163, 2017. (To appear in 50th ACM Symposium on Theory of Computing (STOC), 2018). 3

[FSS14] Michael A. Forbes, Ramprasad Saptharishi, and Amir Shpilka. Hitting sets for multilinear read-once algebraic branching programs, in any order. In *Symposium on Theory of Computing (STOC), New York, NY, USA, May 31 - June 03, 2014*, pages 867–875, 2014. 2, 21

[FSV17] Michael A Forbes, Amir Shpilka, and Ben Lee Volk. Succinct hitting sets and barriers to proving algebraic circuits lower bounds. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 653–664. ACM, 2017. 3

[Für09] Martin Fürer. Faster integer multiplication. *SIAM Journal on Computing*, 39(3):979–1005, 2009. 7

[GGOW16] Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. A deterministic polynomial time algorithm for non-commutative rational identity testing. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016*, pages 109–117, 2016. 3

[GKKS13] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth three. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 578–587, 2013. 5, 8, 9, 18, 19

[GKS17] Rohit Gurjar, Arpita Korwar, and Nitin Saxena. Identity testing for constant-width, and any-order, read-once oblivious arithmetic branching programs. *Theory of Computing*, 13(2):1–21, 2017. (Preliminary version in CCC'16). 2

[GKST16] Rohit Gurjar, Arpita Korwar, Nitin Saxena, and Thomas Thierauf. Deterministic identity testing for sum of read-once oblivious arithmetic branching programs. *Computational Complexity*, pages 1–46, 2016. (Conference version in CCC 2015). 2

[GMQ16] Joshua A. Grochow, Ketan D. Mulmuley, and Youming Qiao. Boundaries of VP and VNP. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 34:1–34:14, 2016. 2

[Gro15] Joshua A Grochow. Unifying known lower bounds via geometric complexity theory. *computational complexity*, 24(2):393–475, 2015. 2

[GSS18] Zeyu Guo, Nitin Saxena, and Amit Sinhababu. Algebraic dependencies and PSPACE algorithms in approximative complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:19, 2018. 3

[GT16] Rohit Gurjar and Thomas Thierauf. Linear matroid intersection is in quasi-NC. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:182, 2016. (To appear in STOC'17). 2

[GTV17] Rohit Gurjar, Thomas Thierauf, and Nisheeth K. Vishnoi. Isolating a vertex via lattices: Polytopes with totally unimodular faces. *CoRR*, abs/1708.02222, 2017. 2

[HS80] Joos Heintz and Claus-Peter Schnorr. Testing polynomials which are easy to compute (extended abstract). In *Proceedings of the 12th Annual ACM Symposium on Theory of Computing, April 28-30, 1980, Los Angeles, California, USA*, pages 262–272, 1980. 2, 4, 8

[Kal89] Erich Kaltofen. Factorization of polynomials given by straight-line programs. *Randomness and Computation*, 5:375–412, 1989. 3, 10, 14, 28, 30

[KI03] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. In *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 355–364, 2003. 2, 8

[KI04] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004. 4, 6, 8, 10

[Koi11] Pascal Koiran. Shallow circuits with high-powered inputs. In *Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 7-9, 2011. Proceedings*, pages 309–320, 2011. 21

[KP09] Pascal Koiran and Sylvain Perifel. VPSPACE and a transfer theorem over the reals. *Computational Complexity*, 18(4):551–575, 2009. 27, 29

[KP11] Pascal Koiran and Sylvain Perifel. Interpolation in Valiant's theory. *Computational Complexity*, 20(1):1–20, 2011. 21

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

[KS16a] Mrinal Kumar and Shubhangi Saraf. Arithmetic circuits with locally low algebraic rank. In *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*, pages 34:1–34:27, 2016. 2

[KS16b] Mrinal Kumar and Shubhangi Saraf. Sums of products of polynomials in few variables: Lower bounds and polynomial identity testing. In *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*, pages 35:1–35:29, 2016. 2

[KSS14] Swastik Kopparty, Shubhangi Saraf, and Amir Shpilka. Equivalence of polynomial identity testing and deterministic multivariate polynomial factorization. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 169–180, 2014. 2

[LG14] François Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th international symposium on symbolic and algebraic computation*, pages 296–303. ACM, 2014. 7

[LLS17] Guillaume Lagarde, Nutan Limaye, and Srikanth Srinivasan. Lower bounds and PIT for non-commutative arithmetic circuits with restricted parse trees. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:77, 2017. (Appeared in MFCS'17). 3

[LMP16] Guillaume Lagarde, Guillaume Malod, and Sylvain Perifel. Non-commutative computations: lower bounds and polynomial identity testing. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:94, 2016. 3

[Muk16] Partha Mukhopadhyay. Depth-4 identity testing and Noether's normalization lemma. In *International Computer Science Symposium in Russia*, pages 309–323. Springer, 2016. 2

[Mul12a] Ketan D Mulmuley. The GCT program toward the P vs. NP problem. *Communications of the ACM*, 55(6):98–107, 2012. 2

[Mul12b] Ketan D. Mulmuley. Geometric complexity theory V: Equivalence between blackbox derandomization of polynomial identity testing and derandomization of Noether's normalization lemma. In *FOCS*, pages 629–638, 2012. 2

[Mul17] Ketan Mulmuley. Geometric complexity theory V: Efficient algorithms for Noether normalization. *Journal of the American Mathematical Society*, 30(1):225–309, 2017. 2, 3

[MV17] Daniel Minahan and Ilya Volkovich. Complete derandomization of identity testing and reconstruction of read-once formulas. In *LIPIcs-Leibniz International Proceedings in Informatics, CCC'17*, volume 79. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017. 2

[MVV87] Ketan Mulmuley, Umesh V. Vazirani, and Vijay V. Vazirani. Matching is as easy as matrix inversion. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, pages 345–354, 1987. 2

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994. 2, 4, 8

[PSS16] Anurag Pandey, Nitin Saxena, and Amit Sinhababu. Algebraic independence over positive characteristic: New criterion and applications to locally low algebraic rank circuits. In *41st International Symposium on Mathematical Foundations of Computer Science, MFCS 2016, August 22-26, 2016 - Kraków, Poland*, pages 74:1–74:15, 2016. (In print, Computational Complexity, 2018). 2

[Sap16] Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity. Technical report, https://github.com/dasarpmar/lowerbounds-survey/, 2016. 3, 10, 11, 12, 27, 33

[Sax08] Nitin Saxena. Diagonal circuit identity testing and lower bounds. In *ICALP*, volume 5125 of *Lecture Notes in Computer Science*, pages 60–71. Springer, 2008. 2, 5

[Sax09] Nitin Saxena. Progress on polynomial identity testing. *Bulletin of the EATCS*, 99:49–79, 2009. 2

[Sax13] Nitin Saxena. Progress on polynomial identity testing - II. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:186, 2013. 2

[Sch80] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, October 1980. 2, 9, 11, 14, 17, 19, 20

[SS12] Nitin Saxena and C. Seshadhri. Blackbox identity testing for bounded top-fanin depth-3 circuits: The field doesn't matter. *SIAM Journal on Computing*, 41(5):1285–1298, 2012. 2

[SSS09] Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. The power of depth 2 circuits over algebras. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2009, December 15-17, 2009, IIT Kanpur, India*, pages 371–382, 2009. 20

[SSS13] Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. A case of depth-3 identity testing, sparse factorization and duality. *Computational Complexity*, 22(1):39–69, 2013. 2

[ST17] Ola Svensson and Jakub Tarnawski. The matching problem in general graphs is in Quasi-NC. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:59, 2017. (Appeared in FOCS'17). 2

[SY10] Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3-4):207–388, 2010. 2

[Val79] Leslie G. Valiant. Completeness classes in algebra. In *Proceedings of the 11h Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1979, Atlanta, Georgia, USA*, pages 249–261, 1979. 2

[Wig17] Avi Wigderson. Low-depth arithmetic circuits: technical perspective. *Communications of the ACM*, 60(6):91–92, 2017. 2

[Yao82] Andrew C Yao. Theory and application of trapdoor functions. In *Foundations of Computer Science, 1982. 23rd Annual Symposium on*, pages 80–91. IEEE, 1982. 3

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposium on Symbolic and Algebraic Computation*, EUROSAM '79, pages 216–226, 1979. 2

A Proofs from Section 1

Theorem 0 (RESTATED). If we have poly(s)-time blackbox PIT for size-s degree-s circuits, then Conjecture 1 holds.

Proof sketch. For all $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of polynomials computed by size-s degree-s circuits. Using basic linear algebra, we can construct an m-variate multilinear annihilator q_m, where $m = O(\log s)$, for the hsg of $\mathcal{P}_s$ in $2^{O(m)}$-time. This q_m cannot lie in $\mathcal{P}_s$, otherwise q_m evaluated at hsg would be a nonzero polynomial (contradicting the annihilation property). For details, see the proof of [Agr05, Thm.51]. For the sake of contradiction, assume that it has a circuit of size $s^{o(1)}$. Since the degree of q_m is $O(\log s)$, we can invoke the structural log-depth reduction property (see [Sap16, Thm.5.15]) and get a $s^{o(1)}$-size $O(\log s)$-degree circuit computing q_m. Whence $q_m \in \mathcal{P}_s$, which is a contradiction. So we have the polynomial family $\{q_m\}_{m \geq 1}$ such that its coefficients are computable in $2^{O(m)}$-time (thus E-computable) but the algebraic circuit complexity is $> s^{\Omega(1)} = 2^{\Omega(m)}$.

Wlog we assume q_m to be a multilinear polynomial family with 0/1 coefficients; as, indexing a bit of a coefficient requires $O(m)$ bits and one can use the variable-increasing transformation from the proof of [KP09, Lem.3.9]. Also, if the coefficient function of a polynomial family is in #P/poly, then the polynomial family is in VNP [Bür13, Prop.2.20]. So, if we assume $E \subseteq$ #P/poly, then $\{q_m\}_m$ is in VNP. Thus, either $E \nsubseteq$ #P/poly or VNP has a polynomial family $\{q_m\}_m$ of algebraic circuit complexity $2^{\Omega(m)}$. □

B Proofs from Section 2– Preliminaries

LEMMA 5 (RESTATED). *Let $\mathbf{f}(y) = (f_1(y), \ldots, f_n(y))$ be a (t,d)-hsg for a set of n-variate polynomial $\mathcal{P}$. Then, there exists an n-variate polynomial $g(\mathbf{x})$ that is not in $\mathcal{P}$, is computable in* poly(tdn)*-time, has individual degree less than $\delta := \lceil d^{3/n} \rceil$, and is homogeneous of degree $(\delta - 1)n/2$.*

Proof. A natural candidate for $g(\mathbf{x})$ is any annihilating polynomial of the n polynomials $\mathbf{f}(y) = (f_1(y), \ldots, f_n(y))$, since for every nonzero $h \in \mathcal{P}$, $h(\mathbf{f})$ is nonzero. Define $\delta \geq 2$ as the smallest integer such that $\delta^{n/3} > d$. Consider $g(\mathbf{x})$ as a n-variate polynomial with individual degree less than δ and homogeneous of degree $(\delta - 1)n/2$. Then, $g(\mathbf{x})$ can be written as:

$$g(\mathbf{x}) = \sum_{|\mathbf{e}|=(\delta-1)n/2\ ,\ 0 \leq \mathbf{e}_i < \delta} c_{\mathbf{e}} \mathbf{x}^{\mathbf{e}} \tag{1}$$

where, $c_{\mathbf{e}}$'s are unknown to us. Note that the number of summands is at least $(\delta/2)^{n/2} \cdot \binom{n}{n/2} > \delta^{n/2}$ (for $n \geq 4$). The former estimate can be obtained by picking a subset $S \in \binom{[n]}{n/2}$ and considering all monomials in $\mathbf{x}_S$ of individual-degree $< \delta/2$. For every such monomial in $\mathbf{x}_S$ we can pick a (complementary) monomial $\mathbf{x}_{[n]\setminus S}$ with exponents from $\{\delta/2, \ldots, \delta-1\}$ such that the product of these two monomials has degree *exactly* $(\delta-1)n/2$.

We can fix all the $c_{\mathbf{e}}$'s to zero except the ones corresponding to an index-set I of size $\delta_0 := dn(\delta-1)/2+2 < \delta^{n/3}n(\delta-1)/2+2 \leq \delta^{n/2}$. This way we have exactly δ_0 unknown $c_{\mathbf{e}}$'s. To be an annihilating polynomial of $\mathbf{f}(y)$, we need $g(\mathbf{f}) = 0$. By comparing the coefficients of the monomials in y, both sides of Equation 1, we get a linear system in the unknowns.

Suppose that δ_1 is the degree of y in $g(\mathbf{f})$. Then, $g(\mathbf{f})$ can be written as $g(\mathbf{f}) = \sum_{i=0}^{\delta_1} p_i \cdot y^i$, where p_i's are linear polynomials in $c_{\mathbf{e}}$'s. The constraint $g(\mathbf{f}) = 0$ gives us a system of linear equations with the number of unknowns δ_0 and the number of equations $\delta_1 + 1$. The value of δ_1 can be at most $d \cdot n \cdot (\delta-1)/2$, which means that the number of unknowns δ_0 is greater than the number of equations δ_1. So, our system of *homogeneous* linear equations always has a nontrivial solution, which gives us a nonzero g as promised.

Computing $\mathbf{f}(y)$ takes t time and a solution of the linear equations can be computed in poly(tdn)-time. So, $g(\mathbf{x})$ can be computed in poly(tdn)-time. □

Lemma 9 (restated). *For some constant δ, let $\{q_m\}_{m\geq 1}$ be a multi-δ-ic polynomial family computable in $\delta^{O(m)}$ time, but it has no $\delta^{o(m)}$-size algebraic circuit.*

Then, for n-variate d-degree polynomials computed by size-s circuits we have a $\delta^{O(\log(sd))}$-time variable-reducing polynomial map, from n to $O(\log(sd))$, that preserves nonzeroness. Furthermore, after variable reduction, the degree of the new polynomial is poly(sd).

Proof. Note that there is a constant $c_0 > 0$ such that q_m requires $\Omega(\delta^{c_0 m})$-size algebraic circuits. Otherwise $\{q_m\}_{m\geq 1}$ will be in $\cap_{c>0}\text{Size}(\delta^{cm})$, and hence in Size$(\delta^{o(m)})$.

Let $\mathcal{P}$ be the set n-variate and d-degree polynomials computed by size-s circuits. Let $n' := sd \geq n$. Let $\mathcal{D} = \{S_1, \ldots, S_{n'}\}$ be a $(c_2 \log n', c_1 \log n', 10 \log n')$-design on the variable set $Z = \{z_1, \ldots, z_{c_2 \log n'}\}$. Constants $c_2 > c_1 > 10$ will be fixed later (guided by Lemma 8). Our hitting-set generator for $\mathcal{P}$ is defined as: for all $i \in [n]$, $x_i = q_{c_1 \log n'}(S_i) =: p_i$ with S_i as variables. Then, we show that for any nonzero polynomial $P(\mathbf{x}) \in \mathcal{P}$, $P(p_1, \ldots, p_n)$ is also nonzero.

For the sake of contradiction, assume that $P(p_1, \ldots, p_n)$ is zero. Since $P(\mathbf{x})$ is nonzero, we can find the smallest $j \in [n]$ such that $P(p_1, \ldots, p_{j-1}, x_j, \ldots, x_n) =: P_1$ is nonzero, but $P_1\big|_{x_j=p_j}$ is zero. Thus, $(x_j - p_j)$ divides P_1. Let $\mathbf{a}$ be an assignment on all the variables in P_1, except x_j and the variables S_j in p_j, with the property: P_1 at $\mathbf{a}$ is nonzero. Since P_1 is nonzero, we can find such an assignment. Now our new polynomial P_2 on the variables S_j is of the form:

$$P_2(S_j, x_j) \;=\; P(p'_1, \ldots, p'_{j-1}, x_j, a_{j+1}, \ldots, a_n)$$

where, for each $i \in [j-1]$, p'_i is the polynomial on the variables $S_i \cap S_j$, and a_i's are field constants decided by our assignment $\mathbf{a}$. By the design, for each $i \in [j-1]$, $|S_i \cap S_j| \leq 10 \log n'$. Since p'_i are polynomials on variables $S_i \cap S_j$ of individual-degree $< \delta$, each p'_i has a circuit (of trivial form $\Sigma\Pi$) of size at most $n\delta^{10 \log n'} \leq \delta^{11 \log n'}$. Then we have a circuit for P_2 of size at most $s_1 := s + n \cdot \delta^{11 \log n'}$, and degree at most $d_1 := d \cdot \delta c_1 \log n'$. Since $(x_j - p_j)$ divides P_2, we can invoke the Kaltofen's factorization algorithm [Kal89] (see [Bür13, Theorem 2.21] for the algebraic circuit complexity of factors) and get an algebraic circuit for p_j of size $(s_1 d_1)^{c_3}$, for some absolute constant c_3 (independent of c_1, c_2).

Now we fix constants c_1, c_2. Pick c_1 such that $\delta^{c_0 \cdot c_1 \log n'}$ is asympotically larger than $(2sn\delta^{11 \log n'} \cdot d\delta c_1 \log n')^{c_3} > (s_1 d_1)^{c_3}$. Since $sd = n'$ and $\delta \geq 2$, the absolute constant $c_1 := 15c_3/c_0$ (independent of c_2) satisfies the above condition.

Pick c_2, following Lemma 8, such that $c_2 \log n' > 10 \cdot (c_1 \log n')^2/(10 \log n')$. So, $c_2 := 1 + c_1^2$ works. With these values of c_1, c_2, we have a design that 'stretches' $c_2 \log n'$ variables to n subsets with the required 'low' intersection property. It is computable in poly(n')-time.

Moreover, if $P(p_1, \ldots, p_n)$ is zero then, by the above discussion, $p_j = q_{c_1 \log n'}(S_j)$ has a circuit of size $(s_1 d_1)^{c_3} = o(\delta^{c_0 \cdot c_1 \log n'})$. This violates the lower bound hypothesis. Thus, $P(p_1, \ldots, p_n)$ is nonzero.

The time for computing $(p_1, \ldots, p_n)$ depends on: **(1)** computing the design (i.e. poly(n')-time), and **(2)** computing $q_{c_1 \log n'}$ (i.e. $\delta^{O(\log n')}$-time). Thus, the nonzeroness-preserving variable-reducing polynomial map is computable in $\delta^{O(\log n')}$ time. After variable reduction, the degree of the new polynomial is $\leq nd \cdot \deg(q_{c_1 \log n'}) = (sd)^{O(1)}$. □

Corollary 6 (restated). *In the proof of Lemma 5, if $td = 2^{O(n)}$ then the polynomial family $g_n := g$, indexed by the variables, is E-computable.*

Proof. The linear system that we got can be solved in poly(tdn)-time. As it is homogeneous we can even get an integral solution in the same time-complexity. Thus, assuming $td = 2^{O(n)}$, the time-complexity of computing coef$_{\mathbf{x}^{\mathbf{e}}}(g)$ is poly(tdn)=poly(2^n) and g is multi-δ-ic ($\because \delta = \lceil d^{3/n} \rceil = O(1)$). In other words, if we consider the polynomials $g_n := g$, indexed by the variables, then the family $\{g_n\}_n$ is E-computable. □

Lemma 10 (restated). *If we have an E-computable polynomial family $\{f_n\}_{n\geq 1}$ of algebraic circuit complexity $2^{\Omega(n)}$, then either E$\not\subseteq$#P/poly or VNP has polynomials of algebraic circuit complexity $2^{\Omega(n)}$.*

Proof. Say, for a constant $\delta \geq 1$, we have an E-computable multi-δ-ic polynomial family $\{f_n\}_{n\geq 1}$ with algebraic circuit complexity $2^{\Omega(n)}$. Clearly, the coefficients in f_n have bitsize $2^{O(n)}$. By using a simple transformation, given in [KP09, Lem.3.9], we get a multilinear polynomial family $\{h_n\}_{n\geq 1}$, that is E-computable and has algebraic complexity $2^{\Omega(n)}$, such that its coefficients are $\{0, 1\}$.

Assume E$\subseteq$#P/poly. Since each coefficient of h_n is 0 or 1 that is computable in E, we deduce that the coefficient-function of h_n is in #P/poly. Thus, by [Bür13, Prop.2.20], $\{h_n\}_{n\geq 1}$ is in VNP and has algebraic circuit complexity $2^{\Omega(n)}$. □

C Proofs from Section 3– Subexp bootstrapping

Claim 14 (restated). *Let $f(s)$ be a function in $\Omega(s)$. Let ε be a function defined as $\varepsilon(s) := 10\lceil \log f(s) \rceil$. Suppose that we have a family $\mathcal{F} := \{q_{m,s} \mid s \in \mathbb{N},\, m = \varepsilon(s)\}$ of multi-δ-ic, where δ is a constant, m-variate polynomials such that $q_{m,s}$ is computable in poly$(f(s))$ time but has no size $< s^{1/2e_0}$ circuit.*

Then, for some constant c, we have a exp$(O(\varepsilon((sd)^c)^2/\log s))$-time computable variable reduction map, from n to $\lceil \varepsilon((sd)^c)^2/\log s \rceil$, that preserves nonzeroness for degree-d n-variate polynomials computed by size-s circuits. Furthermore, after variable reduction, the degree of the new polynomial will be poly(sd).

Proof. Idea is the same as in the proof of Lemma 9. Technical difference is due to the parameters of the Nisan-Wigderson design. Here we provide the details.

Let $\mathcal{P}$ be the set of degree-d polynomials computed by size-s circuits that depend only on first n variables. The number of variables $n \le s$. Let $\varepsilon' = \varepsilon'(s,d) := \lceil \varepsilon((sd)^c)^2/\log s \rceil$. Constant c will be fixed later. Next we describe how to reduce number of variables for every $P \in \mathcal{P}$, from n to ε'.

Let $s' := (sd)^c$. Let $\mathcal{D} := \{S_1, \ldots, S_n\}$ be an $(\varepsilon', \varepsilon(s'), 10\lceil \log s \rceil)$-design on the variable set $Z := \{z_1, \ldots, z_{\varepsilon'}\}$ (Lemma 8). Our hitting-set generator for $\mathcal{P}$ is defined as: for all $i \in [n]$, $x_i = q_{\varepsilon(s'),s'}(S_i) =: p_i$ on variables S_i. For p_i, we do not have circuits of size $< s'^{1/2e_0}$. Then, we show that for any nonzero polynomial $P(\mathbf{x}) \in \mathcal{P}$, $P(p_1, \ldots, p_n)$ is also nonzero.

For the sake of contradiction, assume that $P(p_1, \ldots, p_n) = 0$. Since $P(\mathbf{x})$ is nonzero, we can find the smallest $j \in [n]$ such that $P(p_1, \ldots, p_{j-1}, x_j, \ldots, x_n) =: P_1$ is nonzero, but $P_1\big|_{x_j=p_j}$ is zero. Thus, $(x_j - p_j)$ divides P_1. Let $\mathbf{a}$ be an assignment on all the variables in P_1, except x_j and the variables S_j in p_j, with the property: P_1 at $\mathbf{a}$ is nonzero. Since P_1 is nonzero, we can find such an assignment. Now our new polynomial P_2 on the variables S_j and x_j is of the form:

$$P_2(S_j, x_j) \;=\; P(p'_1, \ldots, p'_{j-1}, x_j, a_{j+1}, \ldots, a_n)$$

where, for each $i \in [j-1]$, p'_i is the polynomial on the variables $S_i \cap S_j$, and a_i's are field constants decided by our assignment $\mathbf{a}$. By the design, for each $i \in [j-1]$, $|S_i \cap S_j| \le 10\lceil \log s \rceil$. Since p'_is are polynomials on variables $S_i \cap S_j$ of individual-degree $< \delta$, each p'_i has a circuit (of trivial form $\Sigma\Pi$) of size at most $10\lceil \log s \rceil \delta \cdot \delta^{10\lceil \log s \rceil} \le s^{c_0}$ for some constant c_0. Then we have a circuit for P_2 of size at most $s_1 := s + n \cdot s^{c_0}$, and the degree of P_2 is at most $d_1 := nd \cdot 10\lceil \log s \rceil \delta$. Since $(x_j - p_j)$ divides P_2, we can invoke Kaltofen's factorization algorithm [Kal89] (see [Bür13, Thm.2.21] for the algebraic circuit complexity of factors) and get an algebraic circuit for p_j of size $(s_1 d_1)^{c_1}$, for some absolute constant c_1. Consequently, we have a circuit for p_j of size $\le (sd)^{c'_1}$ for some constant c'_1 (independent of c). On the other hand, p_j has no circuit of size $< (sd)^{c/2e_0}$. Pick c greater than $2e_0c'_1$. Then we get a contradiction. So, for $c > 2e_0c'_1$, $P(p_1, \ldots, p_n)$ remains nonzero.

The time for computing $(p_1, \ldots, p_n)$ depends on: **(1)** computing the design (i.e. $2^{O(\varepsilon')}$-time), and **(2)** computing p_i's (i.e. $2^{O(\varepsilon(s'))}$-time). Thus, the variable reduction map is computable in $\exp(O(\varepsilon'))$ time, as $\varepsilon(s') > \log f(s') > \log s$.

After variable reduction, the degree of the new polynomial will be $< dn \cdot \deg(q_{\varepsilon(s'),s'}) = (sd)^{O(1)}$. $\square$

D Proofs from Section 4– Constant variate

Lemma 15 (restated). *There exists an algorithm which takes (ℓ, n, d), with $\ell \ge 100$ and $d \ge 13$, and a base set S of size $\ell := \lceil 4n^2/d \rceil$ as input, and outputs an (ℓ, n, d)-design $\mathcal{D}$ having $\ge 2^{d/4}$ subsets, in time $O((4\ell/n)^n)$.*

Proof. Proof is similar to that of Lemma 8 (see Chap.16 [AB09]). We describe a greedy algorithm to construct $\mathcal{D}$.

$\mathcal{D} \leftarrow \emptyset$;
while $|\mathcal{D}| < 2^{d/4}$ **do**
  Find the first n-subset D of S such that $\forall I \in \mathcal{D}$, $|I \cap D| \le d$;
  $\mathcal{D} \leftarrow \mathcal{D} \cup \{D\}$;
end while

What is the running time? In each iteration, we go through all n-subsets of S and for every n-subset D, we have to check whether $\forall I \in \mathcal{D}$, $|I \cap D| \le d$. So, every iteration takes at most

$\binom{\ell}{n} \cdot 2^{d/4} \cdot \ell^2$ time. (Note– the `while` loop can run at most $2^{d/4}$ times.) So, the total running time is $(4\ell/n)^n$, for $\ell \geq 100, n > d \geq 13$. Next, we give the proof of correctness of the algorithm.

Using probabilistic method, we show that for $|\mathcal{D}| < 2^{d/4}$, we can always find an n-subset D of S such that $\forall I \in \mathcal{D}$, $|I \cap D| \leq d$. Let D be a random subset of S, constructed by the following procedure: For all $s \in S$, s will be in D with probability $2n/\ell$. So, $\mathrm{Exp}[|D|] = 2n$ and for each $I \in \mathcal{D}$, $\mathrm{Exp}[|I \cap D|] = 2n^2/\ell \leq d/2$. Using 'Chernoff bounds' [AB09, Thm.A.14] we have that:

$$\Pr[|D| < n] = \Pr\left[|D| < \left(1 - \frac{1}{2}\right) 2n\right] \leq \left(\frac{2}{e}\right)^n \text{ and}$$

$$\begin{aligned}
\Pr[|D \cap I| > d] &\leq \Pr\left[|D \cap I|| \geq \frac{4n^2}{\ell}\right] \quad \left(\because d \geq \frac{4n^2}{\ell}\right) \\
&\leq \left(\frac{e}{4}\right)^{\frac{2n^2}{\ell}} \quad \text{(applying Chernoff bound)} \\
&\leq \left(\frac{e}{4}\right)^{\frac{d}{2}(1-\frac{1}{\ell})} \quad \left(\because \ell \leq \frac{4n^2}{d} + 1\right).
\end{aligned}$$

Let E denote the event that $|D| < n$. For all $I \in \mathcal{D}$, let E_I denote the event that $|D \cap I| > d$. Then,

$$\begin{aligned}
\Pr[E \text{ or } \exists I \in \mathcal{D}, E_I] &\leq \Pr[E] + |\mathcal{D}| \cdot \Pr[E_I] \qquad \text{(Union bound)} \\
&\leq \left(\frac{2}{e}\right)^d + 2^{\frac{d}{4}} \left(\frac{e}{4}\right)^{\frac{d}{2}(1-\frac{1}{\ell})} \\
&< 1 \; (\because \ell > d \geq 13).
\end{aligned}$$

Thus, an n-subset D does exist and the algorithm will grow the collection $\mathcal{D}$. □

Proposition 16 (restated). *If, for some $i \geq 0$, $m_i \geq 192 f_i^2 \cdot \frac{1}{\epsilon} \log(128 f_i^2)$, then the same relation holds between m_{i+1} and f_{i+1}.*

Proof. From the definition,

$$\begin{aligned}
m_{i+1} &= 2^{\frac{m_i}{64 f_i^2}} \\
&\geq 2^{\frac{192 f_i^2 \cdot \frac{1}{\epsilon} \log(128 f_i^2)}{64 f_i^2}} \qquad \text{(from hypothesis)} \\
&= 2^{\frac{3}{\epsilon} \log(128 f_i^2)} \\
&= 2^{\frac{21}{\epsilon}} f_i^{\frac{6}{\epsilon}}
\end{aligned}$$

and $192 f_{i+1}^2 \cdot \frac{1}{\epsilon} \log(128 f_{i+1}^2) = 3 \times 2^{14} f_i^4 \cdot \frac{1}{\epsilon} \log(2^{15} f_i^4)$. Since $m_{i+1} \geq 2^{21/\epsilon} f_i^{6/\epsilon} > 2^{21} f_i^6 \geq 3 \times 2^{14} f_i^4 \cdot \frac{1}{\epsilon} \log(2^{15} f_i^4)$, we prove the required statement. □

Proposition 17 (restated). *Suppose that $m_0 \geq \max\{(8 f_0)^{\frac{2}{\epsilon}},\ 192 f_0^2 \cdot \frac{1}{\epsilon} \log(128 f_0^2)\}$. Then for all $i \geq 0$: 1) $m_{i+1} \geq 2^{m_i^{1-\epsilon}}$ and 2) $m_{i+1} \geq 2 m_i > 3456 f_i^2$.*

Proof. (1) Proving $m_{i+1} \geq 2^{m_i^{1-\epsilon}}$ is equivalent to proving $m_i \geq (8 f_i)^{\frac{2}{\epsilon}}$. So, at $i = 0$, the hypothesis implies that $m_1 \geq 2^{m_0^{1-\epsilon}}$.

For $i \geq 1$, to prove $m_i \geq (8f_i)^{\frac{2}{\epsilon}}$, it is sufficient to prove that $m_{i-1} \geq 64f_{i-1}^2 \cdot \frac{2}{\epsilon} \log(8 \times 16 f_{i-1}^2) = 128 f_{i-1}^2 \cdot \frac{1}{\epsilon} \log(128 f_{i-1}^2)$. The latter relation is true for $i = 1$. Using Proposition 16, we can claim that for all $i \geq 0$, $m_i \geq 192 f_i^2 \cdot \frac{1}{\epsilon} \log(128 f_i^2)$. Consequently, for all $i \geq 0$, $m_{i+1} \geq 2^{m_i^{1-\epsilon}}$.

(2) From $m_i \geq 192 f_i^2 \cdot \frac{1}{\epsilon} \log(128 f_i^2)$ we get that $m_i > 192 f_i^2 \times 9 = 1728 f_i^2$.

Proving $m_{i+1} \geq 2m_i$ is equivalent to proving $m_i \geq (8f_i)^2 \log(2m_i)$. Note that it holds at $i = 0$ because of the hypothesis on m_0 and since $m_0 / \log(2m_0)$ increases as m_0 increases.

For $i \geq 1$, proving $m_i \geq (8f_i)^2 \log(2m_i)$, is equivalent to proving that $2^{m_{i-1}/(64 f_{i-1}^2)} \geq (8 \times 16 f_{i-1}^2)^2 \left(1 + \frac{m_{i-1}}{64 f_{i-1}^2}\right)$. Equivalently, we need to show that

$$m_{i-1} \geq 64 f_{i-1}^2 \left(2 \log(128 f_{i-1}^2) + \log\left(1 + \frac{m_{i-1}}{64 f_{i-1}^2}\right)\right). \tag{2}$$

Note that *if* we substitute $m_{i-1} \mapsto \alpha := (128 f_{i-1}^2 - 1) 64 f_{i-1}^2$ then RHS becomes $192 f_{i-1}^2 \cdot \log(128 f_{i-1}^2) =: \beta$. Recall that, using the hypothesis, we have shown the claim: $m_{i-1} \geq 192 f_{i-1}^2 \cdot \frac{1}{\epsilon} \log(128 f_{i-1}^2) > \beta$.

Now, we consider three cases: (i) If $m_{i-1} < \alpha$ then in Eqn.2 RHS $\leq \beta$, in which case we are done by the claim. (ii) If $m_{i-1} = \alpha$ then the inequality in Eqn.2 holds as $\alpha > \beta$. (iii) If $m_{i-1} > \alpha$ then the inequality will continue to hold as the difference function LHS-RHS increases with m_{i-1}. ☐

Proposition 18 (restated). *The least i, for which $m_i \geq s$, is $\leq \frac{3}{1-\epsilon} \log\left(\frac{3}{1-\epsilon}\right) + 2\log^\star s$.*

Proof. By Proposition 17 we know that m_i more than doubles as $i \mapsto i+1$. Thus, at $i_0 := \frac{3}{1-\epsilon} \log\left(\frac{3}{1-\epsilon}\right)$, we have $m_{i_0} > 2^{i_0}$. Note that $2^{i_0(1-\epsilon)} > \frac{i_0}{1-\epsilon}$. Thus, $m_{i_0}^{1-\epsilon} > \frac{\log m_{i_0}}{1-\epsilon}$, and $2^{(1-\epsilon) m_{i_0}^{1-\epsilon}} > m_{i_0}$.

Also, from the same proposition we have that $m_{i+1} \geq 2^{m_i^{1-\epsilon}}$. Hence,

$$m_{i+2} \geq 2^{2^{(1-\epsilon)\cdot m_i^{1-\epsilon}}}.$$

Since, for $i \geq i_0$, we have $m_i \geq m_{i_0}$. The above inequality gives us $m_{i+2} \geq 2^{m_i}$. Consequently, beyond $i = i_0 + 2\log^\star s$, we have $m_i \geq s$. ☐

E Proofs from Section 5– Shallow depths

Theorem 20 (restated). *Suppose that for some constant $n \geq 2$ and some arbitrarily slow growing function $a(s)$, we have a $\left(poly(s), O(s^n/\log^2 s)\right)$-hsg for* multilinear *polynomials computed by size-s $\Sigma \wedge^a \Sigma\Pi$ circuits that depend only on the first $n \log s$ variables.*

Then, for N-variate d-degree size-s circuits, we have a poly(sd)-time nonzeroness preserving variable reducing polynomial map ($N \mapsto O(\log(sd))$) and Conjecture 1 holds. Furthermore, after variable reduction, the degree of the new polynomial will be poly(sd).

Proof. The proof is along the lines of [AV08, Theorem 3.2]. For all $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of multilinear polynomials computed by size-s $\Sigma \wedge^a \Sigma\Pi$ circuit that depend only on the first $n \log s$ variables. First, using the $\left(\text{poly}(s), O(s^n/\log^2 s)\right)$-hsg we show how to construct a family of multilinear polynomial which is E-computable but not computable by $2^{o(n)}$-size circuit. Using this hard polynomial family we get both the nonzeroness-preserving variable-reducing polynomial map and Conclusion 1. Invoking Lemma 9, in poly(sd) time, we can convert a

d-degree N-variate polynomial computed by a size-s circuit to $O(\log(sd))$-variate poly(sd)-degree polynomial. Conjecture 1 follows from the Lemma 10.

Now, we discuss how to construct that hard polynomial family from the hsg $\mathbf{f}(y)$ of $\mathcal{P}_s$. Let $m := n \log s$. Now we consider the annihilator of $\mathbf{f}(y)$ to get a hard polynomial. Let k be the number of m-variate $m/2$-degree multilinear monomials. Since $k = \binom{m}{m/2} \geq 2^m/\sqrt{2m} = s^n/\sqrt{2m} > O(s^n/\log^2 s)\cdot m$ (for large enough s), we get an m-variate and $m/2$-degree multilinear homogeneous annihilating polynomial $q_m \notin \mathcal{T}_s$ and computable in $s^{O(1)}$ time. The analysis is similar to Lemma 5. Importantly $q_m \notin \mathcal{P}_s$, so, no $\Sigma\wedge^a\Sigma\Pi$ circuit of size $< s = 2^{\Theta(m)}$ can compute it. Next we show that it is also not computable by any $2^{o(m)}$-size algebraic circuit.

For the sake of contradiction, assume that q_m has a $2^{o(m)}$-size circuit. From log-depth reduction result [Sap16, Chapter 5] we get a circuit C, of $d_m = \Theta(\log m)$-depth and $s_m = 2^{o(m)}$ size, with the additional properties:

1. alternative layers of addition/multiplication gates with the top-gate (root) being addition,
2. below each multiplication layer the related polynomial degree at least halves, and
3. fanin of each multiplication gate is at most 5.

Now we cut the circuit C at the t-th layer of multiplication gates from the top, where $t = t(d_m)$ will be fixed later, to get the following two parts:

Top part: the top part computes a polynomial of degree at most 5^t and the number of variables is at most s_m. So it can be reduced to a trivial $\Sigma\Pi$ circuit of size $\binom{s_m+5^t}{5^t} = s_m^{O(5^t)}$ (Stirling's approximation, see [Sap16, Proposition 4.4]).

Bottom part: in the bottom part, we can have at most s_m many top-multiplication gates that feed into the top part as input. Each multiplication gate computes a polynomial of degree at most $m/2 \cdot 2^{-t}$ and the number of variables is at most m. So each multiplication gate can be reduced to a trivial $\Sigma\Pi$ circuit of size $\binom{m+m/2^{t+1}}{m/2^{t+1}} = 2^{O(mt/2^t)}$.

From the above discussion, we have a $\Sigma\Pi^{5^t}\Sigma\Pi^{m/2^{t+1}}$ circuit C', computing q_m, that has size $s_m^{O(5^t)} + s_m \cdot 2^{O(mt/2^t)}$.

Now we fix t. The second summand becomes $2^{o(m)}$ if we pick $t = \omega(1)$ (recall that $s_m = 2^{o(m)}$). To get a similar upper bound on the first summand we need to pick $5^t \log s_m = o(m)$. Finally, we also want $5^t \leq a(s)$, to satisfy the fanin bound of the top multiplication layer. A function $t = t(d_m)$, satisfying the three conditions, exists as $\log s_m = o(m)$ and $a(\cdot)$ is an increasing function. Let us fix such a function t. (As C has super-constant depth, we can also assume that the cut at depth t will be possible.) Thus the circuit C', computing q_m, has size $s'_m = 2^{o(m)}$.

Relabel $a := 5^t$. The condition $5^t \log s_m = o(m)$ also ensures that $a = o(m)$. So now we have a shallow circuit for q_m of the form $\Sigma\Pi^a\Sigma\Pi$. Applying Lemma 11, we get a circuit computing q_m of the form $\Sigma\wedge^a\Sigma\Pi$ and size $s'_m \cdot 2^a = 2^{o(m)} \cdot 2^{o(m)} = 2^{o(m)}$ which is $< s$. This contradicts the hardness of q_m. Thus, there is no algebraic circuit for q_m of size $2^{o(m)}$.

Now we have a family of multilinear polynomials $\{q_m\}_{m\geq 1}$ such that is $2^{O(m)}$ time computable but has no $2^{o(m)}$ size circuit. $\square$