SIAM J. Comput.
Vol. 45, No. 4, pp. 1533–1562

JACOBIAN HITS CIRCUITS: HITTING SETS, LOWER BOUNDS FOR DEPTH-*D* OCCUR-*k* FORMULAS AND DEPTH-3 TRANSCENDENCE DEGREE-*k* CIRCUITS*

MANINDRA AGRAWAL†, CHANDAN SAHA‡, RAMPRASAD SAPTHARISHI§, AND NITIN SAXENA¶

Abstract. We present a single common tool to strictly subsume all known cases of polynomial time black box polynomial identity testing (PIT), that have been hitherto solved using diverse tools and techniques, over fields of zero or large characteristic. In particular, we show that polynomial (in the size of the circuit) time hitting-set generators for identity testing of the two seemingly different and well studied models—depth-3 circuits with bounded top fanin, and constant-depth constant-read multilinear formulas—can be constructed using one common algebraic-geometry theme: *Jacobian* captures algebraic independence. By exploiting the Jacobian, we design the first efficient hitting-set generators for broad generalizations of the above-mentioned models, namely, (a) depth-3 ($\Sigma\Pi\Sigma$) circuits with constant *transcendence degree* of the polynomials computed by the product gates (no bounded top fanin restriction), and (b) constant-depth constant-*occur* formulas (no multilinear restriction). Constant *occur* of a variable, as we define it, is a more general concept than constant read. Also, earlier work on the latter model assumed that the formula is multilinear. Thus, our work goes further beyond the related results obtained by Saxena and Seshadhri [*STOC*, ACM, New York, 2011, pp. 431–440], Saraf and Volkovich [*STOC*, ACM, New York, 2011, pp. 421–430], Anderson, van Melkebeek, and Volkovich, [*IEEE Conference on Computational Complexity*, IEEE, Piscataway, NJ, 2011, pp. 273–282], Beecken, Mittmann, and Saxena [*ICALP*, Springer, New York, 2011, pp. 134–148] and Grenet et al. [*Proceedings of the 30th Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, Schloss Dagstuhl–Liebniz–Zentrum für Informatik, Wadern, Germany, 2011, pp. 127–139] and brings them under one unifying technique. In addition, using the same Jacobian-based approach, we prove exponential lower bounds for the immanant (which includes permanent and determinant) on the same depth-3 and depth-4 models for which we give efficient PIT algorithms. Our results reinforce the intimate connection between identity testing and lower bounds by exhibiting a concrete mathematical tool—the Jacobian. The Jacobian is equally effective in solving both the problems on certain interesting and previously well-investigated (but not well understood) models of computation.

Key words. algebraic independence, black box, circuits, depth, identity testing, immanant, Jacobian, lower bound, Vandermonde

AMS subject classifications. 68W30, 68Q25

DOI. 10.1137/130910725

1. Introduction and examples.

A polynomial in many variables, when written down verbosely as a sum of monomials, might have a large expression. Arithmetic circuits, on the other hand, provide a succinct way to represent multivariate polynomials. An arithmetic circuit, consisting of addition (+) and multiplication (×) gates, takes several variables as input and computes a polynomial in those variables. The study of arithmetic circuits—as to which algorithmic questions on polynomials can

*Received by the editors February 25, 2013; accepted for publication (in revised form) November 16, 2015; published electronically August 30, 2016. A preliminary version appeared in *STOC*, ACM, New York, 2012, pp. 599–614.
http://www.siam.org/journals/sicomp/45-4/91072.html
†Indian Institute of Technology, Kanpur, India (manindra@iitk.ac.in). This author was supported by Humboldt Forschungspreis and J C Bose Fellowship.
‡Indian Institute of Science, Bangalore, India (chandan@csa.iisc.ernet.in). This author was supported by the IMPECS Fellowship.
§Chennai Mathematical Institute, Chennai, India (ramprasad@cmi.ac.in). This author was supported by IMPECS and MSR (India) Ph.D. Fellowship.
¶Indian Institute of Technology, Kanpur, India (nitin@cse.iitk.ac.in).

be resolved efficiently in this model of computation, and which polynomials do not admit any polynomial-sized circuit representation—forms the foundation of algebraic complexity theory.

One particular algorithmic question, the problem of *polynomial identity testing* (PIT), occupies a pivotal position in the theory of arithmetic circuit complexity. It is the problem of deciding if the output of a given arithmetic circuit is the identically zero polynomial. Being such an elementary problem, identity testing has enjoyed its status of prime importance by appearing in several fundamental results including primality testing [AKS04], the PCP theorem [ALM+98] and the IP = PSPACE result [LFKN90, Sha90], among many others like graph matching [Lov79, MVV87], polynomial interpolation [CDGK91], matrix completion [IKS10], polynomial solvability [KY08], factorization [SV10], learning of arithmetic circuits [KS06], and the geometric complexity theory approach [Mul12, Mul11]. What is more intriguing is that there is an intimate connection between identity testing and lower bounds [KI03, HS80, AvM10], especially the problem of separating the complexity classes VP from VNP (which must necessarily be shown before showing $\mathsf{P} \neq \mathsf{NP}$ [Val79, SV85]). Proving $\mathsf{VP} \neq \mathsf{VNP}$ amounts to showing that an explicit class of polynomials, like the permanent, cannot be represented by polynomial-sized arithmetic circuits, which in turn would follow if identity testing can be derandomized using a certain kind of pseudo random generator [Agr05, KI03]. (Note that identity testing has a simple and efficient randomized algorithm—pick a random point and evaluate the circuit at it [Sch80, Zip79, DL78].)

During the past decade, the quest for derandomization of PIT has yielded several results on restricted models of circuits. But, fortunately, the search has been made more focused by a line of work [GKKS13, Koi12, AV08, VSBR83] which states that a polynomial time *black box* derandomization of identity testing for depth-3 circuits (via a certain pseudorandom generator) implies a quasi-polynomial time derandomization of PIT for *polydegree*[1] circuits. By a polynomial time black box test for a circuit class $\mathcal{C}$, we mean the following.

- Construct a polynomial-sized list of points with small integer coordinates such that any nonzero circuit in $\mathcal{C}$ evaluates to a nonzero value on one of the points. (For characteristic $p > 0$, one works with a small field extension, where each coordinate of a point is an element of the extension field.)
- A Turing machine that runs in time polynomial in the parameters defining $\mathcal{C}$ (precisely, size of circuits in $\mathcal{C}$) and outputs such a list of points is also called a *polynomial time hitting-set generator* for $\mathcal{C}$.

With depth-3 as the final frontier, the results that have been achieved so far include polynomial time hitting-set generators for the following models:

- depth-2 ($\Sigma\Pi$) circuits (equivalently, the class of *sparse* polynomials) [KS01];
- depth-3 ($\Sigma\Pi\Sigma$) circuits with constant top fanin [SS11];
- constant-depth constant-read multilinear formulas [AvMV11, SV11] (and their sparse-substituted variants);
- circuits *generated* by sparse polynomials with constant transcendence degree [BMS11].

To our knowledge, these are the only instances for which *polynomial* time hitting-set generators are known. The result on depth-3 bounded top fanin circuits is based upon the Chinese remaindering technique of [KS07] and the ideal-theory framework studied in [SS10]. Their work followed after a sequence of developments in rank

[1]Circuits computing polynomials with degree bounded by a polynomial function in the size of the circuit.

bound estimates [DS05, KS08, SS09, KS09b, SS10], some using incidence geometry—although, this result [SS11] in particular is not rank based. On the other hand, the work on constant-depth multilinear formulas [AvMV11, SV11] is obtained by building upon and extending the techniques of other earlier results [KMSV10, SV09, SV08] on "read-once" models. At a high level, this involved a study of the structure of multilinear formulas under the application of partial derivatives with respect to a carefully chosen set of variables and invoking depth-3 rank bounds (survey [SY10]). More recently, a third technique has emerged in [BMS11] which is based upon the abstract concept of *algebraic independence* of polynomials and a related computational handle called the *Jacobian*. They showed that for any given polydegree circuit C and sparse polynomials $f_1, \ldots, f_m$ with constant transcendence degree, a hitting set for $C(f_1, \ldots, f_m)$ can be constructed in polynomial time.

Our contribution. With these diverse techniques floating around the study of hitting-set generators, one wonders, could there be one single tool that is sufficiently powerful to capture all these models? Is there a common feature in these different models that can be used to construct unified PITs? The answer to both these questions, as we show in this work, is *yes*. The key to this lies in studying the properties of the Jacobian, a mathematical object lying at the very core of algebraic geometry. As for the "unique feature," notice that in the above four models some "parameter" of the circuit is "bounded"—be it bounded top fanin, bounded read of variables, or bounded transcendence degree. (Bounded depth should not be seen as an extra restriction on the circuit model because of [GKKS13, AV08]). At an intuitive level, it seems to us that it is this "bounded parameter"-ness of the circuit that makes the Jacobian perform at its best.

In the process of finding a universal technique, we significantly strengthen the earlier results. We construct hitting-set generators not only for depth-3 circuits with bounded top fanin, but also for circuits of the form $C(T_1, \ldots, T_m)$, where C is a polydegree circuit and $T_1, \ldots, T_m$ are products, of linear polynomials, with bounded transcendence degree. In case of depth-3 circuits, $C(T_1, \ldots, T_m)$ is simply $T_1 + \cdots + T_m$. Further, we remove the restriction of multilinearity totally from the constant-depth constant-read model and construct the first efficient hitting-set generator for this class. The condition of constant read is also replaced by the more general notion of constant *occur*.

At this point, one is faced with a natural question, how effective is the Jacobian in proving lower bounds? The intimate connection between efficient algorithms and lower bounds has recurrently appeared in various contexts [Wil11, Rag08, Uma03, PSZ00, IW97]. For arithmetic circuits, this link is provably tight [KI03, Agr05, AV08]: Derandomizing identity testing is equivalent to proving circuit lower bounds. This means, one might have to look for techniques that are powerful enough to handle the dual worlds of algorithm design and lower bounds with equal effectiveness—e.g., the *partial derivative technique* has been used to prove lower bounds and identity testing (albeit non-black box) on restricted models (survey [CKW11]); the τ*-conjecture* is another such example [GKPS11]. In this work, we demonstrate the utility of the Jacobian in proving exponential lower bounds for the immanant (which includes determinant and permanent) on the same depth-3 and depth-4 models for which we give efficient PIT algorithms. In particular, this includes depth-4 constant-occur formulas, depth-4 circuits with constant transcendence degree of the underlying sparse polynomials (which significantly generalizes the lower bound result in [GKPS11]), and depth-3 circuits with constant transcendence degree of the polynomials computed by the product gates. To our knowledge, all these lower bounds are new and it is not

TABLE 1
Comparison with the earlier efficient hitting sets.

Previous best		This paper		
Model	*Running time*[1]	*Extended Model*[2]	*Running time*[1]	*Imm_n lower bound*
$\Sigma\Pi\Sigma(k)$ circuits: $T_1 + \cdots + T_k \overset{?}{=} 0$	s^k [SS11]	$C(T_1, \ldots, T_m) \overset{?}{=} 0$ polydegree C and $\mathrm{trdeg}\,\{T_i\} \leq k$	s^k	$\mathrm{trdeg}\,\{T_i\} = \Omega(n)$
$\Sigma\Pi\Sigma\Pi(k)$ `multilinear` circuits	s^{k^3} [SV11]	$\Sigma\Pi\Sigma\Pi$ occur-k formulas	s^{k^2}	$s = 2^{\Omega(n/k^2)}$
depth-D, read-k `multilinear` formulas	s^R where $R = k^{k^2} + kD$ [AvMV11]	depth-D, occur-k formulas	s^R where $R = k^{2^D}$	$s = 2^{\Omega(n)}$ for constant k, D assuming Conjecture 6.1
$C(f_1, \ldots, f_m) \overset{?}{=} 0$ polydegree C, $\Sigma\Pi$ circuits f_i's & $\mathrm{trdeg}\,\{f_i\} \leq k$	s^k [BMS11]	–	–	$s = 2^{\Omega(n/k)}$

[1]Estimates the bit complexity of the hitting-set generator; constant factors not stressed (also in higher exponents).
[2]We assume a zero or large characteristic.

known how to prove them using earlier techniques. A summary of the results in this paper is provided in Table 1.

Remark. The algorithms of [SS11], [SV11], and [AvMV11] work over any field, whereas ours work under the assumption that the field characteristic is zero or large. Also, [AvMV11] presents a quasi-polynomial time algorithm for *arbitrary* depth, constant read, multilinear formula—a result which we do not know yet how to capture using our technique. Two other quasi-polynomial hitting sets that our work does not capture are the hitting sets for constant-depth set-multilinear circuits [ASS13] and read-once oblivious algebraic branching programs [FS13].

1.1. A tale of two PITs (and three lower bounds). A set of polynomials $\mathbf{f} = \{f_1, \ldots, f_m\} \subset \mathbb{F}[x_1, \ldots, x_n]$ (in short, $\mathbb{F}[\mathbf{x}]$) is said to be *algebraically independent* over $\mathbb{F}$ if there is no nonzero polynomial $H \in \mathbb{F}[y_1, \ldots, y_m]$ such that $H(f_1, \ldots, f_m)$ is identically zero. A maximal subset of $\mathbf{f}$ that is algebraically independent is a *transcendence basis* of $\mathbf{f}$ and the size of such a basis is the *transcendence degree*[2] of $\mathbf{f}$ (denoted $\mathrm{trdeg}_{\mathbb{F}}\, \mathbf{f}$). Our first theorem states the following.

THEOREM 1.1. *Let C be a polydegree circuit of size s and each of $T_1, \ldots, T_m$ be a product of d linear polynomials in $\mathbb{F}[x_1, \ldots, x_n]$ such that* $\mathrm{trdeg}_{\mathbb{F}}\, \{T_1, \ldots, T_m\} \leq r$. *A hitting set for such $C(T_1, \ldots, T_m)$ can be constructed in time polynomial in n and $(sd)^r$, assuming $char(\mathbb{F}) = 0$ or $> d^r$.*

If C is a single $+$ gate, we get a hitting-set generator for depth-3 circuits with constant *transcendence degree* of the polynomials computed by the product gates (there is *no* restriction on top fanin).

Our second result uses the following generalization of *read-k* formulas (where every variable appears in at most k leaf nodes of the formula) to *occur-k* formulas. Two reasons behind this generalization are one, to accommodate the power of exponentiation—if we take the eth power of a read-k formula using a product gate, the "read" of the resulting formula goes up to ek—we would like to avoid this superfluous blow up in read. Two, a read-k formula has size $O(kn)$, which severely hinders its power of computation—for instance, determinant and permanent cannot even be expressed in this model when k is a constant [Kal85]. This calls for the following definition.

[2]Since algebraic independence satisfies the matroid property (cf. [Oxl92]), transcendence degree is well-defined.

DEFINITION 1.2. *An* occur-k formula *is a rooted tree with internal nodes labeled by* $+$ *and* $\times\curlywedge$ *(power-product gate). A* $\times\curlywedge$ *gate, on inputs* $g_1, \ldots, g_m$ *with incoming edges labeled* $e_1, \ldots, e_m \in \mathbb{N}$*, computes* $g_1^{e_1} \cdots g_m^{e_m}$*. At the leaves of this tree are depth-2 formulas computing sparse polynomials (*leaf nodes*), where every variable occurs in at most* k *of these sparse polynomials.*

The size of a $\times\curlywedge$ gate is defined as the integer $(e_1 + \cdots + e_m)$ associated with its incoming edges, while the size of a $+$ gate is counted as one. The size of a leaf node is the size of the corresponding depth-2 formula. With these conventions, the *size* of an occur-k formula is defined to be the total size of all its gates (and leaf nodes) plus the number of edges. Note that any polynomial can be expressed as an occur-1 formula, albeit of exponential size.

Depth is defined to be the number of layers of $+$ and $\times\curlywedge$ gates plus 2 (the "plus 2" accounts for the depth-2 formulas at the leaves). Thus, occur-k is more relaxed than the traditional read-k as it packs the "power of powering" (to borrow from [GKPS11]), and the leaves are sparse polynomials (at most kn many) whose dependence on its variables is arbitrary; e.g., $(x_1^3x_2 + x_1^2x_3^2 + x_1x_4)^e$ is *not* read-1 but is trivially depth-3 occur-1.

THEOREM 1.3. *A hitting set for any depth-D occur-k formula of size s can be constructed in time polynomial in s^R, where $R = (2k)^{2D\cdot 2^D}$ (assuming char$(\mathbb{F}) = 0$ or $> s^R$).*

A tighter analysis for depth-4 occur-k formulas yields a better time complexity. Note that a depth-4 occur-k formula allows unbounded top fanin. Also, it can be easily seen to subsume $\Sigma\Pi\Sigma\Pi(k)$ multilinear formulas studied by [SV11, KMSV10]. This is because any multilinear $\Sigma\Pi\Sigma\Pi(k)$ formula is a sum of k products of sparse polynomials and every variable appears in at most one of the sparse polynomials in every such product.

THEOREM 1.4. *A hitting set for any depth-4 occur-k formula of size s can be constructed in time polynomial in s^{k^2} (assuming char$(\mathbb{F}) = 0$ or $> s^{4k}$).*

For constant depth, the above theorems not only remove the restriction of multilinearity (and relax read-k to occur-k), but further improve upon the time complexity of [AvMV11] and [SV11]. The hitting-set generator of [AvMV11] works in time $n^{k^{O(k^2)}+O(kD)}$, and hence is superexponential when $k = \Omega(s^{\varepsilon/2D\cdot 2^D})$ for any positive $\varepsilon < 1$ and a constant D, whereas the generator in Theorem 1.3 runs in subexponential time for the same choice of parameters. The running time of [SV11] is $s^{O(k^3)}$, which is slightly worse than that of Theorem 1.4.

Since any polynomial has an exponential-sized depth-2, occur-1 formula (just the sparse representation), proving lower bounds on this model is an interesting proposition in its own right.

DEFINITION 1.5 (see [LR34]). *Let S_n denote the permutation group on n points and $\mathbb{C}^\times$ be the nonzero complex numbers. For any map $\chi : S_n \to \mathbb{C}^\times$, the* immanant *of a matrix $M = (x_{ij})_{n\times n}$ with respect to χ is defined as* $\mathrm{Imm}_\chi(M) = \sum_{\sigma\in S_n} \chi(\sigma) \prod_{i=1}^n x_{i\sigma(i)}$.

Determinant and permanent are special cases of the immanant with χ as the alternating sign character and the identity character, respectively. Denote $\mathrm{Imm}_\chi(M)$ by Imm_n for an arbitrarily fixed χ.

THEOREM 1.6. *Any depth-4 occur-k formula that computes Imm_n must have size $s = 2^{\Omega(n/k^2)}$ over any field of characteristic zero (even counting each $\times\curlywedge$ gate as size one).*

Thus, if each variable occurs in at most $n^{1/2-\varepsilon}$ $(0 < \varepsilon < 1/2)$ many underlying sparse polynomials, it takes an exponential-sized depth-4 circuit to compute Imm_n. Our next result is an exponential lower bound on the model for which the hitting set was developed in [BMS11] (but no lower bound was shown). It is also an improvement over the result obtained in [GKPS11] which holds only for more restricted depth-4 circuits over reals.

THEOREM 1.7. *Let C be any circuit. Let $f_1, \dots, f_m$ be sparse polynomials (of any degree) with sparsity bounded by s and their* trdeg *bounded by r. If $C(f_1, \dots, f_m)$ computes Imm_n, then $s = 2^{\Omega(n/r)}$ over any field of characteristic zero.*

Which means, any circuit involving fewer than $n^{1-\varepsilon}$ $\Sigma\Pi$-polynomials at the last levels, must have exponential size to compute Imm_n. (The models of Theorems 1.6 and 1.7 are incomparable.) The next result is on the model for which the hitting set is given by Theorem 1.1.

THEOREM 1.8. *Let C be any circuit and $T_1, \dots, T_m$ be products of linear polynomials. If $C(T_1, \dots, T_m)$ computes Imm_n, then* $\mathrm{trdeg}_{\mathbb{F}} \{T_1, \dots, T_m\} = \Omega(n)$ *over any field of characteristic zero.*

Which means, any circuit involving only $o(n)$ many algebraically independent $\Pi\Sigma$-polynomials at the last levels *cannot* compute Imm_n.

A related lower bound is one by Shpilka and Wigderson [SW02] who showed that any depth-3 circuit computing Det_n (the determinant of an $n \times n$ symbolic matrix) requires top fanin $\Omega(n^2)$ and size $\Omega(\frac{n^4}{\log n})$. Theorem 1.8 implies a top fanin lower bound of $\Omega(n)$ and size lower bound of $\Omega(n^2)$ for depth-3 circuits computing Det_n. In this regard, Shpilka and Wigderson's result is stronger than the above theorem. On the other hand, Theorem 1.8 states that Det_n cannot be computed by a depth-3 circuit with a large (possibly $\omega(n^2)$) number of product gates $T_1, \dots, T_m$ whose transcendence degree is $o(n)$. In this sense, the theorem says something stronger than a top fanin lower bound.

1.2. Deterministic testing of algebraic independence. The construction of hitting-set generators (stated in the previous section) also implies deterministic algorithms for certain special cases of the following problem: Given a set of polynomials as arithmetic circuits, check deterministically if they are algebraically independent. In fact, for these special cases we only require a black box access to the input circuits.[3] In this respect, it can be said that our hitting-set generators (and to some extent the lower bounds) exist because these independence testers exist.

The proof of Theorem 1.1 yields the following tester.

THEOREM 1.9. *Given black box access to polynomials $T_1, \dots, T_r$ that are products of d linear polynomials in $\mathbb{F}[x_1, \dots, x_n]$, there is a poly$((nd)^r)$-time algorithm to test whether they are algebraically independent, assuming char$(\mathbb{F}) = 0$ or $> d^r$.*

Similarly, the proof of Theorem 1.3 yields the following tester.

[3]The classical (Jacobian-based) efficient algorithm for testing the algebraic independence of general circuits, over large or zero characteristic, is *randomized and whitebox*.

THEOREM 1.10. *Let $T_1, \ldots, T_r$ be n-variate degree d polynomials computed by depth-D occur-k formulas of size s and presented as black boxes. There is an* $(sdn)^R$*-time algorithm, where* $R = r \cdot (2k)^{2D \cdot 2^D}$*, to test whether* $\{T_1, \ldots, T_r\}$ *are algebraically independent, assuming* $char(\mathbb{F}) = 0$ *or* $> s^R$.

1.3. Our ideas. The exact reasons why our techniques work, where older ones failed, are extremely technical. However, we now give the motivating, but imprecise, ideas. To a set of products of sparse polynomials $\{T_1, \ldots, T_m\}$ we associate a polynomial—the Jacobian $J(T_1, \ldots, T_r)$. It captures the algebraic independence of $T_1, \ldots, T_r$ (assuming this to be a transcendence basis of the T_i's). If we could find an r-variate linear map φ that keeps $\varphi \circ J(T_1, \ldots, T_r)$ nonzero, then $\varphi(T_1), \ldots, \varphi(T_r)$ are again algebraically independent and it can be shown that for *any* C: $C(T_1, \ldots, T_m) = 0$ iff $C(\varphi(T_1), \ldots, \varphi(T_m)) = 0$. Since the T_i's are not sparse, the Jacobian is usually a difficult polynomial to work with, and so is finding φ. However, for the special models in this paper we are able to design φ—mainly because the Jacobian (being defined via partial derivatives) has a nice "linearizing effect," on the circuit product gates, that factors itself. The map φ ultimately provides a hitting set for $C(T_1, \ldots, T_m)$, as we reduce to a PIT of a polynomial over "few" (roughly equal to r) variables.

The initial idea for lower bounds is similar. Suppose $\mathrm{Imm}_n = C(T_1, \ldots, T_m)$. Then, by algebraic dependence, $J(\mathrm{Imm}_n, T_1, \ldots, T_r) = 0$. Our proofs then exploit the nature of this identity for the special models. This part requires proving certain combinatorial properties of the immanant.

Remark. The dependence of our results on the field characteristic is because the Jacobian criterion, which involves taking derivatives, is used to characterize algebraic independence. We believe that the condition on the field characteristic in our results is probably not a fundamental requirement—rather, lifting this condition is perhaps a technical hurdle due to the lack of a suitable criterion that captures algebraic independence for low characteristic fields. (Intuitively, a low characteristic enables more "cancellations," polynomial identities, and configurations [SS10].) Recently, this direction of research was investigated in [MSS14], where a new criterion for algebraic independence, namely, the *Witt–Jacobian*, is presented that works even for small characteristic. Applying this new criterion it is shown in [MSS14] that the problem of testing algebraic independence is in $\mathsf{NP}^{\#\mathsf{P}}$. However, it seems that this criterion is not yet effective enough to be applied to our problem and lift the restriction on field characteristic from our results.

2. Preliminaries: Jacobian and faithful homomorphisms. Our contribution, in this section, is an elementary proof of Theorem 2.4, which was originally proved in [BMS11] using Krull's *hauptidealsatz*. Here, we state the main properties of the Jacobian and faithful homomorphisms without proofs—for details, refer to [BMS13].

DEFINITION 2.1. *The* Jacobian *of a set of polynomials* $\mathbf{f} = \{f_1, \ldots, f_m\}$ *in* $\mathbb{F}[x_1, \ldots, x_n]$ *is defined to be the matrix* $\mathcal{J}_{\mathbf{x}}(\mathbf{f}) = (\partial_{x_j} f_i)_{m \times n}$*, where* $\partial_{x_j} f_i = \partial f_i / \partial x_j$*. Let* $S \subseteq \mathbf{x} = \{x_1, \ldots, x_n\}$ *and* $|S| = m$*. Then* $J_S(\mathbf{f})$ *denotes the minor (i.e., determinant of the submatrix) of* $\mathcal{J}_{\mathbf{x}}(\mathbf{f})$ *formed by the columns corresponding to the variables in* S.

FACT 2.2 (Jacobian criterion). *Let* $\mathbf{f} \subset \mathbb{F}[\mathbf{x}]$ *be a finite set of polynomials of degree at most d, and* $\mathrm{trdeg}_{\mathbb{F}} \mathbf{f} \leq r$*. If* $char(\mathbb{F}) = 0$ *or* $char(\mathbb{F}) > d^r$*, then* $\mathrm{trdeg}_{\mathbb{F}} \mathbf{f} = \mathrm{rank}_{\mathbb{F}(\mathbf{x})} \mathcal{J}_{\mathbf{x}}(\mathbf{f})$.

The proof of this fact may be found in [BMS13].

Definition 2.3. *A homomorphism* $\Phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{y}]$ *($\mathbf{y}$ is another set of variables) is said to be* faithful *for a finite set of polynomials* $\mathbf{f} \subset \mathbb{F}[\mathbf{x}]$ *if* $\operatorname{trdeg}_{\mathbb{F}} \mathbf{f} = \operatorname{trdeg}_{\mathbb{F}} \Phi(\mathbf{f})$.

Theorem 2.4 (faithful is useful). *Let Φ be a homomorphism faithful for $\mathbf{f} = \{f_1, \ldots, f_m\} \subset \mathbb{F}[\mathbf{x}]$. Then for any $C \in \mathbb{F}[y_1, \ldots, y_m]$, $C(\mathbf{f}) = 0 \Leftrightarrow C(\Phi(\mathbf{f})) = 0$.*

Proof. It is trivial to see that $C(\mathbf{f}) = 0 \Rightarrow C(\Phi(\mathbf{f})) = 0$. Since Φ is faithful for $\mathbf{f}$, there is a transcendence basis (say, $f_1, \ldots, f_s$) of $\mathbf{f}$ such that $\Phi(f_1), \ldots, \Phi(f_s)$ is a transcendence basis of $\Phi(\mathbf{f})$. Since $\{f_1, \ldots, f_s\}$ are algebraically independent, the field $\mathbb{F}(f_1, \ldots, f_s)$ is isomorphic to $\mathbb{F}(y_1, \ldots, y_s)$. Further, since every other f_i is algebraically dependent on $\{f_1, \ldots, f_s\}$, it is also algebraic over $\mathbb{F}(f_1, \ldots, f_s)$. Hence,

$$\mathbb{F}(f_1, \ldots, f_m) \equiv (\mathbb{F}(f_1, \ldots, f_s))\,(f_{s+1}, \ldots, f_m) \equiv (\mathbb{F}(f_1, \ldots, f_s))\,[f_{s+1}, \ldots, f_m].$$

In other words, the elements of the field $\mathbb{K} = \mathbb{F}(\mathbf{f})$ can be written as polynomials in $f_{s+1}, \ldots, f_m$ with coefficients from $\mathbb{F}(f_1, \ldots, f_s)$. Suppose $C(\mathbf{f})$ is a nonzero element of $\mathbb{K}$, then there is an inverse $Q \in \mathbb{K}$ such that $Q \cdot C(\mathbf{f}) = 1$. Since Q is a polynomial in $f_{s+1}, \ldots, f_m$ with coefficients from $\mathbb{F}(f_1, \ldots, f_s)$, by clearing off the denominators of these coefficients in Q, we get an equation $\tilde{Q} \cdot C(\mathbf{f}) = P(f_1, \ldots, f_s)$, where $\tilde{Q}$ is a nonzero polynomial in $\mathbf{f}$ and P is a nonzero polynomial in $f_1, \ldots, f_s$. Applying Φ to both sides of the equation, we conclude that $C(\Phi(\mathbf{f})) = \Phi(C(\mathbf{f})) \neq 0$, otherwise $P(\Phi(f_1), \ldots, \Phi(f_s)) = \Phi(P(f_1, \ldots, f_s)) = 0$ which is not possible as $\Phi(f_1), \ldots, \Phi(f_s)$ are algebraically independent and P is a nontrivial polynomial. ∎

Recipe for faithful homomorphisms. All the PITs in this paper proceed by constructing faithful homomorphisms for a certain set of polynomials. The following fact describes the changes in the Jacobian after a "change of variables."

Fact 2.5 (chain rule). *For any finite set of polynomials $\mathbf{f} \subset \mathbb{F}[\mathbf{x}]$ and a homomorphism $\Phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{y}]$, we have $\mathcal{J}_{\mathbf{y}}(\Phi(\mathbf{f})) = \Phi\left(\mathcal{J}_{\mathbf{x}}(\mathbf{f})\right) \cdot \mathcal{J}_{\mathbf{y}}(\Phi(\mathbf{x}))$ (where Φ applied to a matrix/set refers to the matrix obtained by applying Φ to every entry).*

Proof. Follows directly from the chain rule of differentiation. ∎

The recipe for faithful homomorphisms uses the following "rank preserving linear maps" studied by Gabizon and Raz [GR05, Theorem 5].

Lemma 2.6 (Theorem 5 of [GR05]). *Let A be an $r \times n$ matrix with entries in a field $\mathbb{F}$, and let t be an indeterminate. Then, $\operatorname{rank}_{\mathbb{F}(t)}(A \cdot (t^{ij})_{i\in[n], j\in[r]}) = \operatorname{rank}_{\mathbb{F}} A$.*

Lemma 2.7 (recipe for faithful maps). *Let $\mathbf{f} \subset \mathbb{F}[\mathbf{x}]$ be a finite set of polynomials of degree at most d, $\operatorname{trdeg}_{\mathbb{F}} \mathbf{f} \leq r$, and char$(\mathbb{F}) = 0$ or $> d^r$. Let $\Psi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}]$ be a homomorphism such that $\operatorname{rank}_{\mathbb{F}(\mathbf{x})} \mathcal{J}_{\mathbf{x}}(\mathbf{f}) = \operatorname{rank}_{\mathbb{F}(\mathbf{z})} \Psi(\mathcal{J}_{\mathbf{x}}(\mathbf{f}))$.*

Then, the map $\Phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}, t, y_1, \ldots, y_r]$ that maps, for all i, $x_i \mapsto (\sum_{j=1}^{r} y_j t^{ij}) + \Psi(x_i)$ is a homomorphism faithful for $\mathbf{f}$.

We stress that in the above lemma all we require from Ψ is that the rank of $\mathcal{J}_{\mathbf{x}}(\mathbf{f})$ equals the rank of $\Psi(\mathcal{J}_{\mathbf{x}}(\mathbf{f}))$, which is just Ψ applied on each entry of $\Psi(\mathcal{J}_{\mathbf{x}}(\mathbf{f}))$. Note that $\Psi(\mathcal{J}_{\mathbf{x}}(\mathbf{f}))$ is *not* the Jacobian of $\Psi(\mathbf{f})$, and hence Ψ is not necessarily faithful for $\mathbf{f}$. Here Ψ is just a map that preserves the nonzeroness of some minor of $\mathcal{J}_{\mathbf{x}}(\mathbf{f})$ and hence Ψ could in principle be a scalar map. Of course, a scalar map can never be faithful to any nontrivial set of polynomials $\mathbf{f}$. Nevertheless, the above recipe allows us to take any such map Ψ and modify it to a map Φ that is faithful for $\mathbf{f}$.

Proof. Without loss of generality, let $\operatorname{trdeg}_{\mathbb{F}} \mathbf{f} = r$, which then (by the Jacobian criterion) is the rank of $\mathcal{J}_{\mathbf{x}}(\mathbf{f})$. We show that the matrix $\mathcal{J}_{\mathbf{y}}(\Phi(\mathbf{f}))$ is of rank r, which would imply (by the Jacobian criterion) that $\operatorname{trdeg}_{\mathbb{F}(t,\mathbf{z})} \Phi(\mathbf{f}) = r$. Note that if

$\mathrm{trdeg}_{\mathbb{F}(t,\mathbf{z})}\,\Phi(\mathbf{f}) = r$, then we also have $\mathrm{trdeg}_{\mathbb{F}}\,\Phi(\mathbf{f}) \geq r$. Since we know $\mathrm{trdeg}_{\mathbb{F}}(\mathbf{f}) = r$, this would force $\mathrm{trdeg}_{\mathbb{F}}\,\Phi(\mathbf{f}) = r$ as well.

Consider the projection $\mathcal{J}'$ of $\mathcal{J}_{\mathbf{y}}(\Phi(\mathbf{f}))$ obtained by setting $y_1 = \cdots = y_r = 0$.

$$\begin{aligned}\mathcal{J}' &= [\mathcal{J}_{\mathbf{y}}(\Phi(\mathbf{f}))]_{\mathbf{y}=\mathbf{0}}\\ &= [\Phi(\mathcal{J}_{\mathbf{x}}(\mathbf{f}))\cdot\mathcal{J}_{\mathbf{y}}(\Phi(\mathbf{x}))]_{\mathbf{y}=\mathbf{0}}\\ &= \Psi(\mathcal{J}_{\mathbf{x}}(\mathbf{f}))\cdot\mathcal{J}_{\mathbf{y}}(\Phi(\mathbf{x})),\end{aligned}$$

where the second line follows simply by Fact 2.5, and the third line uses the fact that $\Phi(x_i)$ is linear in $\mathbf{y}$.

Observe that the matrix $\mathcal{J}_{\mathbf{y}}(\Phi(\mathbf{x}))$ is exactly the Vandermonde matrix that is present in Lemma 2.6. Also, $\Psi(\mathcal{J}_{\mathbf{x}}(\mathbf{f}))$ has entries in $\mathbb{F}(\mathbf{z})$, and by assumption has the same rank as $\mathcal{J}_{\mathbf{x}}(\mathbf{f})$. Hence, by Lemma 2.6,

$$\begin{aligned}\mathrm{rank}_{\mathbb{F}(t,\mathbf{z})}\mathcal{J}' &= \mathrm{rank}_{\mathbb{F}(t,\mathbf{z})}\left(\Psi(\mathcal{J}_{\mathbf{x}}(\mathbf{f}))\cdot\mathcal{J}_{\mathbf{y}}(\Phi(\mathbf{x}))\right)\\ &= \mathrm{rank}_{\mathbb{F}(\mathbf{z})}\Psi(\mathcal{J}_{\mathbf{x}}(\mathbf{f})) = r.\end{aligned}$$

And since $\mathcal{J}'$ is just a projection of $\mathcal{J}_{\mathbf{y}}(\Phi(\mathbf{f}))$, the rank of the latter must also be r. Hence, Φ is indeed faithful. $\square$

3. Hitting set for depth-3 circuits of constant transcendence degree. Let C be any circuit and D be the circuit $C(T_1, \ldots, T_m)$, where each T_i is of the form $\prod_{j=1}^{d} \ell_{ij}$, every ℓ_{ij} is a linear polynomial in $\mathbb{F}[x_1, \ldots, x_n]$. For simplicity, assume without loss of generality that all the ℓ_i's are monic with respect to the lexicographic ordering $x_1 \succ \ldots \succ x_n$. Denote by $\mathbf{T}$ the set $\{T_1, \ldots, T_m\}$ and by $L(T_i)$ the multiset of linear polynomials that constitute T_i. Suppose $\mathrm{trdeg}_{\mathbb{F}}\,\mathbf{T} = k$ and $\mathbf{T}_k = \{T_1, \ldots, T_k\}$ is a transcendence basis of $\mathbf{T}$.

Since $\mathcal{J}_{\mathbf{x}}(\mathbf{T}_k)$ has full rank ($\mathrm{char}(\mathbb{F}) = 0$ or $\mathrm{char}(\mathbb{F}) > d^k$), without loss of generality assume that the columns corresponding to $\mathbf{x}_k = \{x_1, \ldots, x_k\}$ form a nonzero $k \times k$ minor of $\mathcal{J}_{\mathbf{x}}(\mathbf{T}_k)$. By Lemma 2.7, if we construct a $\Psi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}]$ that keeps $J_{\mathbf{x}_k}(\mathbf{T}_k)$ nonzero, then Ψ can easily be extended to a homomorphism $\Phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[\mathbf{z}, t, y_1, \ldots, y_k]$ that is faithful for $\mathbf{T}$. And hence, by Theorem 2.4, it would follow that $\Phi(D) = 0$ iff $D = 0$.

The linearity of the determinant would allow us to express $J_{\mathbf{x}_k}(\mathbf{T}_k)$ as a depth-3 circuit.

FACT 3.1. *For any set of vectors $v_{11}, \ldots, v_{kn} \in \mathbb{F}^n$,*

$$det\left[\sum_{i=1}^{k} v_{1i}, \ldots, \sum_{i=1}^{k} v_{ni}\right] = \sum_{1 \leq i_1, \ldots, i_n \leq k} det\left[v_{1i_1}, \ldots, v_{ni_n}\right].$$

Note that if $T_i = \prod_{j=1}^{d} \ell_{ij}$, then

$$\partial_x T_i = T_i \cdot \left(\sum_{j=1}^{d} \frac{\partial_x \ell_{ij}}{\ell_{ij}}\right).$$

Using this with the linearity of the determinant, $J_{\mathbf{x}_k}(\mathbf{T}_k)$ takes the following form,

$$\begin{aligned} J_{\mathbf{x}_k}(\mathbf{T}_k) &= \sum_{\ell_1 \in L(T_1),\dots,\ell_k \in L(T_k)} \frac{T_1 \cdots T_k}{\ell_1 \cdots \ell_k} \cdot \det \begin{bmatrix} \partial_{x_1}\ell_1 & \dots & \partial_{x_k}\ell_1 \\ \vdots & \ddots & \vdots \\ \partial_{x_1}\ell_k & \dots & \partial_{x_k}\ell_k \end{bmatrix} \\ &= \sum_{\ell_1 \in L(T_1),\dots,\ell_k \in L(T_k)} \frac{T_1 \cdots T_k}{\ell_1 \cdots \ell_k} \cdot J_{\mathbf{x}_k}(\ell_1,\dots,\ell_k). \end{aligned} \tag{3.1}$$

Call a set of linear polynomials *independent* if the corresponding homogeneous linear parts (i.e., the constant-free parts) are $\mathbb{F}$-linearly independent. The term $J_{\mathbf{x}_k}(\ell_1,\dots,\ell_k)$ ensures that the above sum is only over those $\ell_1,\dots,\ell_k$ that are independent linear polynomials (otherwise the Jacobian vanishes). Note that this implies that no ℓ_i can repeat in any denominator of (3.1). The sum has the form of a depth-3 circuit, call it H_0, and we construct a low variate Ψ such that $\Psi(H_0) \neq 0$. We show that this is achieved by a Ψ that preserves the independence of a "small" set of linear polynomials—which we call a *certificate* of H_0. (The certifying path technique is from [SS11].)

Certificate of H_0. We can assume that each of the terms $J_{\mathbf{x}_k}(\ell_1,\dots,\ell_k)$ in (3.1) is a *nonzero* field constant. Let $\mathcal{L}(H_0)$ be the set of all linear polynomials occurring in the denominator terms "$\ell_1 \cdots \ell_k$" of all the summands in sum (3.1). Hence, $\mathcal{L}(H_0)$ is the set of all distinct ℓ_i's that occur in the denominator of (3.1). This means, the depth-3 circuit H_0 has the form $H_0 = T \cdot \sum_L \alpha_L/\ell_1 \cdots \ell_k$, where $T := \prod_{i=1}^k T_i$, α_L is a nonzero field constant and the sum runs over some sets $L = \{\ell_1,\dots,\ell_k\}$ of k independent linear polynomials contained in $\mathcal{L}(H_0)$.

Define, *content* of a depth-3 circuit $G = \sum_i P_i$, where P_i is a product of linear polynomials, as $\mathrm{cont}(G) := \gcd_i\{P_i\}$, and let the *simple part*, denoted by $\mathrm{sim}(G)$, be defined as $G/\mathrm{cont}(G)$. Hence $\mathrm{cont}(H_0) = \gcd_L\{T/\ell_1 \cdots \ell_k\}$ and

$$\mathrm{sim}(H_0) = F_0 \cdot \sum_L \frac{\alpha_L}{\ell_1 \cdots \ell_k}, \text{ where } F_0 = \frac{T}{\mathrm{cont}(H_0)}. \tag{3.2}$$

Note that F_0 is the least common multiple of the denominators in (3.1), and hence equal to the product of the linear polynomials in $\mathcal{L}(H_0)$. Thus, $\deg(F_0) = |\mathcal{L}(H_0)|$. For any $\ell \in \mathcal{L}(H_0)$, the terms in $\mathrm{sim}(H_0)$ that survive modulo ℓ are those with ℓ in the denominator "$\ell_1 \cdots \ell_k$" of the above expression. Hence,

$$H_1 := \mathrm{sim}(H_0) \bmod \ell_1 \quad = \quad \frac{F_0}{\ell_1} \cdot \sum_{L := \{\ell_2,\dots,\ell_k\}} \frac{\alpha_L}{\ell_2 \cdots \ell_k}.$$

We can treat H_1 as a depth-3 circuit in one less variable: Suppose that $\ell_1 = c_1x_1 + \sum_{i=2}^n c_ix_i$, where c_i's $\in \mathbb{F}$ and $c_1 \neq 0$, then we can replace x_1 by $-\sum_{i=2}^n c_ix_i/c_1$ in $\mathrm{sim}(H_0)$, particularly in F_0/ℓ_1 (of course, after dividing F_0 by ℓ_1) as well as in each of $\ell_2,\dots,\ell_k$ in the denominators, so that H_1 becomes a depth-3 circuit in $\mathbb{F}[x_2,\dots,x_n]$. Therefore, it makes perfect sense to talk about $\mathrm{cont}(H_1)$ and $\mathrm{sim}(H_1)$. Observe that $\ell_2,\dots,\ell_k$ remain independent linear polynomials modulo ℓ_1, and so H_1 is a depth-3 circuit of the "same nature" as H_0 but with one less linear polynomial in the denominators. Also, the set of linear polynomials $\mathcal{L}(H_1)$ is a subset of the set of linear polynomials $\mathcal{L}(H_0)$ modulo ℓ_1. Extending the above argument, it is possible to define a sequence of circuits $H_i := \mathrm{sim}(H_{i-1}) \bmod \tilde{\ell}_i$, $(1 \leq i \leq k)$, where $\tilde{\ell}_i \in \mathcal{L}(H_{i-1})$.

Further, $\mathcal{L}(H_i)$ is a subset of $\mathcal{L}(H_{i-1})$ modulo $\tilde{\ell}_i$, which implies that essentially there are independent linear polynomials, say $\ell_1, \ldots, \ell_k$, in $\mathcal{L}(H_0)$ such that $\tilde{\ell}_i = \ell_i \bmod (\ell_1, \ldots, \ell_{i-1})$ and therefore $H_i = \mathrm{sim}(H_{i-1}) \mod (\ell_1, \ldots, \ell_i)$.

LEMMA 3.2 (certifying path). *There exist independent linear polynomials* $\{\ell_1, \ldots, \ell_k\} \subseteq \mathcal{L}(H_0)$ *such that* $H_i \neq 0 \bmod (\ell_1, \ldots, \ell_i)$ $\forall i \in [k]$, *and* H_k *is a nonzero product of linear polynomials in* $\mathcal{L}(H_0)$ *modulo* $(\ell_1, \ldots, \ell_k)$.

Proof. Since $T_1, \ldots, T_k$ is a transcendence basis, we start with $H_0 \neq 0$. The proof is by induction on k and follows the sketch given while defining $\mathrm{sim}(\cdot)$.

The degree of the nonzero polynomial $\mathrm{sim}(H_0)$ is $|\mathcal{L}(H_0)| - k$. By Chinese remaindering, there exists an $\ell_1 \in \mathcal{L}(H_0)$ such that $H_1 := \mathrm{sim}(H_0) \bmod \ell_1 \neq 0$. In the base case ($k = 1$), it is easy to see that H_1 is a nonzero product of linear polynomials modulo ℓ_1. For any larger k, the depth-3 polynomial H_1 has exactly the same form as H_0 but with $k-1$ independent linear polynomials in the denominators. Inducting on this smaller value $k-1$, keeping in mind that $\mathcal{L}(H_i) \subset \mathcal{L}(H_0)$ modulo $(\ell_1, \ldots, \ell_i)$, completes the proof. ∎

A set $\{\ell_1, \ldots, \ell_k\}$, satisfying Lemma 3.2, is called a *certifying path* of H_0. Fix a certifying path $\{\ell_1, \ldots, \ell_k\}$. Let $\Psi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[z_1, \ldots, z_{k+1}]$ be such that $\Psi(\ell_1), \ldots, \Psi(\ell_k)$ are independent linear polynomials in $\mathbb{F}[\mathbf{z}]$ and for every $\ell \in \cup_{i=1}^k L(T_i)$, $\ell \neq 0 \bmod (\ell_1, \ldots, \ell_k)$ iff $\Psi(\ell) \neq 0 \bmod (\Psi(\ell_1), \ldots, \Psi(\ell_k))$. We call such a Ψ a *rank-*$(k+1)$ *preserving map* for $\mathcal{L}(H_0)$. It can be shown that one of the maps $\Psi_b : x_i \mapsto \sum_{j=1}^{k+1} z_j b^{ij}$, where b runs over $dkn(k+1)^2$ distinct elements of $\mathbb{F}$, is a rank-$(k+1)$ preserving map for H_0. (It is a simple corollary of Lemma 2.6).

THEOREM 3.3. *If* $\Psi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[z_1, \ldots, z_{k+1}]$ *is a rank-*$(k+1)$ *preserving map for* $L(H_0)$, *then* $\Psi(H_0) \neq 0$.

Proof. Let $\{\ell_1, \ldots, \ell_k\}$ be the certifying path of H_0 fixed above. Let $\mathcal{I}_i := \langle \ell_1, \ldots, \ell_i \rangle$, the ideal generated by the linear forms $\{\ell_1, \ldots, \ell_i\}$. The proof is by reverse induction on k: Assuming $\Psi(H_i) \neq 0 \bmod \Psi(\mathcal{I}_i)$, we show that $\Psi(H_{i-1}) \neq 0 \bmod \Psi(\mathcal{I}_{i-1})$ for $k \geq i \geq 2$. The base case is easy, as by Lemma 3.2, H_k is a *nonzero* product of linear polynomials in $\mathcal{L}(H_0)$ modulo $\mathcal{I}_k$. Hence, by the definition of a rank-$(k+1)$ preserving map, $\Psi(H_k) \neq 0 \mod \Psi(\mathcal{I}_k)$ (the ideal generated by independent linear polynomials is an integral domain).

The proof of the inductive step proceeds as follows. Assume that we know $\Psi(H_i) \neq 0 \bmod \Psi(\mathcal{I}_i)$. By construction,

$$H_{i-1} = \mathrm{cont}(H_{i-1}) \cdot \mathrm{sim}(H_{i-1}) = \mathrm{cont}(H_{i-1}) \cdot [q_i \ell_i + H_i] \mod \mathcal{I}_{i-1}$$

for some polynomial q_i, which means, $\Psi(H_{i-1}) = \Psi(\mathrm{cont}(H_{i-1})) \cdot [\Psi(q_i)\Psi(\ell_i) + \Psi(H_i)] \bmod \Psi(\mathcal{I}_{i-1})$.

$$\begin{aligned} \text{If} \quad & [\Psi(q_i)\Psi(\ell_i) + \Psi(H_i)] = 0 \bmod \Psi(\mathcal{I}_{i-1}), \\ \text{then} \quad & \Psi(H_i) = [\Psi(q_i)\Psi(\ell_i) + \Psi(H_i)] = 0 \bmod \Psi(\mathcal{I}_i) \end{aligned}$$

which contradicts the induction hypothesis. Also, by Lemma 3.2, $H_{i-1} \neq 0 \bmod \mathcal{I}_{i-1}$ implies that $\mathrm{cont}(H_{i-1}) \neq 0 \bmod \mathcal{I}_{i-1}$. Note that $\mathrm{cont}(H_{i-1})$ is a product of linear polynomials in $\mathcal{L}(H_0)$ modulo $\mathcal{I}_{i-1}$, and hence each factor ℓ of $\mathrm{cont}(H_{i-1})$ is not in $\mathcal{I}_{i-1}$. Since Ψ is a rank-$(k+1)$ preserving map, $\Psi(\ell)$ continues to not be in $\Psi(\mathcal{I}_{i-1})$. Therefore,

$$\Psi(H_{i-1}) = \Psi(\mathrm{cont}(H_{i-1})) \cdot [\Psi(q_i)\Psi(\ell_i) + \Psi(H_i)] \neq 0 \bmod \Psi(\mathcal{I}_{i-1})$$

as $\mathcal{I}_{i-1}$ is generated by $(i-1)$ independent linear polynomials.

Finally, to obtain $\Psi(H_0) \neq 0$ from $\Psi(H_1) \neq 0 \bmod \Psi(\mathcal{I}_1)$, use the same argument as above and that $\Psi(\ell) \neq 0$ for every $\ell \in \cup_{i=1}^{k} L(T_i)$. ∎

Since H_0 was just a maximal nonzero minor of $\mathcal{J}(T_1, \ldots, T_r)$, we have the following corollary (using Fact 2.2).

COROLLARY 3.4. *Let* $\Psi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[z_1, \ldots, z_{r+1}]$ *be a rank-*$(r+1)$ *preserving map for* $L(T_1) \cup \cdots \cup L(T_m)$ *(where* $\{T_1, \ldots, T_m\}$ *are products of linear functions with transcendence degree bounded by* r*). Then,* $\Phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[z_1, \ldots, z_{r+1}, t, y_1, \ldots, y_{r+1}]$ *defined as* $\Phi(x_i) \mapsto \Psi(x_i) + \sum_{j=1}^{r+1} t^{ij} y_j$ *is a faithful homomorphism for* $\{T_1, \ldots, T_m\}$.

Proof. Let $\operatorname{trdeg}\{T_1, \ldots, T_m\} = k \leq r$. Since any rank-$(r+1)$ preserving map Ψ is also a rank-$(k+1)$ preserving map, Theorem 3.3 states that Ψ preserves the rank of $\mathcal{J}(T_1, \ldots, T_m)$. Hence by Lemma 2.7 we have that Φ is a faithful homomorphism for $\{T_1, \ldots, T_r\}$. ∎

With this, we can prove both Theorems 1.1 and 1.9.

THEOREM 1.1 (restated). *Let* C *be a polydegree circuit of size* s *and each of* $T_1, \ldots, T_m$ *be a product of* d *linear polynomials in* $\mathbb{F}[x_1, \ldots, x_n]$ *such that* $\operatorname{trdeg}_{\mathbb{F}}\{T_1, \ldots, T_m\} \leq r$. *A hitting set for such* $C(T_1, \ldots, T_m)$ *can be constructed in time polynomial in* n *and* $(sd)^r$*, assuming char*$(\mathbb{F}) = 0$ *or* $> d^r$.

Proof of Theorem 1.1. Corollary 3.4 gives a homomorphism

$$\Phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[z_1, \ldots, z_{r+1}, t, y_1, \ldots, y_{r+1}]$$

that is faithful for $\{T_1, \ldots, T_m\}$. By Theorem 2.4, we have that $C(T_1, \ldots, T_m) = 0$ iff $\Phi(C(T_1, \ldots, T_m)) = 0$. Since C is a polydegree circuit of size s, $\Phi(C(T_1, \ldots, T_m))$ is a polynomial of degree at most $ds^{O(1)}$ and $nrds^{O(1)}$ in the variables $\{\mathbf{y}, \mathbf{z}\}$ and t, respectively. Using [Sch80, Zip79, DL78], we can construct a hitting set for $\Phi(D)$ in time polynomial in $n(sd)^r$. Since construction of Ψ takes time poly(ndr), the total time taken is poly$(n, (sd)^r)$. ∎

THEOREM 1.9 (restated). *Given black box access to polynomials* $T_1, \ldots, T_r$ *that are products of* d *linear polynomials in* $\mathbb{F}[x_1, \ldots, x_n]$*, there is a poly*$((nd)^r)$ *time algorithm to test whether they are algebraically independent, assuming char*$(\mathbb{F}) = 0$ *or* $> d^r$.

Proof of Theorem 1.9. Corollary 3.4 gives a homomorphism

$$\Phi : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[z_1, \ldots, z_{r+1}, t, y_1, \ldots, y_{r+1}]$$

that is faithful for $\{T_1, \ldots, T_r\}$. Hence it suffices to check if $\{\Phi(T_1), \ldots, \Phi(T_r)\}$ are algebraically independent or not. Since each $\Phi(T_i)$ is an $O(r)$-variate polynomial of degree poly(nd), they have at most poly$((nd)^r)$ monomials. It is well known [KS01, BOT88] that an n-variate degree d multivariate polynomial with R monomials can be reconstructed from poly(n, d, R) evaluations. Hence, using poly$((nd)^r)$ evaluations, each $\Phi(T_i)$ can be explicitly written down as a sum of monomials. Hence, the Jacobian $\mathcal{J}(\Phi(T_1), \ldots, \Phi(T_r))$ can be written down explicitly, and an application of the Schwartz–Zippel lemma in [Sch80, Zip79, DL78] allows us to check if $\mathcal{J}(\Phi(T_1), \ldots, \Phi(T_r))$ has full rank in deterministic poly$((nd)^r)$ time. ∎

4. Hitting set for constant-depth constant-occur formulas.

Bounding the top fanin. Let C belong to the class $\mathcal{C}$ of depth-D occur-k formulas of size s, with potentially large top fanin. The following easy observation allows us to slightly modify C to work with a bounded top fanin formula (without increasing the other parameters too much).

OBSERVATION 4.1. *If $C(x_1,\ldots,x_n)$ is nonconstant and nonzero, then there is an i such that $\tilde{C} := C(x_1,\ldots,x_{i-1},x_i+1,x_{i+1},\ldots,x_n) - C(x_1,\ldots,x_n) \neq 0$, assuming* $\mathrm{char}(\mathbb{F}) > s^D$ *(i.e., the bound on the degree of C).*

COROLLARY 4.2. *Let $\mathcal{H}$ be a hitting set for the class of occur-$2k$, top fanin $2k$, depth-$(D+1)$, and size (s^2+s) formulas. Then, there is a hitting set $\mathcal{H}'$ with $|\mathcal{H}'| = |\mathcal{H}|^{O(1)}$ for the class of occur-k, depth-D, size s formulas of unbounded top fanin.*

Proof. Let C be an occur-k, depth-D formula of size s that computes a nonzero polynomial. By Observation 4.1, there exists some i such that

$$\tilde{C} := C(x_1,\ldots,x_{i-1},x_i+1,x_{i+1},\ldots,x_n) - C(x_1,\ldots,x_n) \neq 0.$$

If C has a $+$ gate on top, then $C(\mathbf{x}) = \sum_{i=1}^{m} T_i$, where the T_i's are computed by $\times\curlywedge$ gates at the next level. Since x_i occurs in at most k of the T_i's, $\tilde{C}$ has top fanin at most $2k$. If C has a $\times\curlywedge$ gate on top, then $\tilde{C}$ has a $+$ gate on top with fanin 2 and $\mathrm{depth}(\tilde{C}) = D+1$. Therefore, $\tilde{C}$ belongs to the class of depth-$(D+1)$ occur-$2k$ formulas of size at most (s^2+s), and a $+$ gate on top with fanin bounded by $2k$. (The size of $C(x_1,\ldots,x_{i-1},x_i+1,x_{i+1},\ldots,x_n)$ is bounded by s^2 as the total sparsity of the polynomials corresponding to the leaves of $C(x_1,\ldots,x_n)$ can grow at most quadratically as x_i is replaced by x_i+1.)

Since $\mathcal{H}$ is a hitting set for the class of occur-$2k$, top fanin $2k$, depth-$(D+1)$, size (s^2+s) formulas, define $\mathcal{H}' \supset \mathcal{H}$ by including points $(\alpha_1+1,\alpha_2,\ldots,\alpha_n)$, $(\alpha_1,\alpha_2+1,\ldots,\alpha_n),\ldots,(\alpha_1,\ldots,\alpha_{n-1},\alpha_n+1)$ in $\mathcal{H}'$ for every point $(\alpha_1,\alpha_2,\ldots,\alpha_n) \in \mathcal{H}$. Observe that $\mathcal{H}'$ is a hitting set for C and $|\mathcal{H}'| = n \cdot |\mathcal{H}|$. ∎

Therefore, it is sufficient to construct a hitting set for the class of bounded top-fanin bounded-occur formulas. By reusing symbols, assume that C is a depth-D occur-k formula of size s with a $+$ gate on top having top fanin at most k.

In order to construct a hitting set for $C(\mathbf{x}) = \sum_{i=1}^{k} T_i$, we shall solve the following slightly more generalized goal.

> *Goal:* Let $T_1,\ldots,T_k$ be polynomials computed by occur-k depth-D circuits of size s each. Construct a map Φ that is faithful for $\{T_1,\ldots,T_k\}$.

Theorem 2.4 asserts that any such faithful map would preserve the nonzeroness of C. Let $\mathbf{T}_r = \{T_1,\ldots,T_r\}$ be a transcendence basis of $\mathbf{T}$. Since $\mathcal{J}_{\mathbf{x}}(\mathbf{T}_r)$ has full rank ($\mathrm{char}(\mathbb{F}) = 0$ or $> s^{Dr}$, by Lemma 2.7), assume that the columns corresponding to $\mathbf{x}_r = \{x_1,\ldots,x_r\}$ form a nonzero minor of $\mathcal{J}_{\mathbf{x}}(\mathbf{T}_r)$. By Lemma 2.7, it suffices to construct a Ψ that keeps $J_{\mathbf{x}_r}(\mathbf{T}_r) \neq 0$.

Proof idea. Identify a gate with the polynomial it computes, and count *level* of a gate from the top—the gates T_i's are at level 1. Suppose each T_i is a $\times\curlywedge$ gate and $T_i = \prod_{\ell=1}^{d} P_{i,\ell}^{e_{i,\ell}}$, where $P_{i,\ell}$'s are gates at level 2. Since T_i is also an occur-k formula, $x_1,\ldots,x_r$ appear in at most kr of the $P_{i,\ell}$'s, say $P_{i,1},\ldots P_{i,kr}$. Hence,

$$\begin{aligned} \partial_{x_j} T_i &= \left(\prod_{\ell=kr+1}^{d} P_{i,\ell}^{e_{i,\ell}}\right) \cdot \partial_{x_j}\left(\prod_{\ell=1}^{kr} P_{i,\ell}^{e_{i,\ell}}\right) \quad \text{for every } 1 \le i,j \le r \\ \implies J_{\mathbf{x}_r}(\mathbf{T}_r) &= \left(\prod_{i=1}^{r}\prod_{\ell=kr+1}^{d} P_{i,\ell}^{e_{i,\ell}}\right) \cdot J_{\mathbf{x}_r}\left(\prod_{\ell=1}^{kr} P_{1,\ell}^{e_{1,\ell}},\ldots,\prod_{\ell=1}^{kr} P_{r,\ell}^{e_{r,\ell}}\right). \end{aligned}$$

Now notice that the Jacobian term on the right-hand side of the last equation is a polynomial in $P_{i,\ell}$ and $\partial_{x_j} P_{i,\ell}$, for $1 \leq i, j \leq r$, and $1 \leq \ell \leq kr$. (Note the irrelevance of the exponents $e_{i,\ell}$'s.) So, if Ψ is faithful for the set $\mathcal{P} := \{P_{i,\ell}, \partial_{x_j} P_{i,\ell} : 1 \leq i, j \leq r, 1 \leq \ell \leq kr\}$ and the singleton sets $\{P_{i,\ell}\}$ for $1 \leq i \leq r$, $kr + 1 \leq \ell \leq d$, then $\Psi(J_{\mathbf{x}_r}(\mathbf{T}_r)) \neq 0$. In other words, the task of constructing a faithful homomorphism is instead replaced by the task of constructing a map that keeps every factor in the above product nonzero. To get some more intuition, consider the case of a depth-4 circuit in which case each of the $P_{i,j}$'s are sparse polynomials. Preserving the nonzeroness of sparse polynomials can be achieved via ideas from [KS01].

Observe that the polynomials in $\mathcal{P}$ and the singleton sets are (zeroth and first order) derivatives of the gates at level 2, and further these sets involve (the derivatives of) *disjoint* groups of level 2 gates. This disjointness feature ensures that the number of such sets is at most s.

Thus, we have reduced the problem of constructing a faithful map Φ for $\mathbf{T}$ (gates at level 1) to the problem of constructing a map Ψ that is faithful for at most s many sets each containing derivatives of gates at the second level. Now, the idea is to carry forward this argument recursively to deeper levels: In the next level of the recursion we reduce the problem to constructing a map that is faithful for at most s sets containing (zeroth, first, and second order) derivatives of disjoint groups of gates at level 3, and so on. Eventually, the recursion reaches the level of the sparse polynomials (the leaf nodes) where a faithful map can be constructed using ideas from [KS01].

Let us formalize this proof idea. For any *multiset* of variables S, let $\Delta_S f$ denote the partial derivative of f with respect to the variables in S (including repetitions, as S is a multiset). Let $\mathrm{var}(S)$ denote the set of distinct variables in S.

LEMMA 4.3 (Gcd trick). *Let G be any gate in C and $S_1, \ldots, S_w$ be multisets of variables. Then there exists another occur-k formula G' for which the vector of polynomials $(\Delta_{S_1} G, \ldots, \Delta_{S_w} G) = V_G \cdot (\Delta_{S_1} G', \ldots, \Delta_{S_w} G')$ such that*

1. *if G is a $+$ gate, then G' is also a $+$ gate whose children consist of at most $k \cdot |\cup_{i=1}^{w} \mathrm{var}(S_i)|$ of the children of G, and $V_G = 1$;*
2. *if G is a $\times\curlywedge$ gate, then G' is also a $\times\curlywedge$ gate whose children consist of at most $k \cdot |\cup_{i=1}^{w} \mathrm{var}(S_i)|$ of the children of G, and $V_G = G/G'$.*

Further, the gates constituting G' and V_G are disjoint.

Proof.

1. Suppose $G = H_1 + \cdots + H_m$. Then at most $k \cdot |\cup \mathrm{var}(S_i)|$ of its children depend on the variables present in $\cup \mathrm{var}(S_i)$; let G' be the sum of these children. Then, $\Delta_{S_i} G = \Delta_{S_i} G'$ as the other gates are independent of the variables in $\cup S_i$.
2. Suppose $G = H_1^{e_1} \cdots H_m^{e_m}$. Since G is a gate in an occur-k formula, at most $k \cdot |\cup \mathrm{var}(S_i)|$ of the H_i's depend on the variables in $\cup S_i$; call these $H_1, \ldots, H_t$. Let $G' := H_1^{e_1} \cdots H_t^{e_t}$ and $V_G := G/G'$. Then, $\Delta_{S_i} G = V_G \cdot \Delta_{S_i} G'$ as claimed. ∎

We say that a map is faithful for a collection of sets if it is faithful for every set in the collection. Going by the "proof idea," suppose at the ℓth level of the recursion we want to construct a Ψ_ℓ that is faithful for a collection of (at most) s sets of polynomials, each set containing at most r_ℓ partial derivatives (of order up to c_ℓ) of the gates at level ℓ. Moreover, the sets involve derivatives of disjoint groups of gates. To begin with, $\ell = 1$ and we wish to construct a Ψ_1 that is faithful for just one set $\mathbf{T}$, so $r_1 \leq k$ and $c_1 = 0$. The next lemma captures the evolution of the recursion and is essentially a careful analysis of the growth of the sets as we descend levels.

LEMMA 4.4 (evolution via factoring). *Let $\mathcal{U}$ be a set of r_ℓ derivatives (of orders up to c_ℓ) of gates $\mathcal{G}_{\mathcal{U}}$ at level ℓ, and $\mathcal{U}'$ be a transcendence basis of $\mathcal{U}$. Any $|\mathcal{U}'| \times |\mathcal{U}'|$*

minor of $\mathcal{J}_{\mathbf{x}}(\mathcal{U}')$ is of the form $\prod_i V_i^{e_i}$, where V_i's are polynomials in at most $r_{\ell+1} := (c_\ell+1)\cdot 2^{c_\ell+1}k\cdot {r_\ell}^2$ many derivatives (of order up to $c_{\ell+1} := c_\ell+1$) of disjoint groups of children of $\mathcal{G}_{\mathcal{U}}$.

Proof. Let $r' = |\mathcal{U}'|$ and M be an arbitrary $r'\times r'$ submatrix of $\mathcal{J}_{\mathbf{x}}(\mathcal{U}')$. Let $(\Delta_{S_1}G,\dots,\Delta_{S_{r'}}G)$ be an arbitrary row of M that corresponds to certain derivatives of a node G at level ℓ. Since each of the derivatives is of order at most $c_\ell+1$, we have that $|\cup_{i=1}^{r'}\operatorname{var}(S_i)| \le r'(c_\ell)+r' \le r'(c_\ell+1)$. By applying Lemma 4.3 to this row, we can express this row as $V_G\cdot(\Delta_{S_1}G',\dots,\Delta_{S_{r'}}G')$. So, $\det(M)=\prod_G V_G\cdot\det M'$, where M' is the submatrix obtained after removing the V_G's common from each of the rows. Thus, the elements present inside M' are of the form $\Delta_{S_i}G'$, where G' has at most $kr'(c_\ell+1)$ children.

Since each $|S_i|\le c_\ell+1$ and G is a node in an occur-k formula, at most $k(c_\ell+1)$ children of G' depend on $\operatorname{var}(S_i)$.

If G' is a $+$ gate, then $\Delta_{S_i}G'$ is the sum of the derivatives of at most $k(c_\ell+1)$ of its children (that depend on $\operatorname{var}(S_i)$).

If G' is a $\times\curlywedge$ gate computing $H_1^{e_1}\cdots H_t^{e_t}$ (where $t\le kr'(c_\ell+1)$), then $\Delta_{S_i}G'$ is a polynomial combination of the H_i's and $\{\Delta_T H_j\}_{\emptyset\ne T\subseteq S_i}$ for each H_j depending on $\operatorname{var}(S_i)$.

Hence in either case, $\Delta_{S_i}G'$ is a polynomial in at most $kr'(c_\ell+1)+k(c_\ell+1)\times(2^{c_\ell+1}-1)$ many derivatives (of order at most $(c_\ell+1)$) of the children of G'. Summing across the r' elements in that row, the number of derivatives used are at most $kr'(c_\ell+1)+k(c_\ell+1)(2^{c_\ell+1}-1)r' \le kr'(c_\ell+1)2^{c_\ell+1}$. Summing over all r' rows, the number of derivatives used is at most $kr'^2(c_\ell+1)2^{c_\ell+1}$. Further, each V_G that was removed as a common factor in a row is just a product of gates in level $(\ell+1)$, and are disjoint from the gates whose derivatives constitute M' (by Lemma 4.3). Thus, if r_ℓ was an upper bound for r', then we have that the number of derivatives used is at most

$$r_{\ell+1} = kr_\ell^2(c_\ell+1)2^{c_\ell+1}$$

as claimed. $\square$

Let $\mathcal{C}_\ell$ denote the collection of sets for which we want to construct a faithful map Ψ_ℓ at the ℓth level of the recursion. To begin with, $\mathcal{C}_1 = \{\{T_1,\dots,T_k\}\}$ and the collection $\mathcal{C}_{\ell+1}$ is formed from $\mathcal{C}_\ell$ in the following fashion: For any $S\in\mathcal{C}_\ell$, consider a maximal nonsingular minor of $\mathcal{J}(S)$, and by Lemma 4.4 this minor can be written as a product of $V_i^{(S)}$'s where each $V_i^{(S)}$ is a function of at most $r_{\ell+1}$ derivatives of polynomials computed in level $\ell+1$. Denote the set of derivatives that $V_i^{(S)}$ depends on as $\operatorname{Elem}(V_i^{(S)})$. Then, $\mathcal{C}_{\ell+1} = \{\operatorname{Elem}(V_i^{(S)}) \;:\; S\in\mathcal{C}_\ell\}$.

It follows from the above lemma that the groups of gates whose derivatives form the different $\operatorname{Elem}(V_i)$'s are disjoint and therefore $|\mathcal{C}_{\ell+1}|\le s$. Using Lemmas 2.7 and 4.4, we can lift a map $\Psi_{\ell+1}$ to construct Ψ_ℓ.

COROLLARY 4.5. *If $\Psi_{\ell+1}$ is faithful for $\mathcal{C}_{\ell+1}$, then*

$$\Psi_\ell : x_i \mapsto \left(\sum_{j=1}^{r_\ell} y_{j,\ell}\cdot(t_\ell)^{ij}\right) + \Psi_{\ell+1}(x_i)$$

is faithful for $\mathcal{C}_\ell$, where $\{y_{1,\ell},\dots,y_{r_\ell,\ell},t_\ell\}$ is a fresh set of variables.

Unfolding the recursion until we reach the level of the sparse polynomials (at level $(D-2)$), if Ψ_{D-2} is a faithful homomorphism for $\mathcal{C}_{D-2}$, then Ψ_1 defined by

$$\Psi_1(x_i) \quad \mapsto \quad \sum_{\ell=1}^{D-3}\left(\sum_{j=1}^{r_\ell} y_{j,\ell}t_\ell^{ij}\right) + \Psi_{D-2}(x_i) \tag{4.1}$$

is faithful for $\{T_1,\ldots,T_k\}$. Hence, we are reduced to the task of constructing a faithful map for the collection $\mathcal{C}_{D-2}$ which consists of at most s sets of derivatives of s-sparse polynomials (i.e., consisting of at most s monomials), and each set being of size at most r_{D-2}. Constructing a faithful homomorphism for a collection of sets of sparse polynomials can be achieved using standard techniques used in sparse-PIT.

LEMMA 4.6. *Let $\mathcal{C} = \{W_1,\ldots,W_s\}$, where each W_s is a set of r polynomials that are n-variate, degree d, and s-sparse. For each $1 \le p \le O((s^r nd)^4)$, the map $\Psi^{(p)} : \mathbb{F}[\mathbf{x}] \to \mathbb{F}[y_1,\ldots,y_r,t,u]$ is defined by*

$$\Psi^{(p)} : x_i \quad \mapsto \quad \left(\sum_{j=1}^{r} y_j t^{ij}\right) + u^{(dr+1)^i \bmod p}.$$

Then one of the maps $\Psi^{(p)}$ (as p varies over the specified range) is faithful for the collection $\mathcal{C}$.

Proof. For each $W \in \mathcal{C}$, any maximal nonzero minor of $\mathcal{J}_{\mathbf{x}}(W)$ is a sparse polynomial with sparsity bounded by s^r and degree bounded by dr. Using [KS01], the nonzeroness of this determinant is maintained by one of the maps $\Phi^{(p)} : x_i \mapsto u^{(dr+1)^i \bmod p}$ as p varies from 1 to $O((s^r nd)^4)$. For such a map $\Phi^{(p)}$, Lemma 2.7 asserts that $\Psi^{(p)} : x_i \mapsto \sum_{j=1}^{r} y_j t^{ij} + \Phi^{(p)}(x_i)$ is faithful for $\mathcal{C}$. ∎

With the above lemma and (4.1), we can achieve our goal of constructing a faithful homomorphism for $\{T_1,\ldots,T_k\}$.

THEOREM 4.7. *Let $T_1,\ldots,T_k$ be polynomials computed by depth-D occur-k formulas of size at most s each. Then, a homomorphism that is faithful for $\{T_1,\ldots,T_k\}$ can be constructed in time polynomial in s^R, where $R = (2k)^{2D\cdot 2^D}$ (assuming char$(\mathbb{F}) = 0$ or $> s^R$).*

Proof. Unfolding the recursion in Corollary 4.5 and (4.1), it suffices to construct a map $\Psi^{(p)}$ that is faithful for $\mathcal{C}_{D-2}$ (as defined earlier). Recall that $\mathcal{C}_{D-2}$ is a collection of s sets of at most r_{D-2} derivatives of s-sparse polynomials of degree at most d. Hence, we can apply Lemma 4.6 to this collection $\mathcal{C}_{D-2}$ to get that

$$\Psi^{(p)}(x_i) \quad \mapsto \quad \sum_{\ell=1}^{D-2}\left(\sum_{j=1}^{r_\ell} y_{j,\ell}t_\ell^{ij}\right) + u^{(sr_\ell+1) \bmod p}.$$

Using the relation between $r_{\ell+1}$ and r_ℓ from Lemma 4.4, it is easy to show that $r_{D-2} \le R = (2k)^{2D\cdot 2^D}$ and $\sum_{\ell=1}^{D-2} r_\ell = O(R)$. ∎

With the above construction for a faithful homomorphism, we can prove both Theorem 1.3 and Theorem 1.10.

THEOREM 1.3 (restated). *A hitting set for any depth-D occur-k formula $C(\mathbf{x}) = T_1+\cdots+T_k$ of size s can be constructed in time polynomial in s^R, where $R = (2k)^{2D\cdot 2^D}$ (assuming char$(\mathbb{F}) = 0$ or $> s^R$).*

Proof. Theorem 4.7 gives a homomorphism $\Psi^{(p)}$ that is faithful for $\{T_1, \ldots, T_k\}$ and reduces the number of variables to $O(\sum_{\ell=1}^{D-2} r_\ell) = O(R)$. By Theorem 2.4, we have that $C = T_1 + \cdots + T_k = 0$ iff $\Psi^{(p)}(T_1) + \cdots + \Psi^{(p)}(T_k) = 0$. Since C is a polydegree circuit of size s, $\Psi^{(p)}(C)$ is a polynomial of degree at most poly(s) in the $O(R)$ variables. Using the Schwartz–Zippel lemma in [Sch80, Zip79, DL78], we can construct a hitting set for $\Phi(D)$ in time polynomial in s^R. ∎

THEOREM 1.10 (restated). *Let $T_1, \ldots, T_r$ be n-variate degree d polynomials computed by depth-D occur-k formulas of size s and presented as black boxes. There is an $(sdn)^R$-time algorithm, where $R = r \cdot (2k)^{2D \cdot 2^D}$, to test whether $\{T_1, \ldots, T_r\}$ are algebraically independent, assuming char$(\mathbb{F}) = 0$ or $> s^R$.*

Proof. Theorem 4.7 gives a homomorphism $\Psi^{(p)}$ that is faithful for $\{T_1, \ldots, T_r\}$. Hence it suffices to check whether $\{\Psi^{(p)}(T_1), \ldots, \Psi^{(p)}(T_r)\}$ are algebraically independent or not. Since each $\Psi^{(p)}(T_i)$ is an $O(R)$-variate polynomial of degree poly(sdn), they have at most poly$((sdn)^R)$ monomials. Using the interpolation algorithms from [KS01, BOT88], each $\Psi^{(p)}(T_i)$ can be explicitly written down as a sum of monomials from poly$((sdn)^R)$ evaluations. Hence, the Jacobian $\mathcal{J}(\Psi^{(p)}(T_1), \ldots, \Psi^{(p)}(T_r))$ can be written down explicitly, and an application of the Schwartz–Zippel lemma in [Sch80, Zip79, DL78] allows us to check if $\mathcal{J}(\Psi^{(p)}(T_1), \ldots, \Psi^{(p)}(T_r))$ has full rank in deterministic poly$((sdn)^R)$ time. ∎

4.1. Restriction to the case of depth-4.

THEOREM 1.4 (restated). *A hitting set for any depth-4 occur-k formula of size s can be constructed in time polynomial in s^{k^2} (assuming char$(\mathbb{F}) = 0$ or $> s^{4k}$).*

Proof. Let $C = \sum_{i=1}^{k} T_i$ be a depth-4 occur-k formula, where $T_i = \prod_{j=1}^{d} P_{ij}^{e_{ij}}$, P_{ij}'s are sparse polynomials. The discussion at the beginning of this section justifies the assumption that the top fanin is k. Once again, assuming $\mathbf{T}_r$ to be a transcendence basis of $\mathbf{T}$, we need to design a Ψ such that $\Psi(J_{\mathbf{x}_r}(\mathbf{T}_r)) \neq 0$. Let us count the number of P_{ij}'s that depend on the variables $\mathbf{x}_r$; the remaining $P_{ij}^{e_{ij}}$'s can be taken out common from every row of $\mathcal{J}_{\mathbf{x}_r}(\mathbf{T}_r)$ while computing its determinant—this is the first "taking common" step. Let $c_{i\ell}$ be the number of P_{ij}'s present in T_i that depend on x_ℓ. The total number of sparse polynomials depending on $\mathbf{x}_r$ is therefore $\sum_{1 \le i, \ell \le r} c_{i\ell}$. From the condition of occur-k, $\sum_i c_{i\ell} \le k$ and hence $\sum_{i,\ell} c_{i\ell} \le rk \le k^2$. Let $c_i := \sum_j c_{ij}$ be the number of $\mathbf{x}_r$-dependent P_{ij}'s present in T_i. For an $\mathbf{x}_r$-dependent P_{ij}, we can also take $P_{ij}^{e_{ij}-1}$ common from the ith row of $\mathcal{J}_{\mathbf{x}_r}(\mathbf{T}_r)$—call this the second taking common step. The sparsity of every entry of the ith row of the residual matrix M—after the two taking common steps—is bounded by $c_i s^{c_i}$, where s is the size of C. Thus, $\det(M)$ has sparsity at most $r! \cdot \prod_{i=1}^{r} c_i s^{c_i} = s^{O(k^2)}$, which implies that $J_{\mathbf{x}_r}(\mathbf{T}_r)$ is a product of at most $s+1$ powers of sparse polynomials, each of whose sparsity is bounded by $s^{O(k^2)}$ and degree bounded by sk. As argued before, use [KS01] along with Lemma 2.7 to construct a hitting set for C in time $s^{O(k^2)}$ (assuming char$(\mathbb{F}) = 0$ or $> s^{4k}$). ∎

5. Lower bounds for the immanant.

For the sake of simplicity, we prove the lower bounds for Det$_n$—determinant of an $n \times n$ matrix $M = (\!(x_{ij})\!)$—assuming zero characteristic. All our arguments apply to Imm$_\chi(M)$ for any character χ and this is elaborated in section 5.5. The following two lemmas are at the heart of our approach to proving lower bounds. Let $\mathbf{x} := \{x_{ij} : 1 \le i, j \le n\}$ and $\mathbf{T} := \{T_1, \ldots, T_m\}$, where T_i's are polynomials in $\mathbb{F}[\mathbf{x}]$. We shall defer the proofs of these lemmas to the end of the section.

LEMMA 5.1. *Suppose* $\mathrm{Det}_n = C(T_1, \ldots, T_m)$*, where* C *is any circuit and let* $\mathbf{T}_r = \{T_1, \ldots, T_r\}$ *be a transcendence basis of* $\mathbf{T}$ *with* $r < n$*. Then, there exist a set of* $r+1$ *variables* $\mathbf{x}_{r+1} \subset \mathbf{x}$ *and an equation* $\sum_{i=1}^{r+1} c_i f_i \cdot M_i = 0$ *such that* M_i*'s are distinct first order principal minors of* M*,* f_i*'s are distinct* $r \times r$ *minors of* $\mathcal{J}_{\mathbf{x}_{r+1}}(\mathbf{T}_r)$*, not all* f_i*'s are zero, and* $c_i \in \mathbb{F}^*$*.*

LEMMA 5.2. *If* $M_1, \ldots, M_t$ *are distinct first order principal minors of* M *and* $\sum_{i=1}^{t} f_i \cdot M_i = 0$ *(not all* f_i*'s are zero), then the total sparsity of the* f_i*'s is at least* $2^{n/2-t}$*.*

5.1. Lower bound on depth-4 occur-k formulas.

Proof of Theorem 1.6. Let C be a depth-4 occur-k formula of size s that computes Det_n. Since Det_n is irreducible we can assume a top $+$ gate in C. Then $\tilde{C} := C(x_{11}+1, x_{12}, \ldots, x_{nn}) - C(\mathbf{x})$ is a depth-4 occur-$2k$ formula of size at most $s^2+s \le 2s^2$ and top fanin bounded by $2k$ (similar argument as at the beginning of section 4). Moreover, $\tilde{C}$ computes the minor of M with respect to x_{11} which is essentially Det_{n-1}. By reusing symbols, assume that C is a depth-4 occur-k formula with top fanin bounded by k, and C computes Det_n.

Let $C = \sum_{i=1}^{k} T_i = \mathrm{Det}_n$, where $T_i = \prod_{j=1}^{d} P_{ij}^{e_{ij}}$, P_{ij}'s are sparse polynomials. Let $\mathbf{T}_r$ be a transcendence basis of $\mathbf{T} = \{T_1, \ldots, T_k\}$. By Lemma 5.1, we have an equation $\sum_{i=1}^{r+1} c_i f_i \cdot M_i = 0$ such that f_i's are distinct $r \times r$ minors of $\mathcal{J}_{\mathbf{x}_{r+1}}(\mathbf{T}_r)$ for some set of $r+1$ variables $\mathbf{x}_{r+1}$. Arguing in the same way as in the proof of Theorem 1.4 (in section 4.1), we can throw away certain common terms from the minors f_i's and get another equation $\sum_{i=1}^{r+1} g_i M_i = 0$, where the sparsity of each g_i is $s^{O(k^2)}$. If we apply Lemma 5.2 on this equation, we get our desired result. ∎

5.2. Lower bound on circuits generated by ΣΠ polynomials.

Proof of Theorem 1.7. In Lemma 5.1, take the T_i's to be sparse polynomials with sparsity bounded by s. Then, in the equation $\sum_{i=1}^{r+1} c_i f_i \cdot M_i = 0$, each f_i has sparsity at most $r! \cdot s^r \le (rs)^r$. Finally, by applying Lemma 5.2, we get

$$(r+1) \cdot (rs)^r \ge 2^{n/2-r} \implies s = 2^{\Omega(n/r)}$$

to obtain the desired lower bound. ∎

5.3. Lower bound on circuits generated by ΠΣ polynomials.

Proof of Theorem 1.8. Let $\mathbf{T} = \{T_1, \ldots, T_m\}$ be products of linear polynomials such that $C(T_1, \ldots, T_m) = \mathrm{Det}_n$ with $\mathbf{T}_k = \{T_1, \ldots, T_k\}$ being a transcendence basis (we choose to denote the transcendence degree by k to be consistent with section 3). By Lemma 5.1, we get $\sum_{i=1}^{k+1} c_i f_i M_i = 0$, where the f_i's are $k \times k$ minors of $\mathcal{J}_{\mathbf{x}_{k+1}}(\mathbf{T}_k)$ and without loss of generality $f_1 \neq 0$.

The coefficient of each M_i in this equation is a $k \times k$ minor of $\mathcal{J}(T_1, \ldots, T_k)$. By expanding each such minor using Fact 3.1, we get that the above equation $\sum_{i=1}^{k+1} c_i f_i M_i = 0$ can be expressed as

$$H_0 := T \cdot \sum_{L} \frac{\alpha_L(\mathbf{M}_{k+1})}{\ell_1 \cdots \ell_k} = 0,$$

where each $\alpha_L(\mathbf{M}_{k+1}) := \sum_{i=1}^{k+1} \alpha_{L,i} M_i$ is an $\mathbb{F}$-linear combination of the distinct minors $\mathbf{M}_{k+1} := \{M_1, \ldots, M_{k+1}\}$. Observe that H_0 is a sum of products of linear polynomials, with "coefficients" being $\mathbb{F}$-linear combinations of $\mathbf{M}_{k+1}$. And since $f_1 \neq 0$, the coefficient of M_1 in H_0 is a nonzero depth-3 circuit.

The idea is to apply a similar treatment as in section 3 to evolve H_0. The invariant that shall be maintained is that the coefficient of M_1 (modulo some linear polynomials), which is a depth-3 circuit, would stay nonzero. This would finally yield a nontrivial linear combination $\alpha_L(\mathbf{M}_{k+1}) = 0 \bmod \boldsymbol{\ell}_k$ (where $\boldsymbol{\ell}_k$ is a set of k independent linear polynomials), whence we can apply the following lemma (whose proof shall be deferred to the end of the section as well).

LEMMA 5.3. *If $M_1, \ldots, M_t$ are distinct first order principal minors of M and $\sum_{i=1}^{t} \alpha_i M_i = 0 \bmod \boldsymbol{\ell}_k$ (not all $\alpha_i = 0$) for independent linear polynomials $\boldsymbol{\ell}_k$, then $t + k \geq n$.*

Formally, define the *content* of $H = T \sum_L \alpha_L(\mathbf{M}_{k+1})/\ell_1 \cdots \ell_k$ as $\operatorname{cont}(H) := \gcd_L\{T/\ell_1 \cdots \ell_k\}$, and also define $\operatorname{sim}(H) := H/\operatorname{cont}(H)$. Therefore,

$$\operatorname{sim}(H_0) = F_0 \sum_L \frac{\alpha_L(\mathbf{M}_{k+1})}{\ell_1 \cdots \ell_k} = 0,$$

where F_0 is the product of all distinct linear polynomials appearing in the denominator.

The coefficient of M_1 in the above expression is $F_0 \sum_L \alpha_{L,1}/\ell_1 \cdots \ell_k$, which by assumption is a nonzero depth-3 circuit of degree at most $|\mathcal{L}(H_0)| - k$. Therefore by Chinese remaindering, $\exists \ell_1 \in \mathcal{L}(H_0)$ such that this coefficient remains nonzero modulo ℓ_1. Hence, we can define $H_1 := \operatorname{sim}(H_0) \bmod \ell_1$ which has the form

$$H_1 = \frac{F_0}{\ell_1} \cdot \sum_{L \ni \ell_1} \frac{\alpha_L(\mathbf{M}_{r+1})}{\ell_2 \cdots \ell_k} = 0 \bmod \ell_1.$$

Thus, we may write $H_1 = F_1 \sum_L \alpha_L(\mathbf{M}_{k+1} \bmod \ell_1)/\ell_2 \cdots \ell_k = 0$, and the choice of ℓ_1 maintains the invariant that the coefficient of $(M_1 \bmod \ell_1)$ is nonzero.

The above argument can be repeated inductively. In general, we have $H_i = \operatorname{sim}(H_{i-1}) \bmod \ell_i$ for a similar choice of ℓ_i via Chinese remaindering, and H_i has the form

$$H_i = F_i \sum_L \frac{\alpha_L(\mathbf{M}_{k+1} \bmod \boldsymbol{\ell}_i)}{\ell_{i+1} \cdots \ell_k} = 0$$

with the coefficient of $M_1 \bmod \ell_1, \ldots, \ell_i$ continuing to remain nonzero. Eventually, we obtain $H_k := F_k \cdot \alpha_L(\mathbf{M}_{k+1} \bmod \boldsymbol{\ell}_k) = 0$ while the coefficient of $M_1 \bmod \boldsymbol{\ell}_k$ is nonzero. This implies that $\alpha_L(\mathbf{M}_{k+1}) = 0 \bmod \boldsymbol{\ell}_k$ is a nontrivial equation, and Lemma 5.3 asserts that this is not possible unless $2k + 1 \geq n$ or $k \geq (n-1)/2$. ∎

5.4. Proofs of the technical lemmas.

LEMMA 5.1 (restated). *Suppose* $\operatorname{Det}_n = C(T_1, \ldots, T_m)$*, where C is any circuit and let $\mathbf{T}_r = \{T_1, \ldots, T_r\}$ be a transcendence basis of $\mathbf{T}$ with $r < n$. Then, there exist a set of $r+1$ variables $\mathbf{x}_{r+1} \subset \mathbf{x}$ and an equation $\sum_{i=1}^{r+1} c_i f_i \cdot M_i = 0$ such that M_i's are distinct first order principal minors of M, f_i's are distinct $r \times r$ minors of $\mathcal{J}_{\mathbf{x}_{r+1}}(\mathbf{T}_r)$, not all f_i's are zero, and $c_i \in \mathbb{F}^*$.*

Proof. In a column of a Jacobian matrix $\mathcal{J}_{\mathbf{x}}(\cdot)$, all the entries are differentiated with respect to a variable x, we will say that the column is *indexed* by x. Let $\mathbf{T}_r = \{T_1, \ldots, T_r\}$ be a transcendence basis of $\mathbf{T}$. Amongst the nonzero $r \times r$ minors of $\mathcal{J}_{\mathbf{x}}(\mathbf{T}_r)$ (they exist by Jacobian criterion), pick one (call the matrix associated with the minor, N) that maximizes the number of diagonal variables $\{x_{ii} : 1 \leq i \leq n\}$

indexing the columns of N. Let S denote the set of variables indexing the columns of N. Since $r < n$, there exists a diagonal variable $x_{jj} \notin S$. Consider the $(r+1) \times (r+1)$ minor of $\mathcal{J}_{\mathbf{x}}(\{\mathrm{Det}_n\} \cup \mathbf{T}_r)$ corresponding to the columns indexed by $S' := S \cup \{x_{jj}\}$—call the associated $(r+1) \times (r+1)$ matrix $\tilde{N}$. Since, $\mathrm{Det}_n = C(\mathbf{T})$, the polynomials Det_n and $T_1, \ldots, T_r$ are algebraically dependent and hence $\det(\tilde{N}) = 0$. Expanding $\det(\tilde{N})$ along the first row of $\tilde{N}$, which contains signed first order minors (cofactors) of M, we have an equation $\sum_{i=1}^{r+1} c_i f_i M_i = 0$, where M_i's are distinct minors of M, f_i's are distinct $r \times r$ minors of $\mathcal{J}_{S'}(\mathbf{T}_r)$, and $c_i \in \mathbb{F}^*$. If M_i is the principal minor of M with respect to the variable x_{jj}, then $f_i = \det(N) \neq 0$ (by construction).

It suffices to show that if M_i is a nonprincipal minor of M, then $f_i = 0$. Consider any nonprincipal minor M_i in the above sum, say it is the minor of M with respect to $x_{k\ell}$. The corresponding f_i is precisely the $r \times r$ minor of $\mathcal{J}_{S'}(\mathbf{T}_r)$ with respect to the columns $S' \setminus \{x_{k\ell}\} = (S \setminus \{x_{k\ell}\}) \cup \{x_{jj}\}$. Hence, by the maximality assumption on the number of diagonal elements of M in S, $f_i = 0$. ∎

Lemma 5.2 (restated). *If $M_1, \ldots, M_t$ are distinct first order principal minors of M and $\sum_{i=1}^{t} f_i \cdot M_i = 0$ (not all f_i's are zero), then the total sparsity of the f_i's is at least $2^{n/2-t}$.*

Proof. The proof is by contradiction. The idea is to start with the equation $\sum_{i=1}^{t} f_i M_i = 0$ and apply two steps—*sparsity reduction* and *fanin reduction*. We shall apply these two steps alternately until we arrive at an equation $f_j \cdot M_j = 0$, where neither f_j nor M_j is zero. We shall show that this would always be possible if the total sparsity of the f_i's is less than $2^{n/2-t}$.

With an equation of the form $\sum_{i=1}^{\tau} g_i N_i = 0$, we associate four parameters τ, $\mathcal{S}$, η, and c. These parameters are as follows: τ is called the *fanin* of the equation, $\mathcal{S}$ is the total sparsity of the g_i's (we always assume that not all the g_i's are zero), every N_i is a distinct first order principal minor of a symbolic $\eta \times \eta$ matrix $N = (x_{ij})$, and c is the maximum number of entries of N that are set as constants. To begin with, $g_i = f_i$ and $N_i = M_i$ for all $1 \le i \le t$, so $\tau = t$, $\mathcal{S} = s$ (the total sparsity of the f_i's), $\eta = n$, $N = M$, and $c = 0$. In the "sparsity reduction" step, we start with an equation $\sum_{i=1}^{\tau} g_i N_i = 0$, with parameters τ, $\mathcal{S}$, η, c and arrive at an equation $\sum_{i=1}^{\tau'} g'_i N'_i = 0$ with parameters τ', $\mathcal{S}'$, η', c' such that $\tau' \le \tau$, $\mathcal{S}' \le \mathcal{S}/2$, $\eta - 1 \le \eta' \le \eta$, and $c' \le c + 1$. In the "fanin reduction" step, we start with an equation $\sum_{i=1}^{\tau} g_i N_i = 0$, with parameters τ, $\mathcal{S}$, η, c and arrive at an equation $\sum_{i=1}^{\tau'} g'_i N'_i = 0$ with parameters τ', $\mathcal{S}'$, η', c' such that one of the two cases happens—Case 1: $\tau' \le \tau - 1$, $\mathcal{S}' \le \mathcal{S}$, $\eta' = \eta - 1$, and $c' = c$; Case 2: $\tau' = 1$, $\mathcal{S}' \le \mathcal{S}$, $\eta' = \eta$, and $c' \le c + \tau$.

Naturally, starting with $\sum_{i=1}^{t} f_i M_i = 0$, the sparsity reduction step can only be performed at most $\log s$ many times (since the total sparsity of the g_i's reduces by at least a factor of half every time this step is executed), whereas the fanin reduction step can be performed at most $t-1$ times (as the fanin goes down by at least one for every such step). Finally, when this process of alternating steps ends, we have an equation of the form $g_i \cdot N_i = 0$ (Case 2 of the fanin reduction step), where $g_i \neq 0$ and N_i is a principal minor of a symbolic matrix N of dimension at least $n-(\log s+t-1)$ such that at most $(\log s+t)$ entries of N are set as constants. Now, if $\log s+t \le n-(\log s+t)$ the N_i can never be zero (by Fact 5.8) and hence we arrive at a contradiction. Therefore, $s > 2^{n/2-t}$. Now, we give the details of the sparsity reduction and the fanin reduction steps.

Suppose, we have an equation $\sum_{i=1}^{\tau} g_i N_i = 0$ as mentioned above. We may assume that no variable x divides all the g_i's as we can divide the above equation by x if that were the case. Without loss of generality, assume that the minor N_i is the minor of N with respect to the ith diagonal element of N. Call all the variables x_{ij} in N with both $i, j > \tau$ as the *white variables*. These are the variables that are present in every minor N_i in the sum $\sum_{i=1}^{\tau} g_i N_i$. The variables x_{ij}, where both $i, j \le \tau$ are called the *black variables*, and the remaining are the *gray variables*. By assumption, c of the variables in N are set as constants.

CLAIM 5.4 (sparsity reduction). *Suppose we have an equation $\sum_{i=1}^{\tau} g_i N_i = 0$, where each N_i is a minor of a symbolic $\eta \times \eta$ matrix with at most c variables set to constants and the sum of the sparsities of the g_i's is $s > 1$. If one of the g_i's depends on a white variable, then we can derive an equation of the form*

$$\sum_{i=1}^{\tau'} g_i' N_i' = 0,$$

where $\tau' \le \tau$, each N_i' is a minor of an $\eta' \times \eta'$ matrix for $\eta - 1 \le \eta' \le \eta$ with at most $c' \le c+1$ variables set to constants. Furthermore, the sum of the sparsities of the g_i's is at most $s/2$.

Proof of Claim 5.4. Say x is a white variable that one of the g_i's depends on. Writing each g_i as a polynomial in x. Note that x cannot divide all the g_i's.

Each of the g_i's and N_i's can be expressed as, $g_i = g_{i,0} + x \cdot g_{i,1} + \cdots + x^h \cdot g_{i,h}$ and $N_i = N_{i,0} + x \cdot N_{i,1}$, where $g_{i,j}$'s and $N_{i,j}$'s are x-free. This is possible as x is a white variable and it occurs in every N_i.

Looking at the coefficients of x^0 and x^{h+1} in the equation yields $\sum_{i=1}^{\tau} g_{i,0} \cdot N_{i,0} = 0$ and $\sum_{i=1}^{\tau} g_{i,h} \cdot N_{i,1} = 0$. Note that $N_{i,0}$'s can be thought of as principal minors of the $\eta \times \eta$ matrix N' obtained by setting $x = 0$ in N. And each of the $N_{i,1}$'s can be thought of as minors of the $(\eta-1) \times (\eta-1)$ matrix N' which is the matrix associated with the minor of N with respect to x. Since the monomials in $g_{i,0}$ and $x^h g_{i,h}$ are disjoint, either the total sparsity of the $g_{i,0}$'s or the total sparsity of the $g_{i,h}$'s is $\le \mathcal{S}/2$. Thus, one of the equations $\sum_{i=1}^{\tau} g_{i,0} \cdot N_{i,0} = 0$ or $\sum_{i=1}^{\tau} g_{i,h} \cdot N_{i,1} = 0$ yields an equation of the form $\sum_{i=1}^{\tau'} g_i' N_i' = 0$ with parameters τ', $\mathcal{S}'$, η', c' as claimed before. (In case, we choose $\sum_{i=1}^{\tau} g_{i,h} \cdot N_{i,1} = 0$ as our next equation, we also set the variables in the same columns and rows of x to constants in such a way that a $g_{i,h}$ stays nonzero. This is certainly possible over a characteristic zero field [Sch80, Zip79, DL78].) ∎

The sparsity reduction step is performed whenever the starting equation $\sum_{i=1}^{\tau} g_i N_i = 0$ has a white variable among the g_i's. When all the g_i's are free of white variables, we perform the *fanin reduction step*.

CLAIM 5.5 (fanin reduction). *Suppose we have an equation $\sum_{i=1}^{\tau} g_i N_i = 0$, where each N_i is a minor of a symbolic $\eta \times \eta$ matrix with at most c variables set to constants and the sum of the sparsities of the g_i's is $s > 1$. If none of the g_i's depend on a white variable, then we can derive an equation of one of the two forms:*

- *$g' \cdot N' = 0$ for some minor N' of an $\eta \times \eta$ symbolic matrix with at most $c + \tau$ entries set to constants and $g' \ne 0$;*
- *$\sum_{i=1}^{\tau'} g_i' N_i' = 0$, where $\tau' \le \tau - 1$, each N_i' is a minor of an $\eta' \times \eta'$ matrix for $\eta - 1 \le \eta' \le \eta$ with at most $c' \le c + 1$ variables set to constants, and the sum of sparsities of g_i's bounded by s.*

Proof of Claim 5.5. When we perform this step, all the g_i's consist of black and gray variables. Pick a row R from N barring the first τ rows. Let $y_1, \ldots, y_\tau$ be the gray variables occurring in R (these are, respectively, the variables in the first τ columns of R). We shall first do some preprocessing to ensure that some g_i is nonzero when $y_2 = y_3 = \cdots = y_\tau = 0$. This may not always be true, but we shall slightly modify the equation to enforce this property.

Starting with y_2, divide the equation $\sum_{i=1}^{\tau} g_i N_i = 0$ by the largest power of y_2 common across all monomials in the g_i's, and then set $y_2 = 0$. With some abuse in notation, let us call the residual equation also $\sum_{i=1}^{\tau} g_i N_i$. This process lets us assume that there exists at least one g_i which is not zero at $y_2 = 0$. On the residual equation, repeat the same process with y_3 and then with y_4 and so on till y_τ. Thus, in the residual equation (again, abusing notation and using the same symbols), $\sum_{i=1}^{\tau} g_i N_i = 0$ there is at least one g_i that is not zero when $y_2, \ldots, y_\tau$ are set to zero.

Suppose that exactly one g_i stays nonzero under the projection $y_2 = \cdots = y_\tau = 0$, then $(g_i N_i)_{(y_2=\cdots=y_\tau=0)} = 0$. This is the first form of the derived equation as claimed.

Now, assume that there are at least two g_i's (say g_1 and g_2) that are nonzero under the projection $y_2 = \cdots = y_\tau = 0$. Set all the remaining variables of row R to zero except y_1—these are the white variables in R. Since the g_i's are free of white variables (or else, we would have performed the sparsity reduction step), none of the g_i's are effected by this projection. However, N_1 being a minor with respect to the first diagonal element of N, vanishes completely after the projection. Any other N_i takes the form $y_1 \cdot N_i'$, where N_i' is a principal minor of an $(\eta-1)\times(\eta-1)$ matrix N' which is the matrix associated with the minor of N with respect to y_1. Therefore, after the projection, the equation $\sum_{i=1}^{\tau} g_i N_i = 0$ becomes $\sum_{i=2}^{\tau} \tilde{g}_i \cdot y_1 N_i' = 0 \Rightarrow \sum_{i=2}^{\tau} \tilde{g}_i \cdot N_i' = 0$, where $\tilde{g}_i$ is the image of g_i under the above-mentioned projection and further $\tilde{g}_2 \neq 0$. The $\tilde{g}_i$'s might still contain variables from the first column of N. So, as a final step, set these variables to values so that a nonzero $\tilde{g}_i$ remains nonzero after this projection (the [Sch80, Zip79, DL78] lemma asserts that such values exist in plenty). This gives us the desired form $\sum_{i=1}^{\tau'} g_i' N_i' = 0$ with parameters τ', $\mathcal{S}'$, η', c' as claimed. ∎

These two claims together complete the proof of the lemma. ∎

LEMMA 5.3 (restated). *If $M_1, \ldots, M_t$ are distinct first order principal minors of M and $\sum_{i=1}^{t} \alpha_i M_i = 0 \bmod \boldsymbol{\ell}_k$ (not all $\alpha_i = 0$) for independent linear polynomials $\boldsymbol{\ell}_k$, then $t + k \geq n$.*

Proof. Assume that $t + k < n$ (with $t \geq 1$ it means $k \leq n-2$) and $\alpha_i M_i = 0 \bmod \boldsymbol{\ell}_k$. Recall that reducing an expression modulo $\ell = c_1 x_1 + \sum_{i>1} c_i x_i$ (with $c_1 \neq 0$) is equivalent to replacing x_1 by $\sum_{i>1}(-c_i/c_1)x_i$. Hence, as $\ell_1, \ldots, \ell_k$ are independent linear polynomials, the equation may be rewritten as $\sum_{i=1}^{t} \alpha_i M_i' = 0$, where the (M_i')s are minors of the matrix M' obtained by replacing k entries $(x_1, \ldots, x_k)$ of M by linear polynomials in other variables. We shall call these entries as *corrupted* entries. Without loss of generality, we shall assume that M_i' is the minor corresponding to the ith diagonal variable and that all the α_i's are nonzero.

CLAIM 5.6. *Each of the first t rows and columns of M must have a corrupted entry.*

Proof of Claim 5.6. Suppose the first row (without loss of generality) is free of any corrupted entry. Then, setting the entire row to zero would make all $M_i' = 0$ for $i \neq 1$. But since $\sum \alpha_i M_i' = 0$, this forces M_1' to become zero under the projection as well. This leads to a contradiction as M_1' is a determinant of an $(n-1)\times(n-1)$ symbolic matrix under a projection, and this cannot be zero unless $k \geq n-1$ (by Fact 5.8). ∎

Since $n - k > t$, there must exist a set of $t-1$ rows (actually we would have $t+1$ such rows, but $t-1$ would suffice), say $\{R_{i_1}, \ldots, R_{i_{t-1}}\}$, of M that are free of any corrupted entries. Note that by the claim above, none of these rows are the first t rows of M.

For each of these rows, set the jth variable of row R_{i_j} to 1, and every other variable in $R_1, \ldots, R_{t-1}$ to zero. That is, among the rows $R_{i_1}, \ldots, R_{i_{t-1}}$, the only nonzero entries are $\{x_{i_1,1}, x_{i_2,2}, \ldots, x_{i_{t-1},t-1}\}$. This projections make $M'_i = 0$ for all $i \neq t$ (as in these minors an entire row vanishes). And since we had $\sum_{i=1}^{t} \alpha_i M'_i = 0$ to begin with, this forces M'_t to become zero under this projection as well.

But let us take a moment to see what M'_t reduces to under this projection. The minor M'_t just reduces (up to a sign) to the minor obtained from M' by removing the columns $\{1, \ldots, t\}$ and rows $\{R_{i_1}, \ldots, R_{i_{t-1}}\} \cup \{t\}$. This is a determinant of an $(n-t) \times (n-t)$ symbolic matrix, containing at most $k-t$ corrupted entries, thus $k - t \geq n - t$ (by Fact 5.8). But then $k \geq n$, which contradicts our initial assumption. ∎

5.5. Extensions to immanants. All the lower bound proofs use some very basic properties of Det_n. These properties are general enough that they apply to any *immanant*. For any map $\chi : S_n \to \mathbb{C}^{\times}$, recall the definition of the immanant of an $n \times n$ matrix $M = (x_{ij})$:

$$\mathrm{Imm}_{\chi}(M) = \sum_{\sigma \in S_n} \chi(\sigma) \prod_{i=1}^{n} x_{i,\sigma(i)}.$$

From the image of the map χ it is obvious that $\chi(\sigma) \neq 0$ for any $\sigma \in S_n$.

DEFINITION 5.7. *The* minor *of $Imm_{\chi}(M)$ with respect to the (i,j)th entry is defined as*

$$(Imm_{\chi}(M))_{i,j} = \sum_{\substack{\sigma \in S_n \\ \sigma(i)=j}} \chi(\sigma) \prod_{k \neq i} x_{k,\sigma(k)}.$$

This may also be rewritten as $\mathrm{Imm}_{\chi'}(M_{ij})$ for a suitable map $\chi' : S_{n-1} \to \mathbb{C}^{\times}$, where M_{ij} is the submatrix of M after removing the ith row and jth column. From the definition, it follows directly that the partial derivative of $\mathrm{Imm}_{\chi}(M)$ with respect to x_{ij} is precisely the minor with respect to (i,j).

The only crucial fact of determinants that is used in all the proofs is that a symbolic $n \times n$ determinant cannot be zero when less than n of its entries are altered.

FACT 5.8. *Let M' be the matrix obtained by setting $c < n$ entries of M to arbitrary polynomials in $\mathbb{F}[\mathbf{x}]$. Then for any map $\chi : S_n \to \mathbb{C}^{\times}$, we have $Imm_{\chi}(M') \neq 0$.*

Proof. We shall say an entry of M' is *corrupted* if it is one of the c entries of M that has been replaced by a polynomial. We shall prove this by carefully rearranging the rows and columns so that all the corrupted entries are above the diagonal. Then, since all entries below the diagonal are free, we may set all of them to zero and the immanant reduces to a single nonzero monomial.

Since less than n entries of M' have been altered, there exists a column that is free of any corrupted entries. By relabeling the columns if necessary, let the first column be free of any corrupted entry. Pick any row R that contains a corruption and relabel the rows to make this the first row. This ensures that the first column is free of any corrupted entry, and the $(n-1) \times (n-1)$ matrix defined by rows and columns, 2 through n, contain less than $c - 1$ corruptions. By induction, the $c - 1$ corruptions may be moved above the diagonal by suitable row/column relabeling. And since the first column is untouched during the process, we now have all c corruptions above the diagonal. Now setting all entries below the diagonal to zeroes reduces the immanant to a single nonzero monomial. ∎

With this fact, all our lower bound proofs of the determinant can be rewritten for any immanant.

6. Conditional lower bounds for depth-D occur-k formulas. In this section, we present a lower bound for depth-D occur-k formulas similar in spirit to Theorem 1.6 by assuming the following conjecture about determinant minors.

CONJECTURE 6.1. *Let $M = (x_{ij})$ be an $n \times n$ matrix, and let x_i denote the ith diagonal variable x_{ii}. Let M' be a projection of M by setting $c = o(n)$ of the variables in M to constants. Suppose the elements $x_1, \ldots, x_k$, where k is a constant independent of n, are partitioned into nonempty sets $S_1, \ldots, S_t$. Consider $\mathcal{M}(\mathbf{S}_t)$, the set of tth order principal minors of M', each by choosing a t-tuple $B \in S_1 \times \cdots \times S_t$ as pivots. Over all possible choices of B, we get $m := |S_1| \cdots |S_t|$ many minors. Then for any set of diagonal variables $\mathbf{y}_m$ disjoint from $\mathbf{x}_k$, $J_{\mathbf{y}_m}(\mathcal{M}(\mathbf{S}_t)) \neq 0$.*

The conjecture roughly states that the different tth order principal minors are algebraically independent. We will need a generalization of Lemma 5.1 for the purposes of this section.

LEMMA 6.2. *Let $\{f_1, \ldots, f_s, g_1, \ldots, g_t\}$ be algebraically dependent polynomials such that* $\operatorname{trdeg}\{\mathbf{g}_t\} = t$. *Let $\Gamma \subseteq \mathbf{x}$ be a fixed set of variables of size at least $s + t$. Then there exists a set of $s + t$ variables $\mathbf{x}_{s+t} \subset \mathbf{x}$ and an equation of the form*

$$\sum_{i=1}^{r} c_i \cdot F_i \cdot G_i = 0, \qquad \textit{where } r \leq \binom{s+t}{t}$$

such that each $c_i \in \mathbb{F}^*$, each F_i is a distinct $s \times s$ minor of $\mathcal{J}_{\mathbf{x}_{s+t} \cap \Gamma}(\mathbf{f}_s)$, each G_i is a distinct $t \times t$ minor of $\mathcal{J}_{\mathbf{x}_{s+t}}(\mathbf{g}_t)$, and not all G_i's are zero.

Note that we are not asserting the nonzeroness of F_i's. Also, Lemma 5.1 may be obtained from the above lemma by taking $f_1 = \mathrm{Det}_n$, $s = 1$, and Γ to be the set of diagonal variables.

Proof. The proof is along the lines of Lemma 5.1. Amongst the nonzero $t \times t$ minors of $\mathcal{J}_{\mathbf{x}}(\mathbf{g}_t)$, pick one (call the matrix associated with the minor, N) that maximizes the number of variables in Γ indexing the columns of N. Without loss of generality, let $\mathbf{x}_t = \{x_1, \ldots, x_t\}$ be the set of variables indexing the columns of N. Since $|\Gamma| \geq s + t$, there exist s other variables in Γ, say $\{x_{1+t}, \ldots, x_{s+t}\}$. Consider the $(s+t) \times (s+t)$ minor of $\mathcal{J}_{\mathbf{x}}(\mathbf{f}_s \cup \mathbf{g}_t)$ corresponding to the columns indexed by $\mathbf{x}_{s+t}$—call the associated $(s+t) \times (s+t)$ matrix $\tilde{N}$.

Since $\mathbf{f}_s, \mathbf{g}_t$ are algebraically dependent, $\det(\tilde{N}) = 0$. Expanding $\det(\tilde{N})$ over all possible $s \times s$ minors in the first s rows, we have an equation

$$\sum_{U \subseteq \mathbf{x}_{s+t}, |U| = s} c_i \cdot F_U \cdot G_U = 0,$$

where each F_U is an $s \times s$ minor of $\mathcal{J}_{\mathbf{x}_{s+t}}(\mathbf{f}_s)$ with respect to the variables in U, and each G_U the $t \times t$ minor of $\mathcal{J}_{\mathbf{x}_{s+t}}(\mathbf{g}_t)$ corresponding to $\overline{U}$, and $c_i \in \mathbb{F}^*$. If G_U is the minor with respect to variables $\mathbf{x}_t$, then $G_U = \det(N) \neq 0$ (by construction). It suffices to show that if F_U is a minor indexed by variables outside Γ, then $G_U = 0$. This follows, just like in Lemma 5.1, by the maximality assumption on choice of $\mathbf{x}_t$ which we elaborate on now.

Consider some set $U \subset \mathbf{x}_{s+t}$ that contains some $x_i \notin \Gamma$. Then, $\overline{U}$ contains at least one more element of Γ than $\mathbf{x}_t$. Hence, by the choice of $\mathbf{x}_t$, we must have that the Jacobian of $\mathbf{g}_t$ with respect to $\overline{U}$ is singular, i.e., $G_U = 0$. ∎

The rest of this section shall be devoted to the proof of the following theorem.

THEOREM 6.3. *Assuming Conjecture* 6.1, *any depth-D occur-k formula that computes Det_n must have size $s = 2^{\Omega(n)}$ over any field of characteristic zero.*

Proof idea. The proof proceeds on the same lines as Theorem 1.6. If $T_1, \ldots, T_k$ is a transcendence basis of gates at level 2 computing the determinant, then $\mathcal{J}_{\mathbf{x}}(\mathrm{Det}_n, T_1, \ldots, T_k)$ is a matrix of rank k. This yields a nontrivial equation of the form $\sum N_i^{(1)} \cdot G_i^{(1)} = 0$ where each of the $N_i^{(1)}$'s are principal minors of $M = (x_{ij})$ and $G_i^{(1)}$'s are $k \times k$ minors of $\mathcal{J}_{\mathbf{x}}(T_1, \ldots, T_k)$. Here is where we may use Lemma 4.4 to remove common factors and obtain an equation of the form $\sum N_i^{(1)} \cdot \tilde{G_i}^{(1)} = 0$, where $\tilde{G_i}^{(1)}$ is a polynomial of constantly many derivatives of polynomials computed at the next level. The above equation may be thought of as a polynomial relation amongst $\{N_i^{(1)}\} \cup \{\mathrm{Elem}(\tilde{G_i}^{(1)})\}$. Applying Lemma 6.2 (with a suitable choice of S_2), we get an equation of the form $\sum N_i^{(2)} \cdot G_i^{(2)} = 0$, where each $N_i^{(2)}$ is a minor of $\mathcal{J}_{S_2}(\{N_i^{(1)}\})$, and the $G_i^{(2)}$s are Jacobians minors of $\bigcup \mathrm{Elem}(\tilde{G_i}^{(1)})$. Again after removing common factors, this equation may be interpreted as a polynomial relation amongst the entries of $N_i^{(2)}$ (which are minors of order 2) and $\mathrm{Elem}(\tilde{G_i}^{(2)})$.

Repeating this argument, we finally reach the level of sparse polynomials and thus obtain a nontrivial equation $\sum N_i^{(D-2)} \cdot \tilde{G_i}^{(D-2)} = 0$, where each $N_i^{(D-2)}$ is a Jacobian minor of $(D-3)$-order minors, and each $\tilde{G_i}^{(D-2)}$ is a sparse polynomial. With a slightly more careful choice of the sets S_i in Lemma 6.2, each of the minors $N_i^{(D-2)}$ would be a minor of $\mathcal{J}_{S_{D-2}}(\mathcal{M}(S_1, \ldots, S_{D-3}))$. Assuming Conjecture 6.1, we can show that such an equation is not possible unless the sparsity of the f_i's is large, using a similar argument as in Lemma 5.2.

LEMMA 6.4. *Suppose Det_n is computed by a depth-D occur-k formula of size s. Then there exist variables $x_1, \ldots, x_R$, where $R = R(k, D)$, a partition of $\mathbf{x}_R$ into nonempty sets $S_1, \ldots, S_{D'}$, $(D' \leq (D-2))$ polynomials $f_1, \ldots, f_m$ (not all zero), where $m = |\mathcal{M}(\mathbf{S}_D)|^{O(R)}$ and each f_i has sparsity at most s^R, such that*

$$\sum_{i=1}^{m} f_i \cdot N_i = 0,$$

where each N_i is a minor of $\mathcal{J}_{\mathbf{x}}(\mathcal{M}(\mathbf{S}_{D'}))$ indexed by diagonal variables.

Proof. To begin with, suppose $\mathrm{Det}_n = C(T_1, \ldots, T_m)$, where $T_1, \ldots, T_m$ are polynomials computed at the first level. So Lemma 5.1 gives a starting equation, though we do not really have a sparsity bound on the f_i's. The proof shall proceed by transforming this equation into another, involving lower level polynomials, till we get a sparsity bound.

In a general step, we would have an equation of the form

$$C_\ell(\mathcal{M}(S_1, \ldots, S_{\ell-1}), T_1^{(\ell)}, \ldots, T_{r_\ell}^{(\ell)}) = 0,$$

where each $T_i^{(\ell)}$, is a derivative (of order at most ℓ) of a polynomial computed at level ℓ of the circuit. Without loss of generality, we may assume that $\{T_1^{(\ell)}, \ldots, T_{r_\ell}^{(\ell)}\}$

are algebraically independent. Let $m_\ell := |S_1| \cdots |S_{\ell-1}|$. Choose a set of diagonal elements S_ℓ of size $|\mathcal{M}(S_1,\ldots,S_{\ell-1})|+r_\ell$ that is disjoint from $S_1,\ldots,S_{\ell-1}$. Applying Lemma 6.2 with $\Gamma = S_\ell$, we get an equation of the form

$$\sum_i c_i^{(\ell)} N_i^{(\ell)} \cdot G_i^{(\ell)} = 0,$$

where $N_i^{(\ell)}$ is an $m_\ell \times m_\ell$ minor of $\mathcal{J}_{S_\ell}(\mathcal{M}(\mathbf{S}_{\ell-1}))$ indexed by diagonal variables, each $G_i^{(\ell)}$ is an $r_\ell \times r_\ell$ minor of $\mathcal{J}_{S_\ell}(\mathbf{T}^{(\ell)}_{r_\ell})$. Consider the matrix $\mathcal{J}_{S_\ell}(\mathbf{T}^{(\ell)}_{r_\ell})$ restricted to the columns appearing in some $G_i^{(\ell)}$. Applying Lemma 4.4 on this matrix, we can write each $G_i^{(\ell)} = V_\ell \cdot \tilde{G_i}^{(\ell)}$, where each $\tilde{G_i}^{(\ell)}$ is a polynomial function of at most $r_{\ell+1} := (\ell+1)2^{\ell+1} \cdot k(r_\ell + m_\ell)r_\ell$ many derivatives of polynomials computed at level $\ell+1$, and V_ℓ is the part that comes out common from the rows of the Jacobian after applying the gcd trick of Lemma 4.3. Thus,

$$V_\ell \cdot \sum_i c_i^{(\ell)} N_i^{(\ell)} \cdot \tilde{G_i}^{(\ell)} = 0.$$

Note that V_ℓ cannot be zero as at least one $G_i^{(\ell)}$ was guaranteed to be nonzero by Lemma 6.2. Therefore, $\sum_i c_i^{(\ell)} N_i^{(\ell)} \cdot \tilde{G_i}^{(\ell)} = 0$. Since each $\tilde{G_i}^{(\ell)}$ is a polynomial function of $r_{\ell+1}$ derivatives at the next level, we now have

$$C_{\ell+1}(\mathcal{M}(S_1,\ldots,S_\ell), T_1^{(\ell+1)}, \ldots, T_{r_{\ell+1}}^{(\ell+1)}) = 0.$$

Unfolding this recursion, we finally reach the level of sparse polynomials, at which point we have an equation of the form

$$\sum_i c_i^{(D-2)} N_i^{(D-2)} \cdot \tilde{G_i}^{(D-2)} = 0$$

and each $\tilde{G_i}^{(D-2)}$ is an $r_{D-2} \times r_{D-2}$ Jacobian minor of sparse polynomials. Hence, each $\tilde{G_i}^{(D-2)}$ is itself a polynomial of sparsity bounded by $s^{r_{D-2}}$ as claimed. ∎

We now have to show that an equation of the form $\sum f_i \cdot N_i = 0$ is not possible unless one of the f_i's has exponential sparsity. The above lemma guarantees that at least one of the f_i's is nonzero in this equation, but it could be the case that some of the N_i's are zero. This was not the case in the depth-4 lower bound as each N_i was just a determinant minor. However, in this case they are Jacobians of minors. Conjecture 6.1 asserts that the N_i's are nonzero, even if "few" variables are set to zero. This assumption is enough to get the required lower bound.

LEMMA 6.5. *Let $|\mathcal{M}(S_1,\ldots,S_D)| =: m$ be a constant and let $\{N_i\}_{i\le t}$ be distinct $m \times m$ minors of $\mathcal{J}_{\mathbf{x}}(\mathcal{M}(\mathbf{S}_D))$, where the columns of N_i are indexed by a set T_i of diagonal variables of M disjoint from $\bigcup_{j=1}^{D} S_j$. Suppose $f_1,\ldots,f_t$ are polynomials such that $\sum_{i=1}^{t} f_i \cdot N_i = 0$ (not all f_i's are zero). Then, assuming Conjecture* 6.1 *is true, the total sparsity of the f_i's is $2^{\Omega(n)}$.*

Proof. The proof is along lines similar to the proof of Lemma 5.2 and shall proceed by a similar series of *sparsity reduction* and *fanin reduction* steps to arrive at a contradiction. Throughout the proof, Conjecture 6.1 shall assert that the N_i's stay nonzero (even when few variables are set to constants). We briefly describe the *sparsity reduction* and the *fanin reduction* steps and the rest of the proof would follow in essentially an identical fashion to the proof of Lemma 5.2.

Without loss of generality, assume that $\{x_1, \ldots, x_r\}$ is the union of the sets S_i's and T_i's. Let N refer to the matrix of indeterminates that the N_i's are derived from. In our case, N would be obtained by (possibly) setting a few variables to constants in $M = (x_{ij})$. We'll refer to all the variables x_{ij}, where both $i, j > r$ as white variables; these are present in every entry of each N_i. The variables x_{ij} where both $i, j \leq r$ shall be called black variables, and the rest called gray variables. Here again, the *sparsity reduction* step shall be applied whenever one of the f_i's depends on a *white* variable, otherwise the *fanin reduction* steps shall be applied.

Sparsity reduction step. Suppose one of the f_i's depends on a white variable x. Then each N_i can be written as $N_i = N_{i,0} + \cdots + x^m N_{i,m}$, and $f_i = f_{i,0} + \cdots + f_{i,h}x^h$. One of the two equations corresponding to the coefficient of x^0 and x^{h+m} yields a similar equation with sparsity reduced by a factor of $1/2$. Observe that $N_{i,0}$ is just $N_i\mid_{x=0}$, and hence the polynomials $\{N_{i,0}\}$ may be thought of as corresponding Jacobian minors of N' obtained by setting $x = 0$ in N'. Also, $N_{i,m}$ is obtained by replacing every entry of the matrix corresponding to N_i by its minor with respect to x. And hence, $N_{i,m}$ can be thought of as a corresponding Jacobian minor of N_x obtained by taking the minor of N with respect to x. Thus the two equations corresponding to the coefficient of x^0 and x^{h+m} are indeed of the same form as $\sum f_i N_i = 0$. (In the case of the coefficient of x^{h+m}, we need to set other variables in the row/column containing x as in the proof of Lemma 5.2.)

Fanin reduction step. Without loss of generality, let $x_1 \in T_1 \setminus T_2$. Pick a row R of N barring the first r rows, and let $y_1, \ldots, y_r$ be the gray variables in R (where y_1 is in the same column as x_1). By a similar process as in the proof of Lemma 5.2, we can assume that at least one f_i is nonzero when $y_2, \ldots, y_r$ are set to zero.

If one of the f_i's becomes zero when $y_2, \ldots, y_r = 0$, then pick any white variable y in row R and set every variable in row R to zero besides y. This would ensure that the fanin of the equation reduces and each N_i is now $y^m \cdot N_i'$. Each N_i' may be thought of as being obtained from N_y, the minor of N with respect to y. The other variables in the column of y can be set to values to ensure that the f_i's stay nonzero to obtain an equation of the form $\sum f_i' N_i' = 0$ of the reduced fanin.

If none of the f_i's becomes zero when $y_2, \ldots, y_r = 0$, then set every variable in row R other than y_1 to zero. This ensures an entire column of the matrix corresponding to N_1 becomes zero (as x_1 indexes one of the columns of N_1), and hence N_1 becomes zero. On the other hand, N_2 remains nonzero and each surviving N_i can be written as $y_1^m \cdot N_i'$, where N_i' is the corresponding Jacobian minor of N_{y_1}. Again, the other variables in the column of y_1 can be set to values to ensure that f_i's stay nonzero and we obtain an equation $\sum f_i' N_i' = 0$ of the reduced fanin.

As in the proof of Lemma 5.2, we eventually obtain an equation of the form $f_1 N_1 = 0$, where $f_1 \neq 0$ thus implying that $N_1 = 0$. The number of variables that have been set to constants is bounded by $t + \log S$, where S is the initial total sparsity of the f_i's, and N_1 is a Jacobian minor of a symbolic matrix of dimension $n - (\log S + t - 1)$. Conjecture 6.1 asserts that N_1 would be nonzero unless $\log S + t = \Omega(n - (\log S + t - 1))$ or $S = 2^{\Omega(n)}$. ∎

That concludes the proof of Theorem 6.3 as well.

7. Conclusion. Spurred by the success of the Jacobian in solving the hitting-set problem for *constant-trdeg* depth-3 circuits and *constant-occur* constant-depth formulas, one is naturally inspired to investigate the strength of this approach against other "constant parameter" models—the foremost of which is *constant top fanin* depth-4 circuits (PIT even for fanin 2?).

Another problem, which is closely related to hitting sets and lower bounds, is reconstruction of arithmetic circuits [SY10, Chapter 5]. There is a quasi-polynomial time reconstruction algorithm [KS09a], for a polynomial computed by a depth-3 constant top fanin circuit, that outputs a depth-3 circuit with quasi-polynomial top fanin. Could the Jacobian be used as an effective tool to solve reconstruction problems? If yes, then it would further reinforce the versatility of this tool.

Acknowledgments. We would like to thank the anonymous reviewers for several useful suggestions that have helped us improve the presentation of the paper. This work was done when MA, RS, and NS visited the Max-Planck Institute for Informatics (Saarbrücken) and would like to thank the institute for its generous hospitality.

REFERENCES

[Agr05] M. Agrawal, *Proving Lower Bounds Via Pseudo-random Generators*, in FSTTCS, Springer, Berlin, 2005, pp. 92–105.

[AKS04] M. Agrawal, N. Kayal, and N. Saxena, *Primes is in P*, Ann. of Math. (2), 160 (2004), pp. 781–793.

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, *Proof verification and the hardness of approximation problems*, J. ACM, 45 (1998), pp. 501–555.

[ASS13] M. Agrawal, C. Saha, and N. Saxena, *Quasi-polynomial hitting-set for set-depth-Δ formulas*, in STOC, ACM, New York, 2013, pp. 321–330.

[ASSS12] M. Agrawal, C. Saha, R. Saptharishi, and N. Saxena, *Jacobian hits circuits: Hitting-sets, lower bounds for depth-d occur-k formulas and depth-3 transcendence degree-k circuits*, in STOC, ACM, New York, 2012, pp. 599–614.

[AV08] M. Agrawal and V. Vinay, *Arithmetic circuits: A chasm at depth four*, in FOCS, IEEE, Piscataway, NJ, 2008, pp. 67–75.

[AvM10] S. Aaronson and D. van Melkebeek, *A Note on Circuit Lower Bounds from Derandomization*, Electronic Colloquium on Computational Complexity, TR10-105, 2010, p. 105.

[AvMV11] M. Anderson, D. van Melkebeek, and I. Volkovich, *Derandomizing polynomial identity testing for multilinear constant-read formulae*, in IEEE Conference on Computational Complexity, IEEE, Piscataway, NJ, 2011, pp. 273–282.

[BMS11] M. Beecken, J. Mittmann, and N. Saxena, *Algebraic independence and blackbox identity testing*, in ICALP, IEEE, Piscataway, NJ, 2011, pp. 134–148.

[BMS13] M. Beecken, J. Mittmann, and N. Saxena, *Algebraic independence and blackbox identity testing*, Inform. and Comput., 222 (2013), pp. 2–19.

[BOT88] M. Ben-Or and P. Tiwari, *A deterministic algorithm for sparse multivariate polynominal interpolation*, in 20th Annual STOC, ACM, New York, 1988, pp. 301–309.

[CDGK91] M. Clausen, A. W. M. Dress, J. Grabmeier, and M. Karpinski, *On zero-testing and interpolation of k-sparse multivariate polynomials over finite fields*, Theoret. Comput. Sci., 84 (1991), pp. 151–164.

[CKW11] X. Chen, N. Kayal, and A. Wigderson, *Partial derivatives in arithmetic complexity (and beyond)*, Found. Trends Theor. Comput. Sci., 6 (2011), pp. 1–138.

[DL78] R. A. DeMillo and R. J. Lipton, *A probabilistic remark on algebraic program testing*, Inform. Process. Lett., 7 (1978), pp. 193–195.

[DS05] Z. Dvir and A. Shpilka, *Locally decodable codes with 2 queries and polynomial identity testing for depth 3 circuits*, in STOC, ACM, New York, 2005, pp. 592–601.

[FS13] M. A. Forbes and A. Shpilka, *Quasipolynomial-time identity testing of non-commutative and read-once oblivious algebraic branching programs*, in 54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, IEEE, Piscataway, NJ, 2013, pp. 243–252.

[GKKS13] A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi, *Arithmetic circuits: A chasm at depth three*, in 54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, IEEE, Piscataway, NJ, 2013, pp. 578–587.

[GKPS11] B. Grenet, P. Koiran, N. Portier, and Y. Strozecki, *The limited power of powering: Polynomial identity testing and a depth-four lower bound for the permanent*, in Proceedings of the 30th Foundations of Software Technology and Theoretical Computer Science (FSTTCS), Schloss Dagstuhl–Liebniz–Zentrum für Informatik, Wadern, Germany, 2011, pp. 127–139.

[GR05] A. Gabizon and R. Raz, *Deterministic extractors for affine sources over large fields*, in FOCS, IEEE Computer Society, Los Alamitos, CA, 2005, pp. 407–418.

[HS80] J. Heintz and C.-P. Schnorr, *Testing polynomials which are easy to compute (extended abstract)*, in STOC, ACM, New York, 1980, pp. 262–272.

[IKS10] G. Ivanyos, M. Karpinski, and N. Saxena, *Deterministic polynomial time algorithms for matrix completion problems*, SIAM J. Comput., 39 (2010), pp. 3736–3751.

[IW97] R. Impagliazzo and A. Wigderson, *P = BPP if E requires exponential circuits: Derandomizing the XOR lemma*, in STOC, ACM, New York, 1997, pp. 220–229.

[Kal85] K. A. Kalorkoti, *A lower bound for the formula size of rational functions*, SIAM J. Comput., 14 (1985), pp. 678–687.

[KI03] V. Kabanets and R. Impagliazzo, *Derandomizing polynomial identity tests means proving circuit lower bounds*, in STOC, ACM, New York, 2003, pp. 355–364.

[KMSV10] Z. S. Karnin, P. Mukhopadhyay, A. Shpilka, and I. Volkovich, *Deterministic identity testing of depth-4 multilinear circuits with bounded top fan-in*, in STOC, ACM, New York, 2010, pp. 649–658.

[Koi12] P. Koiran, *Arithmetic circuits: The chasm at depth four gets wider*, Theoret. Comput. Sci., 448 (2012), pp. 56–65.

[KS01] A. Klivans and D. A. Spielman, *Randomness efficient identity testing of multivariate polynomials*, in STOC, ACM, New York, 2001, pp. 216–223.

[KS06] A. R. Klivans and A. Shpilka, *Learning restricted models of arithmetic circuits*, Theory Comput., 2 (2006), pp. 185–206.

[KS07] N. Kayal and N. Saxena, *Polynomial identity testing for depth 3 circuits*, Comput. Complexity, 16 (2007), pp. 115–138.

[KS08] Z. S. Karnin and A. Shpilka, *Black box polynomial identity testing of generalized depth-3 arithmetic circuits with bounded top fan-in*, in IEEE Conference on Computational Complexity, IEEE Computer Society, Los Alamitos, CA, 2008, pp. 280–291.

[KS09a] Z. S. Karnin and A. Shpilka, *Reconstruction of generalized depth-3 arithmetic circuits with bounded top fan-in*, in IEEE Conference on Computational Complexity, IEEE Computer Society, Los Alamitos, CA, 2009, pp. 274–285.

[KS09b] N. Kayal and S. Saraf, *Blackbox polynomial identity testing for depth-3 circuits*, in FOCS, IEEE Computer Society, Los Alamitos, CA, 2009, pp. 198–207.

[KY08] S. Kopparty and S. Yekhanin, *Detecting rational points on hypersurfaces over finite fields*, in IEEE Conference on Computational Complexity, IEEE Computer Society, Los Alamitos, CA, 2008, pp. 311–320.

[LFKN90] C. Lund, L. Fortnow, H. J. Karloff, and N. Nisan, *Algebraic methods for interactive proof systems*, in FOCS, 1990, pp. 2–10.

[Lov79] L. Lovász, *On determinants, matchings, and random algorithms*, in FCT, Akademie-Verlag, Berlin, 1979, pp. 565–574.

[LR34] D. E. Littlewood and A. R. Richardson, *Group, characters and algebras*, Philos. Trans. Roy. Soc. Ser. A, 233 (1934), pp. 721–730.

[MSS14] J. Mittmann, N. Saxena, and P. Scheiblechner, *Algebraic independence in positive characteristic – A p-adic calculus*, Trans. Amer. Math. Soc., 366 (2014), pp. 3425–3450.

[Mul11] K. Mulmuley, *On P vs. NP and geometric complexity theory: Dedicated to Sri Ramakrishna*, J. ACM, 58 (2011), 5.

[Mul12] K. Mulmuley, *Geometric complexity theory v: Equivalence between blackbox derandomization of polynomial identity testing and derandomization of Noether's normalization lemma*, in FOCS, IEEE, Piscataway, NJ, 2012, pp. 629–638.

[MVV87] K. Mulmuley, U. V. Vazirani, and V. V. Vazirani, *Matching is as easy as matrix inversion*, in STOC, ACM, New York, 1987, pp. 345–354.

[Oxl92] J. G. Oxley, *Matroid Theory*, Oxford University Press, Oxford, 1992.

[PSZ00] R. Paturi, M. E. Saks, and F. Zane, *Exponential lower bounds for depth three Boolean circuits*, Comput. Complexity, 9 (2000), pp. 1–15.

[Rag08] P. Raghavendra, *Optimal algorithms and inapproximability results for every CSP?*, in STOC, ACM, New York, 2008, pp. 245–254.

[Sch80] J. T. Schwartz, *Fast probabilistic algorithms for verification of polynomial identities*, J. ACM, 27 (1980), pp. 701–717.

[Sha90] A. Shamir, *IP=PSPACE*, in FOCS, IEEE Computer Society, Los Alamitos, CA, 1990, pp. 11–15.

[SS09] N. Saxena and C. Seshadhri, *An almost optimal rank bound for depth-*3 *identities*, in IEEE Conference on Computational Complexity, IEEE Computer Society, Los Alamitos, CA, 2009, pp. 137–148.

[SS10] N. Saxena and C. Seshadhri, *From Sylvester-Gallai configurations to rank bounds: Improved black-box identity test for depth-*3 *circuits*, in FOCS, IEEE Computer Society, Los Alamitos, CA, 2010, pp. 21–29.

[SS11] N. Saxena and C. Seshadhri, *Blackbox identity testing for bounded top fanin depth-*3 *circuits: The field doesn't matter*, in STOC, ACM, New York, 2011, pp. 431–440.

[SV85] S. Skyum and L. G. Valiant, *A complexity theory based on boolean algebra*, J. ACM, 32 (1985), pp. 484–502.

[SV08] A. Shpilka and I. Volkovich, *Read-once polynomial identity testing*, in STOC, ACM, New York, 2008, pp. 507–516.

[SV09] A. Shpilka and I. Volkovich, *Improved polynomial identity testing for read-once formulas*, in APPROX-RANDOM, Springer, Berlin, 2009, pp. 700–713.

[SV10] A. Shpilka and I. Volkovich, *On the relation between polynomial identity testing and finding variable disjoint factors*, in ICALP, Springer, Berlin, 2010, pp. 408–419.

[SV11] S. Saraf and I. Volkovich, *Black-box identity testing of depth-*4 *multilinear circuits*, in STOC, ACM, New York, 2011, pp. 421–430.

[SW02] A. Shpilka and A. Wigderson, *Depth-*3 *arithmetic circuits over fields of characteristic zero*, Comput. Complexity, 10 (2002), pp. 1–27.

[SY10] A. Shpilka and A. Yehudayoff, *Arithmetic circuits: A survey of recent results and open questions*, Found. Trends Theor. Comput. Sci., 5 (2010), pp. 207–388.

[Uma03] C. Umans, *Pseudo-random generators for all hardnesses*, J. Comput. System Sci., 67 (2003), pp. 419–440.

[Val79] L. G. Valiant, *Completeness classes in algebra*, in STOC, ACM, New York, 1979, pp. 249–261.

[VSBR83] L. G. Valiant, S. Skyum, S. J. Berkowitz, and C. Rackoff, *Fast parallel computation of polynomials using few processors*, SIAM J. Comput., 12 (1983), pp. 641–644.

[Wil11] R. Williams, *Non-uniform ACC circuit lower bounds*, in IEEE Conference on Computational Complexity, IEEE, Piscataway, NJ, 2011, pp. 115–125.

[Zip79] R. Zippel, *Probabilistic algorithms for sparse polynomials*, EUROSAM, Springer, Berlin, 1979, pp. 216–226.