Narrow Proofs May Be Maximally Long*

Albert Atserias
Universitat Politècnica de Catalunya
atserias@cs.upc.edu

Massimo Lauria
KTH Royal Institute of Technology
lauria@kth.se

Jakob Nordström
KTH Royal Institute of Technology
jakobn@kth.se

September 9, 2014

Abstract

We prove that there are 3-CNF formulas over n variables that can be refuted in resolution in width w but require resolution proofs of size $n^{\Omega(w)}$. This shows that the simple counting argument that any formula refutable in width w must have a proof in size $n^{\mathrm{O}(w)}$ is essentially tight. Moreover, our lower bound generalizes to polynomial calculus resolution (PCR) and Sherali-Adams, implying that the corresponding size upper bounds in terms of degree and rank are tight as well. Our results do not extend all the way to Lasserre, however, where the formulas we study have proofs of constant rank and size polynomial in both n and w.

1 Introduction

Proof complexity studies how hard it is to prove that propositional logic formulas are tautologies. While the original motivation for this line of research, as discussed in [CR79], was to prove superpolynomial lower bounds on proof size for increasingly stronger proof systems as a way towards establishing $\mathsf{NP} \neq \mathsf{co\text{-}NP}$ (and hence $\mathsf{P} \neq \mathsf{NP}$), it is probably fair to say that most current research in proof complexity is driven by other concerns.

One such concern is the connection to SAT solving. By a standard transformation any propositional logic formula can be converted to another formula in conjunctive normal form (CNF) that has the same size up to constant factors and is unsatisfiable if and only if the original formula is a tautology. Any algorithm for solving SAT defines a proof system in the sense that the execution trace of the algorithm constitutes a polynomial-time verifiable witness of unsatisfiability.[1] In fact, most modern-day SAT solvers can be seen to search for proofs in systems at fairly low levels in the proof complexity hierarchy, and upper and lower bounds for these proof systems hence give information about the potential and limitations of the corresponding SAT solvers. In this work, we focus on such proof systems.

1.1 Background

The dominant strategy in applied SAT solving today is so-called *conflict-driven clause learning (CDCL)* [BS97, MS99, MMZ+01], which is ultimately based on the *resolution* proof system [Bla37]. The most studied complexity measure for resolution is *size* (also referred to as

*This is the full-length version of the paper [ALN14], which appeared in *Proceedings of the 29th Annual IEEE Conference on Computational Complexity (CCC '14)*.

[1] Such a witness is often referred to as a *refutation* rather than a *proof*, and these two terms are sometimes used interchangeably.

length), which gives lower bounds on the running time on CDCL solvers and for which (optimal) exponential lower bounds are known [Hak85, Urq87, CS88]. Another more recently studied measure is *space*, which corresponds to memory usage, and for which (again optimal) linear lower bounds have been proven [ABRW02, BG03, ET01]. For all of these results, the concept of *width*, measured as the size of a largest clause in a resolution proof, has turned out to play a key role. Width was identified as a crucial resource already in [Gal77], and strong lower bounds on proof width have been shown to imply lower bounds on proof size [BW01] and space [AD08].

Interestingly, although the relationships and trade-offs between width and space in resolution are by now fairly well-understood [Ben09, BN08], as are those between size and space [BN08, BN11, BBI12, BNT13], very basic questions about the connections between size and width have remained open. For instance, the argument in [BW01] that width gives a lower bound on size works by transforming a short resolution proof into a narrow one, but this transformation causes an exponential increase in the size. It is not known whether such a blow-up is necessary, i.e., if there are trade-offs between size and width, or whether the analysis in [BW01] can be sharpened to show that short proofs can be made simultaneously narrow. Also, as noted in the same paper, an upper bound w on the refutation width for a formula over n variables implies a proof size of at most $n^{O(w)}$ simply by counting the number of possible distinct clauses of width w. Again, it is not clear how tight this argument is—for all standard formula families in the literature known to be refutable in small enough width w there are refutations in size $n^{O(1)}$ independent of the width complexity (in fact, even in size *linear* in the formula size). To the best of our knowledge, it has been open whether there exist formulas refutable in width $w = O(\sqrt{n})$ that require size $n^{\Omega(w)}$, i.e., with the width complexity appearing in the exponent.

From a theoretical point of view, the ubiquity of CDCL in SAT solving is somewhat puzzling since resolution is a quite weak proof system. A different approach is to translate CNF formulas to multilinear polynomials and do Gröbner basis computations, which corresponds to *polynomial calculus resolution (PCR)* as defined in [CEI96, ABRW02].[2] Intriguingly, although PCR is known to be exponentially stronger than resolution, implementations of search methods for this proof system such as PolyBoRi [BD09, BDG+09] have a hard time competing with CDCL solvers.

Proof size and space in PCR is defined in analogy with resolution, and the measure corresponding to width of clauses is (total) *degree* of polynomials. It is straightforward to show that PCR can simulate resolution efficiently with respect to all of these measures, meaning that the same worst case upper bounds as in resolution apply to PCR. It was proven in [IPS99] that strong degree lower bounds imply strong size lower bounds, which is a close parallel to the size-width relation for resolution in [BW01], and this size-degree relation has been employed to prove exponential lower bounds on size in a number of papers, with [AR03] perhaps providing the most general setting. Optimal (linear) lower bounds on space were obtained in [BG13] building on [ABRW02, FLN+12], but it is worth noting that these bounds are *not* derived from degree lower bounds—it remains unknown whether an analogue of [AD08] holds for PCR (although [FLM+13] recently reported some progress on this and related open questions). Strong trade-offs between size and space as well as between degree and space have been shown in [BNT13], but—again in analogy with resolution—the exact relations between size and degree remains unclear. The same blow-up as in [BW01] occurs in [IPS99] when small size is converted to small degree, but it is not known whether this is necessary or just an artifact of the proof. Also, it was shown in [CEI96] that a degree upper bound of d implies proof size at most $n^{O(d)}$, but it has been open whether this is tight or not.

Yet another way to achieve greater expressivity than in resolution is to translate clauses into linear inequalities and manipulate them using 0-1 linear programming. Perhaps the simplest

[2] The resolution 'R' in PCR stands for the fact that negated literals get their own formal variables when translating CNF formulas to polynomials. Such variables were missing in the original definition in [CEI96] but adding them makes for a more natural and well-behaved proof system.

and most well-known example of this approach is the *cutting planes* proof system introduced in [CCT87] based on ideas in [Chv73, Gom63]. In this paper, however, we will be interested in somewhat related but different *semialgebraic* methods operating on linear programming relaxations of the CNF translations, such as the *Sherali-Adams*, *Lovász-Schrijver*, and *Lasserre* hierarchies used for attacking NP-hard optimization problems. We discuss this next.

The *Sherali-Adams (SA)* method [SA90] provides a hierarchy of linear programming relaxations of any given 0-1 integer program. The nth level of the hierarchy, where n is the number of 0-1 integer variables, wipes out the integrality gap and is thus exact, but also leads to an exponential blow-up in problem size. The main point of the method, however, is that any linear function of the variables can be optimized over the kth level of the hierarchy in time $n^{\mathrm{O}(k)}$, and in particular feasibility of the kth level relaxation can be checked in that time. In the context of proof complexity, what this means is that if the kth level relaxation of the integer programming formulation of a CNF formula in infeasible (the minimal such k is known as the *SA rank* of the integer program), then there is an $n^{\mathrm{O}(k)}$-time algorithm that can detect this. Furthermore, since the kth level of the hierarchy is an explicitly defined linear program, its infeasibility can be certified as a positive linear combination of its defining inequalities. Such a certificate is a rank-k Sherali-Adams refutation of the corresponding CNF formula.

The *Lovász-Schrijver* approach [LS91] can be thought of as (and indeed it is formally equivalent to) an iterated version of the level-2 SA relaxation. The point is again that any linear function can be optimized over the linear program after k iterations in time $n^{\mathrm{O}(k)}$. Lovász and Schrijver also introduced a method LS^+, which uses semidefinite programming instead of linear programming, and which is significantly stronger in some notable cases of interest in combinatorial optimization.

The *Lasserre* method [Las01], finally, is basically the Sherali-Adams method with semidefinite programming conditions at all levels of the hierarchy. Again it stratifies into levels and the kth level can be solved in time $n^{\mathrm{O}(k)}$. Moreover, Lasserre's method is the strongest of all three in the sense that, level by level, it provides the tightest of all three approximations of the integer linear program. We refer to [Lau01, CT12] for a more detailed discussion of Sherali-Adams, Lovász-Schrijver and Lasserre and a comparison of their relative strength.

In view of the important algorithmic applications that these methods have (see, e.g., [Par00] and subsequent work), it is a natural question whether the upper bounds $n^{\mathrm{O}(k)}$ for rank k are tight, just as for resolution and polynomial calculus resolution.

From the proof complexity side, some notable early papers investigating semialgebraic proof systems were published around the turn of the millennium [Pud99, GV01, GHP02], but then this area of research seems to have gone dormant. In the last few years, these proof systems have made an exciting reemergence in the context of hardness of approximation, revealing unexpected and intriguing connections between approximation and proof complexity. Some examples of this is the paper [Sch08] essentially rediscovering results from [Gri01], and more recent papers such as [BBH+12, OZ13]. There have also been papers such as [BPS07] and (the very recent) [GP14] focusing on *semantic* versions of these proof systems, with less attention to the actual syntactic derivation rules used.

1.2 Our results

The main contribution of this paper is showing that the upper bounds on proof size in terms of width for resolution, degree for PCR, and rank for Sherali-Adams are essentially tight (up to constant factors in the exponent). Moreover, an interesting feature of our result is that we can actually use the same formula family to prove tightness simultaneously for all the proof systems. What this means is that we obtain upper bounds on size in resolution that tightly match lower bounds in the much stronger systems PCR and Sherali-Adams (which are in turn tight for these systems since resolution width is an upper bound on both PCR degree and Sherali-Adams rank). The formal statement of this result is as follows.

Theorem 1.1. *Let $w = w(n)$ be such that $w = O(n^c)$ for some positive constant $c < 1/2$. Then there are 3-CNF formulas $F_{n,w}$ with $\mathrm{O}(wn)$ clauses over $\mathrm{O}(n)$ variables such that the following holds:*

1. *$F_{n,w}$ has a resolution refutation in simultaneous size $n^{\mathrm{O}(w)}$, width $\mathrm{O}(w)$ and space $\mathrm{O}(w)$.*
2. *Any refutation of $F_{n,w}$ in resolution, PCR, or Sherali-Adams must have size $n^{\Omega(w)}$.*

For resolution this actually shows something slightly stronger than that the counting upper bound on size in terms of width is tight. Namely, since the formulas in Theorem 1.1 have the same asymptotic upper bound on space as on width, it follows that even for formulas of space complexity $\mathrm{O}(w)$—which is a more stringent requirement than width complexity $\mathrm{O}(w)$—it is still impossible to obtain any size upper bound better than $n^{\mathrm{O}(w)}$ in general.

Theorem 1.1 has an interesting consequence for the analysis of CDCL solver performance, which we state as a formal corollary. By way of background, it was shown in [AFT11] that if a CNF formula F over n variables has a resolution refutation in width w, then with high probability any CDCL solver[3] will only need time $n^{\mathrm{O}(w)}$ to decide that F is indeed unsatisfiable.[4] An obvious question is whether this result is tight. Theorem 1.1 shows that the answer is "yes," since no CDCL solver can run faster than the shortest resolution proof it can possibly find.[5]

Corollary 1.2. *There are formulas F over n variables refutable in resolution in width w for which any resolution-based CDCL solver cannot run faster than $n^{\Omega(w)}$, and hence the result in [AFT11] is optimal up to constants in the exponent.*

Another interesting aspect of our lower bound for resolution is in the context of Berkholz's EXPTIME-completeness result for deciding resolution width [Ber12]. What Berkholz showed is that given a formula F over n variables and a parameter w, it cannot be decided in time less than $n^{(w-3)/12}$ whether F has a resolution refutation in width w or not. Optimizing the constants in Theorem 1.1, we can show that there are 4-CNF formulas refutable in width w for which no resolution refutation can be shorter than $n^{w/2-\mathrm{o}(1)}$. It is worth noting that this bound is stronger than that in [Ber12], although it of course applies only for the more restricted setting where the algorithm has to output a width-w resolution refutation rather than for the general decision problem. Still, we believe this sheds interesting light on Berkholz's result.

1.3 Discussion of proof techniques

We conclude the overview by outlining the proof of the lower bound in Theorem 1.1 for resolution and how it differs from previously used methods. At a high level, our proof is a standard restriction argument, but it turns out to have some twists which we believe might be of interest and could be useful elsewhere.[6]

Before going into the details of our new restriction argument, let us revisit previous lower bounds on size in terms of width and see how they fall short of proving what we are after. On the one hand, the result in [BW01] states that if a 3-CNF formula on n variables requires width w to refute in resolution, then it also requires size $2^{\Omega(w^2/n)}$. This lower bound is vacuous for w smaller than $\sqrt{n}$ and, in any case, can never be larger than $2^{\Omega(w)}$ since w is bounded by n. On

[3] This result holds for a fairly general mathematical model of what a CDCL solver is, which agrees reasonably well with how state-of-the-art solvers are actually implemented in practice.

[4] Perhaps this might not seem so impressive at first sight—after all, exhaustive search in bounded width runs within this time bound deterministically—but the point is that a CDCL solver is very far from doing exhaustive width search and does not care at all about the existence or non-existence of narrow refutations.

[5] This is of course assuming that the solver does not implement features such as, e.g., cardinality reasoning or extended resolution, since these fall outside of the standard CDCL framework and go beyond resolution-based reasoning.

[6] In fact, in a sense this has already happened in that our paper heavily draws on ideas from [AMO13], which used a similar approach in a very different context.

the other hand, for formulas refutable in width w smaller than $\sqrt{n}$, a direct random restriction argument can sometimes still be applied to get meaningful lower bounds. The idea is that setting a random literal to true will kill off a $\frac{w}{2n}$-fraction of the wide clauses on average. After r rounds of such restrictions, the expected number of surviving wide clauses is at most $\left(1-\frac{w}{2n}\right)^r S$, where S is the size of the refutation, and choosing $r = (2n/w)\log S$ brings the number of wide clauses down to zero. A contradiction is then derived by showing that the residual formula still requires width w to refute. Note, however, that we cannot apply the restriction for more than n rounds (or else there will be no residual formula to argue about), and so the best size lower bound this method can achieve is again $2^{\Omega(w)}$, which is smaller than the $n^{\Omega(w)}$ bound that we are after.

In some sense, the problem is that using restrictions in the style of Håstad's switching lemma [Hås87] does not work in our setting. Instead, it turns out that a seemingly weaker argument inspired by Furst-Saxe-Sipser [FSS84] is just what we need. Let us now describe this modified restriction argument and how it overcomes the problems discussed above.

We start with a carefully chosen family of formulas $F_{n,w}$ and an associated distribution over random restrictions ρ_n. Then we assume that we have a resolution refutation π of $F_{n,w}$ in size $n^{o(w)}$ and analyze how a randomly chosen restriction ρ_n affects π. We get two cases:

1. For clauses C in the refutation π that are noticeably wide, ρ_n is very likely to satisfy a literal in C and so the clause disappears.
2. Clauses that are not so wide will not be satisfied by ρ_n, but since they are reasonably small they are very likely to be shortened by ρ to width strictly less than w.

Admittedly, the first case looks no different from the standard restriction argument, and the second case seems quite weak. But the point is that by considering also the second case, we can afford a significantly bigger bound for "wide" than before, thus getting a bigger probability of success. This is the key to our argument. The rest is now standard: $F_{n,w}$ and ρ_n are chosen so that $F_{n,w}$ restricted by ρ_n is a bounded-width version of a pigeonhole principle (PHP) formula with w pigeons that are supposed to fit into $w-1$ holes. Since π is short enough, by a counting argument there is some restriction ρ_n that eliminates all wide clauses to give a resolution refutation of the PHP formula in width less than w. It is a straightforward separate argument that such a narrow refutation cannot exist, and the lower bound on size follows.

The lower bounds for PCR and Sherali-Adams are quite similar. The restriction part of the argument is basically the same, but one has to work a bit harder to prove the final punchline that the restricted refutations have impossibly low degree and rank, respectively.

It should perhaps be stressed that while the final argument is quite straightforward and natural (at least for resolution), a crucial component in the proof is to find the right formulas $F_{n,w}$ and associated restrictions ρ_n to plug into the argument, and to make a case analysis of the action of ρ_n as above. Both of these aspects use the techniques developed in [AMO13] in an essential way.

1.4 Outline of this paper

The rest of this paper is organized as follows. After having given the necessary preliminaries in Section 2, we state the main theorem for resolution and give a full proof in Section 3. We believe this can serve as a useful warm-up to the more complicated proofs for stronger proof systems that follow in Section 4. In Section 5 we show that our lower bounds do *not* extend all the way to Lasserre. We conclude in Section 6 with some final remarks and a discussion of open problems.

2 Preliminaries

A *literal* over a Boolean variable x is either the variable x itself (a *positive literal*) or its negation $\overline{x}$ (a *negative literal*). A *clause* $C = a_1 \vee \cdots \vee a_k$ is a disjunction of literals. A *k-clause* is a clause that contains at most k literals. A *CNF formula* $F = C_1 \wedge \cdots \wedge C_m$ is a conjunction of clauses. A *k-CNF formula* is a CNF formula consisting of k-clauses. We think of clauses and CNF formulas as sets: the order of elements is irrelevant and there are no repetitions. We denote the logical true value as $\top$ and the logical false value as $\bot$. The empty clause (containing no literals) is also denoted $\bot$, since it is always false. For integers m and n, $m < n$, we use the standard notation $[n] = \{1, 2, \ldots, n\}$ and $[m, n] = \{m, m+1, \ldots, n\}$.

A *resolution derivation* of a clause C from a CNF formula F is a sequence of clauses $(C_1, \ldots, C_\tau)$ such that $C_\tau = C$ and for $1 \leq t \leq \tau$ the clause C_t is obtained by one of the following derivation rules:

- **Axiom:** C_t is a clause in F (an *axiom clause*);
- **Inference:** $C_t = A \vee B$, where $C_i = A \vee x$ and $C_j = B \vee \overline{x}$ for $1 \leq i, j < t$;
- **Weakening:** $C_t \supseteq C_i$ for some $1 \leq i < t$.

A *resolution refutation* of F is a derivation of the empty clause $\bot$ from F.

Every resolution derivation $\pi = (C_1, \ldots, C_\tau)$ can be associated with a directed acyclic graph (DAG) G_π with vertices labelled by clauses C_t in π and edges (C_i, C_j) if C_j is obtained by an inference or a weakening step and C_i is used as a premise in that step. The derivation π is said to be *tree-like* if G_π is a tree. The *(clause) space* of π at time t is the number of clauses derived before or at time t that will be used after or at time t, i.e., all clauses C_i, $i \leq t$, in G_π having an outgoing edge to clauses C_j, $j \geq t$ (plus the clause C_t itself). The space of π is the maximum space at any time t in the derivation. The *width* of π is the maximum number of literals in any clause C_t in π, and the *size* (or *length*) of $\pi = (C_1, \ldots, C_\tau)$ is τ. We remark that it is straightforward to show that all applications of the weakening rule can be eliminated from a resolution refutation without any increase in size, width, or space, and while maintaining tree-likeness.

In *polynomial calculus resolution (PCR)* one instead refutes an unsatisfiable formula F over variables $x_1, \ldots, x_n$ by reasoning in terms of polynomials in the ring $\mathbb{F}[x_1, \ldots, x_n, \overline{x}_1, \ldots, \overline{x}_n]$, where $\mathbb{F}$ is some fixed field and x_i, $\overline{x}_i$ are formally independent variables. It is natural to think of polynomials as being satisfied by an assignment when they evaluate to 0, so in PCR the truth values $\top$ and $\bot$ are represented by 0 and 1, respectively, and a clause $\bigvee_{i \in \mathcal{I}} x_i \vee \bigvee_{j \in \mathcal{J}} \overline{x}_j$ is translated into the one-term polynomial $\prod_{i \in \mathcal{I}} x_i \cdot \prod_{j \in \mathcal{J}} \overline{x}_j$. A *PCR derivation* of a polynomial R from a set of polynomials $\mathcal{S} = \{Q_1, \ldots, Q_m\}$ is a sequence $(P_1, \ldots, P_\tau)$ such that $P_\tau = R$ and for $1 \leq t \leq \tau$ the polynomial P_t is obtained by one of the following derivation rules:

- **Boolean axiom:** P_t is $x^2 - x$ for some variable x (or $\overline{x}$);
- **Complementarity axiom:** P_t is $1 - x - \overline{x}$ for some variable x;
- **Initial axiom:** P_t is one of the polynomials $Q_j \in \mathcal{S}$;
- **Linear combination:** $P_t = \alpha P_i + \beta P_j$ for $1 \leq i, j < t$ and some $\alpha, \beta \in \mathbb{F}$;
- **Multiplication:** $P_t = x P_i$ for $1 \leq i < t$ and some variable x.

A *PCR refutation* of F is a PCR derivation of 1 from the set of polynomials representing the clauses of F as explained above. Note that the Boolean axioms make sure that variables can only take values $\top = 0$ and $\bot = 1$, and the complementarity axioms enforce that x and $\overline{x}$ take opposite values.

The *degree* of a PCR derivation π is the maximum of the (total) degrees of the polynomials in π. The *size* of π is the sum of the sizes of the polynomials in π, where the size of a polynomial is defined as its number of terms.[7] The *space* measure can also be generalized from resolution, counting terms instead of clauses, but we will not really need it in this paper.

[7]Just to make terminology precise, in this paper a *monomial* is a product of variables, a *term* is a monomial multiplied by a non-zero coefficient from the field $\mathbb{F}$, and a *polynomial* is a sum of terms with distinct monomials.

Let us next discuss *semialgebraic* proof systems. All such proof systems encode a CNF formula as a set of polynomial inequalities over the reals. A clause $\bigvee_{i\in\mathcal{I}} x_i \vee \bigvee_{j\in\mathcal{J}} \overline{x}_j$ is represented by the inequality $\sum_{i\in\mathcal{I}} x_i + \sum_{j\in\mathcal{J}}(1-x_j) - 1 \geq 0$, where we identify $\top = 1$ and $\bot = 0$—note that this is the opposite of the convention for PCR. A CNF formula F is represented by the inequalities corresponding to its clauses. A *Sherali-Adams (SA) derivation* of an inequality $R \geq 0$ from a set of polynomial inequalities $\{Q_1 \geq 0, \ldots, Q_m \geq 0\}$ is a formula of the form

$$\sum_{t=1}^{\tau} \alpha_t \cdot \prod_{i\in\mathcal{I}_t} x_i \cdot \prod_{i\in\mathcal{J}_t}(1-x_i)\cdot P_t \ , \tag{2.1}$$

that when expanded into a sum of terms gives the polynomial R, where $\alpha_t \in \mathbb{R}^+$ and P_t is one of the original polynomials Q_j, or an *axiom* of the form $x_i^2 - x_i$ or $x_i - x_i^2$, or the constant 1. A *Lasserre derivation* of $R \geq 0$ is a formula of the form (2.1) that expands to R where in addition P_t can be a square Q^2 for any arbitrary polynomial Q. Note that Sherali-Adams and Lasserre are *static* proof systems in that they have "one-shot" derivations, in contrast to resolution and PCR that construct derivations dynamically step by step.

We can augment Sherali-Adams by *twin variables* $\overline{x}_i$ whose intended meaning is the negation of x_i, i.e., $1-x_i$.[8] We define a *Sherali-Adams resolution (SAR) derivation* to be an SA derivation as in (2.1) except that the set of variables is $\{x_1,\ldots,x_n,\overline{x}_1,\ldots,\overline{x}_n\}$ and that P_t can also be a complementarity axiom $1 - x_i - \overline{x}_i$ or $-1 + x_i + \overline{x}_i$.

A *Sherali-Adams (SA), SAR, or Lasserre refutation* of F is a derivation in the respective system of the inequality $-1 \geq 0$ from the inequalities $Q_1 \geq 0, \ldots, Q_m \geq 0$ that encode the clauses of F. The *rank* of the derivation is the maximum of the degrees among the polynomials to which the formulas $\prod_{i\in\mathcal{I}_t} x_i \cdot \prod_{i\in\mathcal{J}_t}(1-x_i)\cdot P_t$ in (2.1) expand, and the *size* of the derivation is the sum of the sizes of those polynomials, where again the size of a polynomial is defined as its number of terms.

A *restriction* (or *partial assignment*) ρ is a partial mapping from variables to $\{\bot,\top\}$. We identify ρ with the set of literals it sets to true. The *domain* of ρ is denoted $\mathrm{dom}(\rho)$ and the size of ρ is $|\rho| = |\mathrm{dom}(\rho)|$. The restriction $C{\restriction_\rho}$ of a clause C by ρ is the trivial clause $\top$ if ρ sets some literal of C to true—such a clause can just be removed from any formula or derivation—and otherwise it is the clause resulting from deleting all literals in C set to false by ρ. The restriction $F{\restriction_\rho}$ of a CNF formula F is the conjunction of its restricted clauses, and a restricted resolution derivation $\pi{\restriction_\rho}$ is the sequence of the restrictions of the clauses in π. It is a basic fact that if π is a refutation of F, then $\pi{\restriction_\rho}$ is a refutation of $F{\restriction_\rho}$.

For PCR derivations and the polynomials therein, restrictions are defined similarly: a restricted term vanishes if one of its variables is set to $\top = 0$ and is otherwise obtained by deleting all variables set to $\bot = 1$, and a restricted polynomial is the sum of its restricted terms. Again, restrictions preserve PCR refutations. For SA and SAR, the definition is analogous except the roles of 0 and 1 are reversed.

3 Upper and lower bounds in resolution

In this section, we establish the special case of our main result for the resolution proof system. Although the lower bound part follows from the stronger results that we will prove in later sections, we believe it is instructive to develop the argument for resolution first. Let us start by stating a slightly more detailed version of Theorem 1.1, but restricted to resolution, which is what we will prove.

[8] As briefly discussed above, this is how PCR was extended in [ABRW02] from the original definition of polynomial calculus (PC) in [CEI96].

Theorem 3.1. *Let $k = k(n)$ be any integer-valued function such that $k(n) \leq n/4\log n$. Then there is a family of 3-CNF formulas $\{F_{n,k}\}_{n\geq 1}$, where $F_{n,k}$ has $\mathrm{O}(n^2)$ variables and $\mathrm{O}(kn^2)$ clauses, such that:*

1. *$F_{n,k}$ has a tree-like resolution refutation in size $\mathrm{O}(k^k n^k)$, width $2k+1$, and space $2k+3$;*
2. *any resolution refutation of $F_{n,k}$ has size $\Omega\big(n^{k-1}/(4k\log n)^k\big)$.*

Straightforward calculations show that if $k(n) = \mathrm{O}(n^c)$ for $c < 1$, then the upper bound is $n^{\mathrm{O}(k)}$ and the lower bound is $n^{\Omega(k)}$.

3.1 Definition of the formula

The CNF formulas we use to establish Theorem 3.1 formalize a *relativized* version of the pigeonhole principle which says that there is a way to choose k out of n pigeons and send them to $k-1$ pigeonholes so that every pigeon gets its own hole. More formally, the formula claims that there are (partial) functions $p : [k] \to [n]$ and $q : [n] \to [k-1]$ such that p is one-to-one and defined on $[k]$, and q is one-to-one and defined on the range of p. Let us first describe a straightforward CNF encoding of this claim with wide clauses that we denote $\mathit{RPHP}^{k,n}_{k-1}$. Once the general idea is clear, we transform this into a slightly more involved 3-CNF formula which is the formula we will work with.

The formula $\mathit{RPHP}^{k,n}_{k-1}$ is over variables $p_{u,v}$ that encode the function p, $q_{v,w}$ that encode the function q, and r_v that encode a superset of the range of p. It consists of the following collection of clauses:

$$
\begin{aligned}
&p_{u,1} \lor p_{u,2} \lor \cdots \lor p_{u,n} && u \in [k], && \text{(3.1a)}\\
&\overline{p}_{u,v} \lor \overline{p}_{u',v} && u,u' \in [k],\ u \neq u',\ v \in [n], && \text{(3.1b)}\\
&\overline{p}_{u,v} \lor r_v && u \in [k],\ v \in [n], && \text{(3.1c)}\\
&\overline{r}_v \lor q_{v,1} \lor \cdots \lor q_{v,k-1} && v \in [n], && \text{(3.1d)}\\
&\overline{r}_v \lor \overline{r}_{v'} \lor \overline{q}_{v,w} \lor \overline{q}_{v',w} && v,v' \in [n],\ v \neq v',\ w \in [k-1]. && \text{(3.1e)}
\end{aligned}
$$

The clauses in (3.1a)–(3.1b) say that p maps $[k]$ injectively into $[n]$; clauses (3.1c) encode the range of p; and clauses (3.1d)–(3.1e) force q to be defined and injective on this range.

Next, we convert $\mathit{RPHP}^{k,n}_{k-1}$ to a 3-CNF formula. This is done in the standard way by using extension variables to break up the wide clauses in (3.1a) and (3.1d) and the 4-clauses in (3.1e). For (3.1a) we obtain the clauses

$$
\begin{aligned}
&p_{u,1} \lor p_{u,2} \lor y_{u,2} && u \in [k], && \text{(3.2a)}\\
&\overline{y}_{u,v} \lor p_{u,v+1} \lor y_{u,v+1} && u \in [k],\ v \in [2,n-3], && \text{(3.2b)}\\
&\overline{y}_{u,n-2} \lor p_{u,n-1} \lor p_{u,n} && u \in [k], && \text{(3.2c)}
\end{aligned}
$$

splitting up (3.1d) yields

$$
\begin{aligned}
&\overline{r}_v \lor q_{v,1} \lor z_{v,1} && v \in [n], && \text{(3.2d)}\\
&\overline{z}_{v,w} \lor q_{v,w+1} \lor z_{v,w+1} && v \in [n],\ w \in [k-4], && \text{(3.2e)}\\
&\overline{z}_{v,k-3} \lor q_{v,k-2} \lor q_{v,k-1} && v \in [n], && \text{(3.2f)}
\end{aligned}
$$

and the rest of the clauses are

$$
\begin{aligned}
&\overline{p}_{u,v} \lor \overline{p}_{u',v} && u,u' \in [k],\ u \neq u',\ v \in [n], && \text{(3.2g)}\\
&\overline{p}_{u,v} \lor r_v && u \in [k],\ v \in [n], && \text{(3.2h)}\\
&\overline{r}_v \lor \overline{r}_{v'} \lor r_{v,v'} && v,v' \in [n],\ v \neq v', && \text{(3.2i)}\\
&\overline{r}_{v,v'} \lor \overline{q}_{v,w} \lor \overline{q}_{v',w} && v,v' \in [n],\ v \neq v',\ w \in [k-1]. && \text{(3.2j)}
\end{aligned}
$$

The 3-CNF formula consisting of the clauses in (3.2a)–(3.2j), which we will denote $ERPHP^{k,n}_{k-1}$, is the formula for which we will prove Theorem 3.1. It is easy to verify that this formula has $\mathrm{O}(kn^2)$ clauses over $\mathrm{O}(n^2)$ variables. We note that if we did not insist on bringing the clause size all the way down to 3, then we could get a 4-CNF formula with $\mathrm{O}(kn^2)$ clauses over $\mathrm{O}(kn)$ variables by not converting the 4-clauses in (3.1e) into the 3-clauses (3.2i) and (3.2j). Our proof of Theorem 3.1 works for this formula as well after straightforward adjustments and gives a slightly better lower bound expressed in terms of the number of variables.

3.2 Proof of the upper bound

Let us first describe how we can refute the formula $ERPHP^{k,n}_{k-1}$ in resolution. In order to do so, we consider all sequences of the form $(v_1, v_2, \ldots, v_k, w_1, w_2, \ldots, w_k)$, where $v_u \in [n]$ and $w_u \in [k-1]$, and the corresponding clauses

$$\bigvee_{u \in [k]} \overline{p}_{u,v_u} \vee \bigvee_{u \in [k]} \overline{q}_{v_u,w_u}. \tag{3.3}$$

We derive all such clauses from the axiom clauses of $ERPHP^{k,n}_{k-1}$, and from these clauses it is then straightforward to obtain a contradiction. All of these derivations are efficient, so the size of the whole refutation is dominated by the number of clauses in (3.3).

For each clause in (3.3) we are in one of two cases: either $v_u = v_{u'}$ holds for some $u \neq u'$, or there must exist a pair $v_u \neq v_{u'}$ with $w_u = w_{u'}$ by the pigeonhole principle. In the former case, the clause (3.3) is just a weakening of the axiom (3.2g), namely $\overline{p}_{u,v} \vee \overline{p}_{u',v}$ with $v = v_u = v_{u'}$. In the latter case, we combine axioms $\overline{p}_{u,v_u} \vee r_{v_u}$ and $\overline{p}_{u',v_{u'}} \vee r_{v_{u'}}$ from (3.2h), $\overline{r}_{v_u} \vee \overline{r}_{v_{u'}} \vee r_{v_u,v_{u'}}$ from (3.2i), and $\overline{r}_{v_u,v_{u'}} \vee \overline{q}_{v_u,w} \vee \overline{q}_{v_{u'},w}$ from (3.2j), where $w = w_u = w_{u'}$, to obtain the clause $\overline{p}_{u,v_u} \vee \overline{p}_{u',v_{u'}} \vee \overline{q}_{v_u,w} \vee \overline{q}_{v_{u'},w}$. It is easy to see that (3.3) can be derived from this clause by weakening. Since a constant number of clauses is involved in this derivation it requires only constant space, and it is straightforward to verify that it can in fact be carried out by a tree-like derivation in space 3 (i.e., keeping one clause in memory and resolving it with a sequence of axioms).

The rest of the refutation consists of derivations of all prefixes of clauses of the form (3.3) by backward induction. For the inductive step we assume that we are able to derive any prefix clause of size t in clause space $(2k-t)+3$ and show how to derive any prefix of size $t-1$ in clause space $(2k-t+1)+3$. The refutation ends when we reach the prefix clause of size 0 (i.e., the empty clause) in clause space $2k+3$.

Suppose first that we can derive each clause of the form

$$\bigvee_{u \in [k]} \overline{p}_{u,v_u} \vee \bigvee_{u \in [k^*-1]} \overline{q}_{v_u,w_u} \vee \overline{q}_{v^*,w^*} = A \vee \overline{q}_{v^*,w^*} \tag{3.4}$$

for some $k^* < k$ in clause space s (writing $v^* = v_{k^*}$ and $w^* = w_{k^*}$ as a shorthand). We want to use the existence of such derivations to derive the clause A in space $s+1$. To this end, start with the axiom $\overline{p}_{k^*,v^*} \vee r_{v^*}$ and note that the literal $\overline{p}_{k^*,v^*}$ also appears in the left-hand part of A in (3.4). We resolve this clause with the axiom $\overline{r}_{v^*} \vee q_{v^*,1} \vee z_{v^*,1}$ to get $\overline{p}_{u,v^*} \vee q_{v^*,1} \vee z_{v^*,1}$. Keeping the latter clause in memory, we invoke a subderivation in space s of the clause $A \vee \overline{q}_{v^*,1}$ and resolve to obtain $A \vee z_{v^*,1}$. Continuing, assume that we have derived $A \vee z_{v^*,w}$ for some $w \geq 1$. Then we resolve this clause with the axiom $\overline{z}_{v^*,w} \vee q_{v^*,w+1} \vee z_{v^*,w+1}$ to obtain $A \vee q_{v^*,w+1} \vee z_{v^*,w+1}$. Keeping the latter clause in memory, we derive $A \vee \overline{q}_{v^*,w+1}$ using no more space than $s+1$ all in all, and then resolve to get $A \vee z_{v^*,w+1}$. When we reach the clause $A \vee z_{v^*,k-3}$ we resolve it with the axiom $\overline{z}_{v^*,k-3} \vee q_{v^*,k-2} \vee q_{v^*,k-1}$ and then with the inductively derived clauses $A \vee \overline{q}_{v^*,k-2}$ and $A \vee \overline{q}_{v^*,k-1}$ to obtain A. We point out again that the clause space of this derivation is $s+1$.

After k steps of this backward induction we get to clauses of the form $\overline{p}_{1,v_1} \vee \overline{p}_{2,v_2} \vee \ldots \vee \overline{p}_{k,v_k}$. To derive the empty clause we do k more steps of backward induction, mimicking the procedure in the previous paragraph. Suppose that we have shown how to derive all clauses

$$\bigvee_{u\in[k^*-1]} \overline{p}_{u,v_u} \vee \overline{p}_{k^*,v^*} = A \vee \overline{p}_{k^*,v^*} \tag{3.5}$$

for $k^* < k$ and want to derive A. To do so, first resolve the axiom $p_{k^*,1} \vee p_{k^*,2} \vee y_{k^*,2}$ with the inductively derived clause $A \vee \overline{p}_{k^*,1}$ and then with $A \vee \overline{p}_{k^*,2}$ to get $A \vee y_{k^*,2}$. Suppose that we have shown how to derive $A \vee y_{k^*,v}$ in this way for $v \geq 2$. In order to obtain $A \vee y_{k^*,v+1}$ we resolve $\overline{y}_{k^*,v} \vee p_{k^*,v+1} \vee y_{k^*,v+1}$ with $A \vee y_{k^*,v}$ and then with $A \vee \overline{p}_{k^*,v+1}$. We iterate up to $A \vee y_{k^*,n-2}$ and finally resolve the axiom $\overline{y}_{k^*,n-2} \vee p_{k^*,n-1} \vee p_{k^*,n}$ with the clauses $A \vee y_{k^*,n-2}$, $A \vee \overline{p}_{k^*,n-1}$, and $A \vee \overline{p}_{k^*,n}$ to obtain A. After k steps of this second stage we reach the empty clause and the refutation is complete. As before, the clause space goes up by an additive one for every inductive step, so the clause space of the whole refutation is $2k+3$.

To analyze the size of the resolution refutation obtained in this way, consider the prefix tree of the sequences $(v_1, v_2, \ldots, v_k, w_1, w_2, \ldots, w_k)$. Each vertex of this tree corresponds to one of the clauses A derived during the backward induction, with the empty clause at the root and clauses (3.3) at the leaves. The length of the derivation of each clause is linear in the number of children, and in addition we derived the leaves with a constant number of steps. Therefore we can charge a constant amount of steps per vertex. The size of the tree is $O(k^k n^k)$, and it follows that this is also the size of the refutation. Furthermore, the refutation is tree-like since no intermediate clause is used more than once. One can also observe that the width of the refutation is $2k+1$ and reaches this maximum at the induction step from sequences of length $2k$ to sequences of length $2k-1$.

3.3 Proof of the lower bound for resolution

As discussed in Section 1.3, we use a random restriction argument to prove our size lower bound for resolution refutations of the formula $\mathit{ERPHP}^{k,n}_{k-1}$. We define a distribution $\mathcal{D}$ on partial assignments ρ by picking a subset $\mathcal{S} = \{v_1, v_2, \ldots, v_k\}$ of k elements from $[n]$ uniformly at random and letting ρ assign values to variables as follows:

- $r_v = \top$ for all $v \in \mathcal{S}$; $r_v = \bot$ otherwise;
- $r_{v,v'} = r_v \wedge r_{v'}$ for all $v \neq v'$;
- $p_{u,v_u} = \top$ and $p_{u,v} = \bot$ for all $u \in [k]$ and all $v \neq v_u$;
- $y_{u,v}$ for all u and v are set arbitrarily so as to satisfy the clauses (3.2a)–(3.2c);
- $q_{v,w}$ and $z_{v,w}$ are left unset for all $v \in \mathcal{S}$ and all w;
- $q_{v,w} = b_v$ and $z_{v,w} = b_v$ for all $v \in [n] \setminus \mathcal{S}$ and all $w \in [k-1]$, where $b_v \in \{\bot, \top\}$ is chosen uniformly and independently at random for every $v \in [n] \setminus \mathcal{S}$.

We want to argue that with high probability such restrictions remove or at least significantly shrink wide clauses.

For $v \in [n]$, let us say that the variables $\{q_{v,1}, \ldots, q_{v,k-1}, z_{v,1}, \ldots, z_{v,k-1}\}$ *mention the pigeon v*. We say that a clause (or term) mentions v if it contains some variable in this set and define the *pigeon-width* to be the number of pigeons mentioned. The next lemma describes the effect of random restrictions ρ from $\mathcal{D}$ on clauses (or terms) depending on their pigeon-width. Namely, a sufficiently wide clause, i.e., mentioning a lot of pigeons, is satisfied by the random restriction with high probability, whereas a narrower clause may not have its truth value fixed by the restriction but will with high probability contain few pigeons afterwards.

Lemma 3.2. *Let k, ℓ, n be natural numbers such that $n \geq 16$ and $\ell \leq k \leq n/4\log n$. Let A be either a clause or term over the variables of $\mathit{ERPHP}^{k,n}_{k-1}$ and let ρ be a random restriction sampled from the distribution $\mathcal{D}$ as defined above. Then the pigeon-width of $A\restriction_\rho$ is less than ℓ with probability at least $1-(4k\log n)^k/n^\ell$.*

Proof. Let us assume that A is a clause—the proof for terms (which will be used for PCR and Sherali-Adams) is completely analogous. Let $v_1, \dots, v_r$ be the pigeons mentioned in A sorted in some order and let $a_1, \dots, a_r$ be a sequence of literals such that a_i witnesses that A mentions v_i.

If $r > 2k \log n$, then the probability that the clause A is not satisfied by the restriction is at most

$$\begin{aligned}\Pr\big[\rho(a_i) \neq \top \text{ for all } i = 1, \dots, r\big] &\leq \Pr\big[\rho(a_i) \neq \top \text{ for all } i = 1, \dots, \lceil 2k \log n \rceil\big] \\ &= \prod_{i=1}^{\lceil 2k \log n \rceil} \Pr\big[\rho(a_i) \neq \top \big| \rho(a_j) \neq \top \text{ for } j < i\big] \\ &\leq \prod_{i=1}^{\lceil 2k \log n \rceil} \Pr\big[\rho(a_i) \neq \top \big| v_j \notin \mathcal{S} \text{ for } j < i\big] \\ &\leq \prod_{i=1}^{\lceil 2k \log n \rceil} \left(\frac{1}{2} + \frac{k}{n-i}\right) \\ &< \left(\frac{5}{8}\right)^{2k \log n} < \frac{1}{n^k} \ .\end{aligned} \tag{3.6}$$

To see this, note that the event $\rho(a_i) \neq \top$ occurs either if the pigeon v_i is not picked or if the literal a_i is set to the wrong value. Assuming that no pigeon $v_1, \dots, v_{i-1}$ has been picked before v_i, the conditional probability of v_i being included in $\mathcal{S}$ is $k/(n-i)$, and is less otherwise. If $v_i \in \mathcal{S}$, then a_i gets the wrong value with probability $1/2$. The final inequalities hold because the ratio $k/(n - 2k \log n)$ is at most $1/(2 \log n)$, and therefore it is at most $1/8$ for $n \geq 16$.

If instead the number of pigeons mentioned by A is $r \leq 2k \log n$, we want to bound the probability that there are at least ℓ pigeons mentioned in A that are chosen in $\mathcal{S}$ and hence survive. The choices of $\mathcal{S}$ with exactly i pigeons mentioned in A are $\binom{r}{i}\binom{n-r}{k-i}$. Considering all possible intersections of size at least ℓ between the set $\mathcal{S}$ and the r pigeons mentioned in A, we obtain that the probability of ℓ surviving pigeons is at most

$$\begin{aligned}\sum_{i=\ell}^{k} \binom{r}{i}\binom{n-r}{k-i}\binom{n}{k}^{-1} &\leq k \binom{\lfloor 2k \log n \rfloor}{k}\binom{n}{k-\ell}\binom{n}{k}^{-1} \\ &\leq \frac{k(2k \log n)^k}{k!} \cdot \frac{n!}{(k-\ell)!(n-k+\ell)!} \cdot \frac{(n-k)!k!}{n!} \\ &\leq k(2k \log n)^k \cdot \frac{1}{(k-\ell)!} \cdot \frac{1}{(n-k)^\ell} < \frac{k(2k \log n)^k}{(n-k)^\ell} \ .\end{aligned} \tag{3.7}$$

To finish the computation we use that $n \geq 16$ and $k \leq n/4 \log n$ to get that $k \leq n/16$, and we observe that $k(16/15)^\ell \leq 2^k$ for every $1 \leq \ell \leq k$. We obtain that

$$\frac{k(2k \log n)^k}{(n-k)^\ell} \leq \frac{k(2k \log n)^k}{(15n/16)^\ell} = k(16/15)^\ell \cdot \frac{(2k \log n)^k}{n^\ell} \leq \frac{(4k \log n)^k}{n^\ell} \ . \tag{3.8}$$

This concludes the proof. ☐

We can use Lemma 3.2 to show that if we hit a sufficiently short resolution refutation of $ERPHP^{k,n}_{k-1}$ with a random restriction ρ, then in the restricted refutation all clauses are likely to have small pigeon-width. The reason this is useful is that the distribution $\mathcal{D}$ is constructed so that the restricted formula is just the standard pigeonhole principle formula, or rather, a 3-CNF version of it (up to renaming of variables). To spell this out explicitly, after renaming the k pigeons in $[n]$ chosen by ρ to $1, \ldots, k$, what remains is the following collection of clauses:

$$\begin{aligned}
&q_{v,1} \vee z_{v,1} && v \in [k], && \text{(3.9a)}\\
&\overline{z}_{v,w} \vee q_{v,w+1} \vee z_{v,w+1} && v \in [k],\, w \in [k-4], && \text{(3.9b)}\\
&\overline{z}_{v,k-3} \vee q_{v,k-2} \vee q_{v,k-1} && v \in [k], && \text{(3.9c)}\\
&\overline{q}_{v,w} \vee \overline{q}_{v',w} && v, v' \in [k],\, v \neq v',\, w \in [k-1]. && \text{(3.9d)}
\end{aligned}$$

But the clauses (3.9a)–(3.9d), which we will denote $EPHP^k_{k-1}$, can easily be shown to require almost maximal pigeon-width in resolution.

Lemma 3.3. *Every resolution refutation of $EPHP^k_{k-1}$ has pigeon-width at least $k-1$.*

Proof. We use a game argument in the style of [Pud00, AD08] adapted to the notion of pigeon-width. The game is played between a *prosecutor* and a *defendant*. At each step of the game the prosecutor queries the defendant for the value of a variable of $EPHP^k_{k-1}$ and stores the answer in his record. The prosecutor is also allowed to erase variable assignments from his record after any query, but if so the defendant can answer differently next time she is asked about an erased variable. The goal of the prosecutor is to force the defendant to falsify a clause from $EPHP^k_{k-1}$, while the goal of the defendant is to answer queries without falsifying any axiom clause in this formula.

To establish the lemma, it is sufficient to show that the prosecutor cannot win unless at some point he holds a record that mentions k pigeons. The reason for this is that if there exists a resolution refutation π of pigeon-width $\ell < k-1$, then the prosecutor can use such a refutation to construct a strategy that never mentions more than $\ell + 1$ pigeons.

To build a winning strategy from a refutation π, the prosecutor walks backwards through the associated graph G_π from the final empty clause all the way to some axiom clause. The invariant maintained is that at each step the current assignment on record is the minimal falsifying assignment for the clause currently visited in G_π. At the beginning of the game the empty record corresponds to the empty clause in the refutation. If the current clause was obtained by resolution, the prosecutor queries the resolved variable (which might temporarily increase the number of pigeons on record by 1), moves to the premise falsified by the answer, and then forgets all assignments not needed to falsify that clause. For a weakening step, the prosecutor just needs to forget variables. The prosecutor wins when the game reaches a source vertex in G_π (if not earlier), since by the invariant the corresponding axiom clause is falsified by the assignment on record at that point.

Switching to the lower-bound perspective, let us now briefly describe a defendant strategy that works against prosecutors mentioning less than k pigeons. The defendant privately keeps a partial matching of the pigeons mentioned in the current record of the prosecutor into holes, making sure that this mapping is compatible with the partial assignment in his record. If the prosecutor asks about a variable which mentions a pigeon already in the domain of the defendant's partial matching, she answers consistently with her matching. If the prosecutor erases all variables mentioning a pigeon, the defendant removes that pigeon from the partial mapping, freeing up the corresponding hole for later reuse. If the prosecutor queries a variable that mentions a new pigeon, we are in one of two cases: either there is at least one free hole, or the record mentions $k-1$ pigeons. In the first case the defendant assigns the new pigeon to some free hole and updates her partial matching accordingly. In the second case the defendant has achieved her goal—although she is now forced to falsify a clause of $EPHP^k_{k-1}$ and loses, the prosecutor was able to win only by compiling a record that mentions k pigeons. ☐

Putting all the pieces together we can now prove the lower bound in Theorem 3.1. Namely, let π be a resolution refutation of $ERPHP^{k,n}_{k-1}$ of size S. Hit π with a random restriction ρ distributed according to $\mathcal{D}$. Since resolution refutations are preserved under restrictions, $\pi{\restriction}_\rho$ is a refutation of $ERPHP^{k,n}_{k-1}\restriction_\rho$ which, as discussed above, is $EPHP^k_{k-1}$ after renaming of variables. By Lemma 3.3, this refutation must have pigeon-width at least $k-1$ with probability 1. On the other hand, using Lemma 3.2 with $\ell = k-1$ and taking a union bound over all clauses in π, the probability that this happens is at most $S \cdot (4k \log n)^k / n^{k-1}$ for large enough n. We can hence conclude that $S \geq n^{k-1}/(4k \log n)^k$, and the proof of Theorem 3.1 is complete.

4 Algebraic and semialgebraic proof systems

Let us now show how the size lower bound for resolution in Section 3 can be generalized to polynomial calculus resolution (PCR) and Sherali-Adams resolution (SAR). The overall structure of the size lower bound proof is very similar to that for resolution in that we first establish a lower bound on a parameter analogous to the pigeon-width in Section 3, which we call *pigeon-degree* for PCR and *pigeon-rank* for SAR, and then plug this bound into the random restriction argument as in the proof of Lemma 3.2.

In this section, we also discuss how upper bounds for PCR and SAR analogous to those for resolution in Theorem 1.1 can be established. The upper bound in resolution more or less immediately carries over to PCR, in the sense that it is very easy to show that a resolution refutation can be simulated easily in PCR in essentially the same size and with PCR degree matching the resolution width. For SA and SAR it requires a bit more work to construct such efficient simulations and we discuss it in some detail below. It should be noted that while PCR degree and SAR rank upper bounds $\mathrm{O}(k)$ are sufficient to obtain refutation of size $n^{\mathrm{O}(k)}$ in both proof systems, using explicit simulations like the ones discussed in this section gives better bounds.

4.1 Lower bound on degree for polynomial calculus resolution

In a natural generalization of the terminology in Section 3, we say that not only the variables $q_{v,w}$ and $z_{v,w}$ of $EPHP^k_{k-1}$ but also their twins $\overline{q}_{v,w}$ and $\overline{z}_{v,w}$ *mention* the pigeon v. The *pigeon-degree* of a monomial is the number of pigeons that are mentioned by its variables, the pigeon-degree of a polynomial is the maximum pigeon-degree of its monomials, and the pigeon-degree of a PCR refutation of $EPHP^k_{k-1}$ is the maximum pigeon-degree of the polynomials in the refutation. The following lower bound for pigeon-degree of PCR refutations is the analogue of Lemma 3.3 for resolution.

Lemma 4.1. *Every PCR refutation of $EPHP^k_{k-1}$ has pigeon-degree at least $\lceil \frac{k-1}{2} \rceil$.*

Proof. We prove the lower bound by studying a different encoding $APHP^k_{k-1}$ of the pigeonhole principle for k pigeons and $k-1$ holes described in [Raz98]. Given any PCR refutation of $EPHP^k_{k-1}$ as defined in (3.9a)–(3.9d) in which all monomials mention at most d pigeons, we show how to transform it into a refutation of degree $d+1$ of $APHP^k_{k-1}$. Since $APHP^k_{k-1}$ requires degree strictly larger than $\lceil \frac{k-1}{2} \rceil$ by Theorem 3.9 in [IPS99], it follows that $d \geq \lceil \frac{k-1}{2} \rceil$.

The alternative formulation $APHP^k_{k-1}$ is defined on variables $x_{v,w}$ for $v \in [k]$ and $w \in [k-1]$, where $x_{v,w} = 1$ means that pigeon v sits in hole w. We stress that this interpretation of the variables is the opposite of the one we use for $EPHP^k_{k-1}$. Also, $APHP^k_{k-1}$ is not a (translation of a) CNF formula but consists of the following polynomials:

$$1 - \sum_{w \in [k-1]} x_{v,w} \qquad v \in [k], \tag{4.1a}$$

$$x_{v,w} x_{v',w} \qquad w \in [k-1],\ v, v' \in [k],\ v \neq v', \tag{4.1b}$$

$$x_{v,w} x_{v,w'} \qquad v \in [k],\ w, w' \in [k-1],\ w \neq w'. \tag{4.1c}$$

To obtain a degree-$(d{+}1)$ refutation for $APHP^k_{k-1}$, the first step is to apply a substitution δ to the variables in the refutation in pigeon-degree d of $EPHP^k_{k-1}$. For q-variables we define $\delta(q_{v,w}) = 1 - x_{v,w}$ and $\delta(\overline{q}_{v,w}) = x_{v,w}$, and for z-variables we let $\delta(z_{v,w}) = 1 - \sum_{j>w} x_{v,j}$ and $\delta(\overline{z}_{v,w}) = 1 - \sum_{j\leq w} x_{v,j}$. This substitution transforms the refutation of $EPHP^k_{k-1}$ into a sequence of polynomials over the variables in $APHP^k_{k-1}$. This is not yet a valid refutation, however, and in order to deal with this we need to show how to derive each substituted polynomial in the sequence. How to do so depends on what rule was used to derive the polynomial before the substitution.

For inference steps, if we derived xP from P then $\delta(xP) = \delta(x)\delta(P)$ can be derived from $\delta(P)$ by a sequence of multiplications and linear combinations, and if the polynomial was derived via a linear combination, then the same derivation step is valid for the substituted polynomials.

If P is an application of the Boolean axiom $x^2 - x$ to a q-variable or z-variable, then $\delta(P)$ can be derived from Boolean axioms combined with polynomials (4.1c). Applications of complementarity axioms are either vacuous (for q-variables) or reduce to (4.1a) (for z-variables).

Finally, we need to show how to derive $\delta(P)$ if P is obtained from one of the clauses in (3.9a)–(3.9d). We describe how to do this for $P = \overline{z}_{v,w} q_{v,w+1} z_{v,w+1}$ as in (3.9b); the other cases are very similar. We have

$$\begin{aligned}\delta(\overline{z}_{v,w} q_{v,w+1} z_{v,w+1}) &= \left(1 - \sum_{j\leq w} x_{v,j}\right)\left(1 - x_{v,w+1}\right)\left(1 - \sum_{j>w+1} x_{v,j}\right) \\ &= 1 - \sum_{j\in[k-1]} x_{v,j} + \sum_{\substack{j\neq j',\, j\leq w+1,\\ j'\geq w+1}} x_{v,j}x_{v,j'}R_{j,j'} \ ,\end{aligned} \tag{4.2}$$

where $1 - \sum_{j\in[k-1]} x_{v,j}$ is (4.1a) and all polynomials $x_{v,j}x_{v,j'}R_{j,j'}$ can be derived by multiplications and linear combinations from (4.1c). Thus, $\delta(\overline{z}_{v,w} q_{v,w+1} z_{v,w+1})$ can be derived from $APHP^k_{k-1}$.

This shows how we can apply the substitution δ to a refutation of $EPHP^k_{k-1}$ to obtain a refutation of $APHP^k_{k-1}$. The substitution exchanges variables indexed by the pigeon v for degree-1 polynomials which mention just v, and therefore each monomial of this refutation mentions at most d pigeons as well. We then postprocess the refutation of $APHP^k_{k-1}$ by removing all monomials that mention the same pigeon twice or more and all the monomials that mention more than one pigeon for the same hole. This is possible using the axioms $x_{v,w}x_{v',w}$ in (4.1b) and $x_{v,w}x_{v,w'}$ in (4.1c), and as a result we obtain a refutation of (total) degree at most $d+1$. The lemma follows. ☐

4.2 Size and rank upper bounds for Sherali-Adams refutations

Let us next switch focus to upper bounds and show that SAR can simulate resolution refutations efficiently in term of size and rank. We remark that a similar simulation is given [DMR09], but since that paper uses a slightly different definition of Sherali-Adams we give a full description of the simulation here for completeness.

We start by introducing notation for two polynomial forms which we will use to represent clauses. For any pair of sets of propositional variables Y, Z, $Y \cap Z = \emptyset$, we let

$$S(Y,Z) = \sum_{y\in Y} y + \sum_{z\in Z} \overline{z} \tag{4.3}$$

and

$$M(Y,Z) = \prod_{y\in Y} \overline{y} \prod_{z\in Z} z \ . \tag{4.4}$$

Consider a clause $C = \bigvee_{y \in V_C^+} y \vee \bigvee_{z \in V_C^-} \overline{z}$ where V_C^+ and V_C^- are the sets of variables appearing positively and negatively in C, respectively. Then we define

$$S(C) = S(V_C^+, V_C^-) \tag{4.5}$$

and

$$M(C) = M(V_C^+, V_C^-) \ . \tag{4.6}$$

Observe that for any assignment of the variables to $\top = 1$ or $\bot = 0$ it holds that $S(C) - 1 \geq 0$ and $-M(C) \geq 0$ if and only if C is satisfied. The former, additive inequality is how clauses are translated to inequalities as discussed in Section 2, but for our simulation of resolution by SAR we will need to work with the latter, multiplicative version.

The following three lemmas show how to efficiently simulate the steps in a resolution derivation.

Lemma 4.2 (Simulation of axiom). *For a clause C of width w the inequality $-M(C) \geq 0$ has a derivation in SAR of rank $w+1$ and size $O(w^2)$ from the inequality $S(C) - 1 \geq 0$.*

Proof. If C is the empty clause then the claim is obvious since in that case $-M(C) = S(C) - 1$. Let C be non-empty and assume for simplicity that it has a positive literal x. Then C has the form

$$x \vee \bigvee_{y \in Y} y \vee \bigvee_{z \in Z} \overline{z} \tag{4.7}$$

with $|Y| + |Z| < w$. By multiplying $S(C) - 1 \geq 0$ by $M(Y, Z)$ we obtain the polynomial inequality

$$xM(Y,Z) + \sum_{y \in Y} yM(Y,Z) + \sum_{z \in Z} \overline{z}M(Y,Z) - M(Y,Z) \geq 0 \ . \tag{4.8}$$

For each $y \in Y$ we can derive

$$\begin{aligned}(1 - y - \overline{y}) \cdot yM(Y \setminus \{y\}, Z) + (y^2 - y) \cdot M(Y \setminus \{y\}, Z) = \\ - y \cdot \overline{y} \cdot M(Y \setminus \{y\}, Z) = -y \cdot M(Y,Z) \ .\end{aligned} \tag{4.9}$$

In essentially the same way we can derive

$$-\overline{z}M(Y,Z) \geq 0 \tag{4.10}$$

for each $z \in Z$. The inequality

$$-M(C) = -\overline{x}M(Y,Z) \geq 0 \tag{4.11}$$

is now the sum of inequality (4.8), all inequalities of the form (4.9) and (4.10) for all $y \in Y$ and $z \in Z$, and of the inequality

$$(1 - x - \overline{x})M(Y,Z) \geq 0 \ . \tag{4.12}$$

This SAR derivation has size $O(w^2)$ and rank w. □

Lemma 4.3 (Simulation of weakening). *For clauses $A \subseteq B$ of width at most w the inequality*

$$M(A) - M(B) \geq 0$$

has a derivation in SAR of rank $w+1$ and size $O(w^2)$.

Proof. Let Y and Z be the set of variables that occur positively and negatively, respectively, in $B \setminus A$, so that $M(B) = M(A) \cdot M(Y,Z)$. Note that $M(Y,Z)$ is the product of the literals in $B \setminus A$, which are all variables in SAR. For ease of notation, let us write this product as $\prod_{i=1}^{|Y|+|Z|} v_i$. Then by using telescoping sums we can derive

$$\sum_{i=1}^{|Y|+|Z|} (1-v_i)M(A)\prod_{j=1}^{i-1} v_i = (1-M(Y,Z))M(A) = M(A) - M(B) \tag{4.13}$$

which establishes the lemma. □

Lemma 4.4 (Simulation of resolution step). *Let A and B be clauses in which the variable x does not appear and let w be the width of $A \vee B$. Then the inequality*

$$M(A \vee x) + M(B \vee \overline{x}) - M(A \vee B) \geq 0$$

has a derivation in SAR of rank $w+1$ and size $O(w^2)$.

Proof. Using Lemma 4.3 twice we derive the two inequalities $M(A \vee x) - M(A \vee B \vee x) \geq 0$ and $M(A \vee \overline{x}) - M(A \vee B \vee \overline{x}) \geq 0$. Then we derive $(x + \overline{x} - 1)M(A \vee B) \geq 0$ from the axiom $x + \overline{x} - 1 \geq 0$. This is the same as

$$M(A \vee B \vee \overline{x}) + M(A \vee B \vee x) - M(A \vee B) \geq 0 \ . \tag{4.14}$$

The inequality that we want to prove is the sum of these three inequalities just derived. This SAR derivation has size $O(w^2)$ and rank w. □

Remark 4.5. In Lemmas 4.2, 4.3 and 4.4 we gave the SAR simulations of the steps of a resolution refutation. To get a simulation in SA it is sufficient to substitute $(1-x_1), \ldots, (1-x_n)$ for the variables $\overline{x}_1, \ldots, \overline{x}_n$. After the substitution we obtain a valid SA proof of the corresponding inequalities of the same rank, but potentially of larger size. Notice that the proofs of the inequalites in Lemmas 4.2, 4.3 and 4.4 have the form of Equation (2.1), with $O(w)$ axioms, each of them multiplied by a degree $w + O(1)$ polynomial. Hence the size of each of these proofs is at most $O(w2^w)$.

Now we can show how resolution refutations can be efficiently simulated in the SA and SAR proof systems.

Lemma 4.6. *If a CNF formula F has a resolution refutation of width w and length L, then it has an SA refutation of rank $w+1$ and size $\mathrm{O}\big(w2^w L\big)$ and an SAR refutation of rank $w+1$ and size $\mathrm{O}\big(w^2 L\big)$.*

Proof. Let $\pi = (C_1, C_2, \ldots, C_L)$ be a resolution refutation of F where all clauses have width at most w. Let us focus first on the SAR simulation. For each clause C_i in the refutation we derive an inequality as follows:

1. If C_i is an axiom clause, then we derive $-M(C_i) \geq 0$.
2. If C_i is obtained by weakening from C_j, then we derive $M(C_j) - M(C_i) \geq 0$.
3. If C_i is obtained by resolving C_j and C_k, then we derive $M(C_j) + M(C_k) - M(C_i) \geq 0$.

All of these inequalities have SAR derivations of rank $w+1$ and size $\mathrm{O}\big(w^2\big)$ by Lemmas 4.2, 4.3 and 4.4 (where we recall that the encoding of an axiom clause C in the SAR proof system is $S(C) - 1 \geq 0$, as required by Lemma 4.2).

Now we have a sequence of inequalities $Q_1 \geq 0, Q_2 \geq 0, \ldots, Q_L \geq 0$, where the inequality $Q_i \geq 0$ corresponds to the clause C_i as explained above. Observe that any positive combination $\sum_{i=1}^{L} \alpha_i Q_i \geq 0$ has a SAR derivation of rank $w+1$ and size $O(L \cdot k^2)$. In order to conclude the proof of the lemma, we just need to argue that there are positive weights α_i such that $\sum_i \alpha_i Q_i = -1$.

The intuition is that if C_i is obtained by weakening from C_j then adding $Q_i = M(C_j) - M(C_i)$ will cancel the term $-M(C_j)$ in Q_j representing C_j, and if C_i is inferred by resolution from C_j and C_k, then adding $Q_i = M(C_j) + M(C_k) - M(C_i)$ will cancel the terms $-M(C_j)$ and $-M(C_k)$ representing C_j and C_k in Q_j and Q_k, respectively. In the end, all monomials representing clauses are cancelled and the only term remaining is -1. However, if a clause is used in several different applications of the resolution or weakening rules we need to set the weights so that it is cancelled the correct number of times.

To do so, consider the DAG of the resolution refutation oriented from the initial clauses towards the empty clause. We assign a weight to each clause C_i in this DAG inductively: the empty clause C_L gets weight 1, and if all immediate successors of a clause have already been assigned weights, then the clause gets the sum of the weights of its immediate successors as the weight for itself. The value of α_i is then the weight assigned to the clause C_i in this way. To verify that $\sum_i \alpha_i Q_i = -1$, notice that every polynomial $M(C_i)$ has negative coefficient in the inequality $Q_i \geq 0$ and positive one in every $Q_j \geq 0$ where C_i appears as a premise in the derivation of C_j. By construction the coefficient of each $M(C_i)$ in the final sum is zero unless $i = L$. Since $\alpha_L = 1$, the final sum is equal to $-M(\emptyset, \emptyset)$ which is -1.

We can obtain a simulation in the SA proof system instead by substituting $(1 - x_i)$ for every negative variable $\overline{x}$ in the SAR simulation described above. Then we can reason as in Remark 4.5 to see that the the size and rank bounds claimed for SA hold. The lemma follows. □

4.3 Lower bound on rank for Sherali-Adams resolution

The *pigeon-rank* of a Sherali-Adams resolution refutation of $EPHP^k_{k-1}$ of the form described in Equation (2.1) is the maximum pigeon-degree of the polynomials to which the formulas $\prod_{i\in\mathcal{I}_t} x_i \cdot \prod_{j\in\mathcal{J}_t}(1 - x_j) \cdot P_t$ expand.

In order to prove a lower bound on pigeon-rank it is useful to generalize this concept to a more abstract notion of rank for SA proofs. Let V be a set of variables and let H be a downward-closed family of subsets of V, i.e., such that if Y belongs to H and $X \subseteq Y$, then X also belongs to H. We say that a polynomial (or polynomial inequality) is *H-bounded*, or has *H-bounded rank*, if H contains the variable set of every monomial in it. We say that an SA derivation as in (2.1) has H-bounded rank if the polynomial to which each formula $\prod_{i\in\mathcal{I}_t} x_i \cdot \prod_{j\in\mathcal{J}_t}(1-x_j) \cdot P_t$ expands is H-bounded. Observe that if an SA derivation has rank r, then it has H-bounded rank where H is the family of all subsets of at most r variables. Similarly, if an SA refutation of $EPHP^k_{k-1}$ has pigeon-rank r, then it has H-bounded rank where H is the family of all subsets of variables that mention at most r pigeons.

Let $\mathcal{P}$ be a set of polynomial inequalities over the variable set V. We say that $\mathcal{P}$ admits an *H-consistent family of distributions* if there exists a collection of probability distributions $\{\Pi_X\}_{X\in H}$ over assignments $\{0,1\}^X$ as X ranges over H that satisfy the following properties:

H1. For every variable set $X \in H$ and every polynomial inequality $Q \geq 0$ in $\mathcal{P}$ that has all its variables in X, it holds that all assignments in the support of Π_X satisfy $Q \geq 0$.

H2. For every pair of variable sets $X, Y \in H$ such that $X \subseteq Y$ and for every assignment $\mu \in \{0,1\}^X$ it holds that
$$\Pi_X(\mu) = \sum_{\substack{\eta\in\{0,1\}^Y \\ \eta\supseteq\mu}} \Pi_Y(\eta) \ , \tag{4.15}$$
where η ranges over all assigments to Y that are consistent with μ.

In the definition above and elsewhere, $\Pi_X(\mu)$ denotes the probability assigned to μ by the distribution Π_X. We will use such H-consistent families of distributions to establish the Sherali-

Adams rank lower bound that we need. Before stating the formal lemma that we will appeal to, let us try to provide some intuition.

If the set of polynomial inequalities $\mathcal{P}$ were satisfiable it would not be hard to come up with a family of probability distributions with properties H1 and H2: we could just fix a global probability distribution over all satisfying assignments, and then let Π_X be the corresponding marginal distribution on any set of variables X. For an unsatisfiable set $\mathcal{P}$ there is no such globally consistent family, but if we can find an H-consistent family of distributions for $\mathcal{P}$, then $\mathcal{P}$ will still "look satisfiable" to any derivation that does not go "outside of H." Whenever we look at a specific inequality $Q \geq 0$ in $\mathcal{P}$, property H1 yields a "marginal distribution" that satisfies the inequality. Furthermore, property H2 ensures that such "marginal distributions" over different sets look locally consistent. The following lemma makes this precise.

Lemma 4.7. *Let H be a downward-closed family of sets of variables and let $\mathcal{P}$ be a set of H-bounded polynomial inequalities. If $\mathcal{P}$ has an SA refutation of H-bounded rank, then $\mathcal{P}$ does not admit an H-consistent family of distributions.*

Proof. Let us think of each X in H as a new formal variable. For each monomial M, let X_M denote the set of variables in M. If R is an H-bounded polynomial, let us write $\widehat{R}$ to denote the linear form on the variables H obtained from R by replacing each term $c \cdot M$ by $c \cdot X_M$ and collecting all terms of the same variable into a single term by adding their coefficients (which could result in cancellations of terms). Note that $\widehat{R}$ can also be thought of as the multilinearization of R, namely the polynomial obtained from R by removing all higher powers in the monomials to get $\widehat{M} = X_M$ instead of M. We write 1_Y to denote the assignment $\{x \mapsto 1 : x \in Y\}$ to a set of variables Y, and for a monomial M (multilinear or not) we define $1_M = 1_{X_M}$.

Let $\mathcal{P} = \{Q_1 \geq 0, \ldots, Q_m \geq 0\}$ be a set of polynomial inequalities and suppose that there exists an SA refutation of $\mathcal{P}$ of the form (2.1) that has H-bounded rank. Let us write R_t for the polynomial to which the formula $\prod_{i \in \mathcal{I}_t} x_i \cdot \prod_{j \in \mathcal{J}_t}(1 - x_j) \cdot P_t$ expands for $1 \leq t \leq \tau$. The assumption that the refutation has H-bounded rank means that every monomial in the polynomial R_t is H-bounded.

Assume for contradiction that $\mathcal{P}$ admits an H-consistent family $\{\Pi_X\}_{X \in H}$. Let $a : H \to \mathbb{R}$ be the real-valued assignment defined by

$$a(X) = \Pi_X(1_X) \ , \tag{4.16}$$

i.e., the probability of the all-ones assignment to the variables in X according to the distribution Π_X, and extend a to all linear forms on the variables $X \in H$ linearly; i.e., if $L = \sum_i c_i X_i$ is such a linear form with coefficients c_i and variables X_i, then $a(L) = \sum_i c_i \cdot a(X_i)$.

We claim that a satisfies $a(\widehat{R}_t) \geq 0$ for every $1 \leq t \leq \tau$. By linearity it then further follows that $a\big(\sum_{t=1}^{\tau} \alpha_t \widehat{R}_t\big) = \sum_{t=1}^{\tau} \alpha_t \cdot a\big(\widehat{R}_t\big) \geq 0$, which is a contradiction since $\sum_{t=1}^{\tau} \alpha_t R_t = -1$ and hence also $\sum_{t=1}^{\tau} \alpha_t \widehat{R}_t = -1$.

Let us prove that the assignment a as defined in (4.16) satisfies every inequality $\widehat{R}_t \geq 0$ for $1 \leq t \leq \tau$. We do so by establishing a stronger claim: if X_t is the set of variables in R_t and $\mathbb{E}_{X_t}$ denotes expectation under the distribution Π_{X_t}, then the following holds:

A1. The assignment $a : H \to \mathbb{R}$ satisifies $a(\widehat{R}_t) = \mathbb{E}_{X_t}[R_t]$.

A2. Every assignment in the support of Π_{X_t} satisfies the inequality $R_t \geq 0$.

To see that A1 holds, we evaluate each monomial M in R_t separately to get

$$a(\widehat{M}) = a(X_M) = \Pi_{X_M}(1_M) = \sum_{\substack{\eta \in \{0,1\}^{X_t} \\ \eta \supseteq 1_M}} \Pi_{X_t}(\eta) = \mathbb{E}_{X_t}[M] \ . \tag{4.17}$$

The first and second equalities in (4.17) hold by definition; the third one follows from property H2 of H-consistent families of distributions; and the final equality is true since a monomial M evaluates to 1 under an assignment $\eta \in \{0,1\}^{X_t}$ if and only if η is compatible with 1_M. Adding over all terms we get $a(\widehat{R}_t) = \mathbb{E}_{X_t}[R_t]$ by applying linearity of a on the left and linearity of expectation on the right.

The verification of the claim in A2 is straightforward. Let η be an assignment in the support of Π_{X_t}. Substituting the values assigned by η to the variables of R_t, we deduce that

$$\eta(R_t) = \eta\left(\prod_{i\in\mathcal{I}_t} x_i \cdot \prod_{j\in\mathcal{J}_t}(1-x_j)\cdot P_t\right) = \prod_{i\in\mathcal{I}_t}\eta(x_i)\cdot\prod_{j\in\mathcal{J}_t}(1-\eta(x_j))\cdot\eta(P_t) \geq 0\ . \tag{4.18}$$

To see this, it suffices to observe that all factors in the final expression in (4.18) are non-negative. First, regardless of what the assignment η is, we clearly have $0 \leq \eta(x) \leq 1$ for any variable x in its domain and hence $\eta(x_i) \geq 0$ and $1-\eta(x_j) \geq 0$. Second, from property H1 we know that if P_t is one of the polynomials Q_i in $\mathcal{P}$ then $\eta(P_t) \geq 0$ since η is in the support of Π_{X_t}. And third, if P_t is one of the axioms $x_i^2 - x_i$ or $x_i - x_i^2$ then $\eta(P_t) = 0$ since the range of η is $\{0,1\}$, and if P_t is the axiom 1 then of course $\eta(P_t) = 1 \geq 0$. This concludes the proof of the lemma. ☐

Dantchev et al. [DMR09] proved a rank lower bound on SAR refutations of PHP^k_{k-1}. Let us show how this result can be extended to a pigeon-rank lower bound for $EPHP^k_{k-1}$.

Lemma 4.8. *Every SAR refutation of $EPHP^k_{k-1}$ has pigeon-rank at least k.*

Proof. First note that by replacing each variable $\overline{x}$ by $1-x$ we transform an SAR proof into an SA proof of the same pigeon-rank. Thus, by Lemma 4.7 it will suffice to build an H-consistent family of distributions where H is the family of sets of variables that mention up to $k-1$ pigeons.

Intuitively, it is clear what the distributions should be: since there is room for up to $k-1$ pigeons in the pigeonholes, we can just choose any one-to-one mapping uniformly at random and set the Boolean variables accordingly. Formally, for every set X of variables than mention at most $k-1$ pigeons we define the distribution Π_X as follows:

1. Let A be the set of at most $k-1$ pigeons that are mentioned by the variables in X.
2. Let φ be a uniformly chosen one-to-one map $\varphi : A \to [k-1]$.
3. For $q_{v,w} \in X$ set $q_{v,w} = 1$ if $\varphi(v) = w$, and $q_{v,w} = 0$ otherwise.
4. For $z_{v,w} \in X$ set $z_{v,w} = 1$ if $\varphi(v) > w$, and $z_{v,w} = 0$ otherwise.

Let us verify that a family of distributions defined in this way satisfy properties H1 and H2.

That property H1 is satisfied is immediate by construction. If C is a clause in $EPHP^k_{k-1}$ with all variables contained in X, then all assignments in the support of Π_X satisfy C since they encode one-to-one mappings (with the extension variables $z_{v,w}$ set appropriately).

Property H2 is also straightforward to verify. Fix any sets X and Y such that $X \subseteq Y$ and that mention up to $k-1$ pigeons and any assignment $\mu \in \{0,1\}^X$. Let A and B be the sets of at most $k-1$ pigeons that are mentioned in X and Y, respectively, and note that $A \subseteq B$. Let us write $a = |A|$ and $b = |B|$. By construction, the assignments $\eta \in \{0,1\}^Y$ in the support of Π_Y are in bijective correspondence with the one-to-one mappings $\psi : B \to [k-1]$ and the same holds for μ in the support of Π_X vis-a-vis $\varphi : A \to [k-1]$. Moreover, each one-to-one mapping $\varphi : A \to [k-1]$ can be chosen in $(k-1)(k-2)\cdots(k-a) = \binom{k-1}{a}a!$ ways, and for a fixed φ the number of one-to-one mappings $\psi : B \to [k-1]$ that extend φ is $(k-a-1)(k-a-2)\cdots(k-b) = \binom{k-1-a}{b-a}(b-a)!$. Since all involved distributions are uniform over their support, for $\mu \in \{0,1\}^X$ in the support of Π_X we have

$$\sum_{\substack{\eta\in\{0,1\}^Y\\ \eta\supseteq\mu}} \Pi_Y(\eta) = \sum_{\substack{\eta\,:\,\Pi_Y(\eta)>0\\ \eta\supseteq\mu}} \frac{1}{\binom{k-1}{b}b!} = \frac{\binom{k-1-a}{b-a}(b-a)!}{\binom{k-1}{b}b!} = \frac{1}{\binom{k-1}{a}a!} = \Pi_X(\mu) \tag{4.19}$$

and for μ outside the support of Π_X the whole summation in (4.19) is zero. This finishes the proof of the lemma. $\square$

4.4 Size bounds for PCR and SAR refutations

Given the lower bounds on pigeon-degree and pigeon-rank for refuting $EPHP^k_{k-1}$ in Lemmas 4.1 and 4.8, respectively, the size lower bounds on refutations of $ERPHP^{k,n}_{k-1}$ in polynomial calculus resolution and Sherali-Adams resolution are straightforward adaptions of the lower bound for resolution in Theorem 3.1. We write down the details here for completeness, starting with the PCR bounds.

Theorem 4.9. *Let $k = k(n)$ be any integer-valued function such that $k(n) \leq n/4\log n$. Then $ERPHP^{k,n}_{k-1}$ can be refuted in PCR in size $\mathrm{O}\big(k^{k+1}n^k\big)$, and any PCR refutation requires size $\Omega\big(n^{\lceil (k-1)/2\rceil}/(4k\log n)^k\big)$.*

Proof. Fix any PCR refutation of $ERPHP^{k,n}_{k-1}$ and let $\mathcal{M}$ be the set of monomials appearing in it. We hit the refutation with a random restriction ρ distributed according to $\mathcal{D}$. Since restrictions preserve PCR derivations we obtain a refutation of $ERPHP^{k,n}_{k-1}{\restriction}_\rho$, which as before is $EPHP^k_{k-1}$ after renaming of variables.

Assume that $|\mathcal{M}| < n^{\lceil (k-1)/2\rceil}/(4k\log n)^k$. Applying Lemma 3.2 with $\ell = \left\lceil \frac{k-1}{2}\right\rceil$ and taking a union bound over the monomials in $\mathcal{M}$, we conclude that there must be at least one restriction ρ in the support of $\mathcal{D}$ such that the pigeon-degree of $\pi{\restriction}_\rho$ is at most $\left\lceil \frac{k-1}{2}\right\rceil - 1$ if n is large enough. This contradicts Lemma 4.1, and hence $|\mathcal{M}|$ must be at least $n^{\lceil (k-1)/2\rceil}/(4k\log n)^k$.

To obtain the upper bound we start with the resolution refutation in Theorem 3.1. It is not hard to see that any resolution refutation of size S and width w translates into a PCR refutation of size wS and degree $w+1$. The additional factor w in the size is due to the fact that while resolution can arbitrary weaken a clause in one step, the way multiplication is defined in PCR means that we need one multiplication step per literal to simulate the same weakening. $\square$

The proof of the bounds for Sherali-Adams is very similar.

Theorem 4.10. *Let $k = k(n)$ be any integer-valued function such that $k(n) \leq n/4\log n$. Then $ERPHP^{k,n}_{k-1}$ can be refuted in SAR in size $\mathrm{O}\big(k^{k+2}n^k\big)$, and any SAR refutation requires size $\Omega\big(n^k/(4k\log n)^k\big)$.*

Proof. Fix any SAR refutation of $ERPHP^{k,n}_{k-1}$ and let $\mathcal{M}$ be the set of monomials appearing in it. Hit the refutation with a random restriction ρ distributed according to $\mathcal{D}$. Since restrictions preserve soundness of SAR proofs, this yields a refutation of $ERPHP^{k,n}_{k-1}{\restriction}_\rho$, which is $EPHP^k_{k-1}$.

Suppose now that $|\mathcal{M}| < n^k/(4k\log n)^k$. Using Lemma 3.2 with $\ell = k$ and a union bound argument for $\mathcal{M}$, we conclude that there exists at least one restriction ρ in the support of $\mathcal{D}$ such that the pigeon-rank of $\pi{\restriction}_\rho$ is at most $k-1$, assuming that n large enough. But this contradicts Lemma 4.8, and hence the lower bound in the theorem follows.

We obtain the upper bound by using the simulation in Lemma 4.6 on the resolution refutation in Theorem 3.1. $\square$

5 An upper bound for relativized PHP formulas in Lasserre

In this section, we show that our lower bound Theorem 1.1 does not generalize to Lasserre but that the formulas $ERPHP^{k,n}_{k-1}$ (and also $RPHP^{k,n}_{k-1}$) have Lasserre refutations in constant rank. To establish this we we will use the easily verified identity

$$\sum_{\substack{i,j\in[n]\\ i\neq j}} \Big(1 - z_i - z_j\Big) z_j + (n-2)\sum_{j\in[n]}\Big(z_j^2 - z_j\Big) + \Big(1 - \sum_{i\in[n]} z_i\Big)^2 = 1 - \sum_{i\in[n]} z_i \tag{5.1}$$

a couple of times. A direct application of (5.1) shows that the inequality $1 - \sum_{i\in[n]} z_i \geq 0$ has a rank-2 Lasserre derivation from the set of all inequalities of the form $1 - z_i - z_j \geq 0$ for $i, j \in [n]$, $i \neq j$. We remark that this fact is a direct consequence of Lemma 1.5 in [LS91]. Let us first use this to get a rank-2 Lasserre refutation of the standard pigeonhole principle PHP^k_{k-1} encoded as the set of clauses

$$x_{u,1} \vee x_{u,2} \vee \cdots \vee x_{u,k-1} \qquad u \in [k], \tag{5.2a}$$
$$\overline{x}_{u,w} \vee \overline{x}_{v,w} \qquad u, v \in [k],\ u \neq v,\ w \in [k-1]. \tag{5.2b}$$

The proof we give next is essentially due to Grigoriev et al. [GHP02].

Lemma 5.1 (**[GHP02]**). *The formulas PHP^k_{k-1} have Lasserre refutations of rank 2.*

Proof. Combining all hole axioms $1 - x_{u,w} - x_{v,w} \geq 0$ in (5.2b) for a fixed hole $w \in [k-1]$ and using (5.1) we can get the inequality $1 - \sum_{u\in[k]} x_{u,w} \geq 0$. Adding these inequality over all holes $w \in [k-1]$ we obtain

$$k - 1 - \sum_{u\in[k]} \sum_{w\in[k-1]} x_{u,w} \geq 0 \ . \tag{5.3}$$

Adding together instead all the pigeon axioms $\sum_{w\in[k-1]} x_{u,w} - 1 \geq 0$ in (5.2a) we get

$$\sum_{u\in[k]} \sum_{w\in[k-1]} x_{u,w} - k \geq 0 \ . \tag{5.4}$$

Summing (5.3) and (5.4) yields $-1 \geq 0$. □

Two more applications of (5.1) will help us get rank-9 Lasserre refutations of $RPHP^{k,n}_{k-1}$ (and $ERPHP^{k,n}_{k-1}$) by reduction to PHP^k_{k-1}. The main idea of the proof is to substitute variables in the derivation in Lemma 5.1 with polynomials defined over the variables of $RPHP^{k,n}_{k-1}$.

Lemma 5.2. *The formulas $RPHP^{k,n}_{k-1}$ and $ERPHP^{k,n}_{k-1}$ have Lasserre refutations of rank 9.*

Proof. Let us first observe that we only need to present the Lasserre refutation of $RPHP^{k,n}_{k-1}$. Once we have a refutation of the original formula $RPHP^{k,n}_{k-1}$ we immediately obtain a refutation of the 3-CNF version $ERPHP^{k,n}_{k-1}$ by using the observation that the encoding of a wide clause $C = a_1 \vee \ldots a_w$ is the the sum of the encodings of the corresponding 3-clauses $a_1 \vee a_2 \vee z_2$, $\overline{z}_2 \vee a_3 \vee z_3$, $\ldots$, $\overline{z}_{w-2} \vee a_{w-1} \vee a_w$. This is so since all extension variables appear exactly once positively and exactly once negatively and so will simply cancel. Thus, once we have a refutation of $RPHP^{k,n}_{k-1}$ we can get a valid refutation of $ERPHP^{k,n}_{k-1}$ of the same rank by substituting the sum of the corresponding short axioms in $ERPHP^{k,n}_{k-1}$ for any long axiom in $RPHP^{k,n}_{k-1}$.

For the rest of the proof we therefore focus on $RPHP^{k,n}_{k-1}$. Let $\mathcal{P}$ be the set of polynomial inequalities that encode it and let us define the shorthand

$$x_{u,w} = \sum_{\ell\in[n]} p_{u,\ell}\, r_\ell\, q_{\ell,w} \ . \tag{5.5}$$

We want to use the proof of the pigeonhole principle in Lemma 5.1 together with the substitution (5.5) for $x_{u,w}$. In order to do so, we need to show how to derive the substituted axioms used in that proof. The inequalities $x_{u,w}^2 - x_{u,w} \geq 0$ can be obtained by summing

$$\sum_{\substack{\ell,m\in[n]\\ \ell\neq m}} (3 - q_{\ell,w} - q_{m,w} - r_\ell - r_m)\, q_{\ell,w}\, q_{m,w}\, p_{u,\ell}\, p_{u,m}\, r_\ell\, r_m \geq 0 \ , \tag{5.6}$$

$$\sum_{\substack{\ell,m\in[n]\\ \ell\neq m}} (r_\ell^2 - r_\ell) r_m\, q_{\ell,w}\, q_{m,w}\, p_{u,\ell}\, p_{u,m} + (r_m^2 - r_m) r_\ell\, q_{\ell,w}\, q_{m,w}\, p_{u,\ell}\, p_{u,m} \geq 0 \ , \tag{5.7}$$

$$\sum_{\substack{\ell,m\in[n]\\ \ell\neq m}} q_{\ell,w}^2\, q_{m,w}\, p_{u,\ell}\, p_{u,m}\, r_\ell\, r_m + q_{\ell,w}\, q_{m,w}^2\, p_{u,\ell}\, p_{u,m}\, r_\ell\, r_m \geq 0 \ , \tag{5.8}$$

and

$$\sum_{\ell\in[n]} (p_{u,\ell}^2 - p_{u,\ell})\, r_\ell^2\, q_{\ell,w}^2 + (r_\ell^2 - r_\ell)\, p_{u,\ell}\, q_{\ell,w}^2 + (q_{\ell,w}^2 - q_{\ell,w})\, p_{u,\ell}\, r_\ell \geq 0 \ , \tag{5.9}$$

and the latter inequalities all have direct rank-7 derivations from $\mathcal{P}$. To derive the inequalities $\sum_{w\in[k-1]} x_{u,w} - 1 \geq 0$ for $u \in [k]$ we can sum up

$$\sum_{\ell\in[n]} \Big(\sum_{w\in[k-1]} q_{\ell,w} - r_\ell \Big) p_{u,\ell}\, r_\ell \geq 0 \ , \tag{5.10}$$

$$\sum_{\ell\in[n]} \Big(r_\ell^2 - r_\ell \Big) p_{u,\ell} + \sum_{\ell\in[n]} \Big(r_\ell - p_{u,\ell} \Big) p_{u,\ell} + \sum_{\ell\in[n]} \Big(p_{u,\ell}^2 - p_{u,\ell} \Big) \geq 0 \ , \tag{5.11}$$

and

$$\sum_{\ell\in[n]} p_{u,\ell} - 1 \geq 0 \ , \tag{5.12}$$

which can all be derived directly from $\mathcal{P}$ in rank 3. The inequality $1 - x_{u,w} - x_{v,w} \geq 0$ is the sum of

$$\sum_{\ell\in[n]} \Big(1 - p_{u,\ell} - p_{v,\ell} \Big)\, r_\ell\, q_{\ell,w} \geq 0 \tag{5.13}$$

and

$$1 - \sum_{\ell\in[n]} r_\ell q_{\ell,w} \geq 0 \ , \tag{5.14}$$

where (5.13) has a direct rank-3 derivation from $\mathcal{P}$. For (5.14) we need to do some more work. Fix indices $\ell, m \in [n]$ with $\ell \neq m$ and observe that

$$\begin{aligned} &\big(1 - r_\ell q_{\ell,w} - r_m q_{m,w}\big)\, r_\ell q_{\ell,w} = \\ &\quad \big(3 - r_\ell - r_m - q_{\ell,w} - q_{m,w}\big)\, q_{\ell,w}\, r_m\, q_{m,w} + \big(q_{\ell,w}^2 - q_{\ell,w}\big)\, r_m\, q_{m,w} + \big(q_{m,w}^2 - q_{m,w}\big)\, r_m\, q_{\ell,w} \\ &\qquad + \big(r_m^2 - r_m\big)\, q_{\ell,w}\, q_{m,w} + \big(r_\ell - r_\ell^2\big)\, q_{\ell,w} + \big(q_{\ell,w} - q_{\ell,w}^2\big)\, r_\ell^2 \ . \end{aligned} \tag{5.15}$$

Note that the first term on the right-hand side of this equation is the polynomial translation of axiom (3.1e). Writing z_ℓ for $r_\ell\, q_{\ell,w}$, this shows that the inequality $(1 - z_\ell - z_m) z_\ell \geq 0$ has a rank-4 derivation from $\mathcal{P}$. Combined with the fact that $z_\ell^2 - z_\ell = (r_\ell^2 - r_\ell) q_\ell^2 + (q_\ell^2 - q_\ell) r_\ell$, equation (5.1) gives a rank-4 derivation of $1 - \sum_{\ell\in[n]} z_\ell \geq 0$. This is precisely (5.14).

Now we mimic the refutation of PHP_{k-1}^k in Lemma 5.1. For a fixed $w \in [k-1]$ we can use the derivations of $1 - x_{u,w} - x_{v,w} \geq 0$ and $x_{v,w}^2 - x_{v,w} \geq 0$ in combination with (5.1) to obtain the inequality $1 - \sum_{u \in [k]} x_{u,w} \geq 0$ by a rank-9 derivation. Adding all such inequalities for $w \in [k-1]$ gives

$$k - 1 - \sum_{w \in [k-1]} \sum_{u \in [k]} x_{u,w} \geq 0 \ . \tag{5.16}$$

On the other hand, adding $\sum_{w \in [k-1]} x_{u,w} - 1 \geq 0$ over all $u \in [k]$ yields

$$\sum_{u \in [k]} \sum_{w \in [k-1]} x_{u,w} - k \geq 0 \tag{5.17}$$

in rank 3 (the rank of the derivation of $\sum_{w \in [k-1]} x_{u,w} - 1 \geq 0$), and a final addition allows us to derive $-1 \geq 0$, never going above rank 9. $\square$

6 Concluding remarks

In this paper, we exhibit a family of 3-CNF formulas over n variables that can be refuted in resolution in width w but require refutations of size $n^{\Omega(w)}$. Furthermore, this lower bound can be extended to polynomial calculus resolution (PCR) and Sherali-Adams. This shows that the seemingly naive counting upper bounds on proof size in terms of width for resolution, degree for PCR, and rank for Sherali-Adams are actually all tight up to small constant factors in the exponent. Furthermore, our lower bound for resolution also implies that the result in [AFT11] that CNF formulas refutable in width w can be decided by CDCL solvers in time $n^{\mathrm{O}(w)}$ is optimal (again up to constant factors in the exponent), since any resolution refutation the solver finds might have to be that large in the worst case.

Regarding open problems, perhaps the most obvious one concerns the tightness of our result. Our formulas have roughly $N = n^2$ variables and are refutable in width roughly $w = 2k$, and our size lower bound is on the order of $n^k = N^{w/4}$. However, the direct counting argument for width w gives an upper bound of about N^w clauses. Could this gap in the exponent be closed? If so, this would have to be for a different formula family since ours has an upper bound of roughly $n^k = N^{w/4}$. One point worth noting is that one can shave a factor 2 off the gap in the exponent by considering the 4-CNF formulas obtained if the 4-clauses in (3.1e) are *not* converted to 3-CNF. In this case, the same upper and lower bounds still hold, but the number of variables is on the order of $N = kn$, which means that we get a lower bound of the form $N^{w/2}$ if we focus on width w upper-bounded by a constant.

A more fundamental question is whether we can find a formula family that exhibits the same kind of hardness for Lasserre. As shown in this paper, the formulas we used for resolution, PCR and Sherali-Adams will not work. For tree-like Lovász-Schrijver (LS), however, we believe that our formulas should be hard (and that the method of proof should be similar, with long *paths* in the refutation tree playing the role of long monomials). In view of the Lasserre upper bound, for tree-like LS^+ we do not know what to believe. The main problem with our formulas is that after restriction we obtain a pigeonhole principle which is hard for resolution, PCR, and Sherali-Adams (in term of rank) but easy for LS^+. A way to get a similar lower bound for Lasserre might be to find a formula that is hard for Lasserre rank and becomes hard for Lasserre size after relativization.

A natural formula for which it would be interesting to prove similar size lower bounds as in this paper is the so-called *clique formula* claiming that there is a k-clique in some fixed n-vertex graph chosen so that this claim is false. It has been conjectured (e.g., in [BGLR12]) that such formulas require resolution refutation size $n^{\Omega(k)}$ for the right kind of graphs, and this has been proven for the restricted case of tree-like resolution [BGL13]. If such a lower bound could be established for general resolution, it would have interesting consequences for parameterized proof complexity.

Finally, while the relations between size, width, and space in resolution are now fairly well-understood, one big open question remains. Namely, it was shown in [BW01] that if a formula has a short resolution refutation then it can also be refuted in small width, but this narrow refutation is obtained at the price of an exponential blow-up in size. Is this inherent, or is it just an artifact of the proof in [BW01]? That is, can size and width be optimized simultaneously in resolution, or are there formulas for which optimizing one of the measures must always cause a stiff penalty for the other? For size vs. space and space vs. width dramatic trade-offs are known [BBI12, Ben09, BN11], and these results extend also to PCR [BNT13], but it remains open whether there are similiar trade-offs between size and width in resolution or between size and degree in PCR.

Acknowledgments

The authors would like to thank Mladen Mikša and Marc Vinyals for interesting discussions related to the topics of this work.

Part of the work of the first author was done while visiting KTH Royal Institute of Technology. The second and third authors were funded by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013) / ERC grant agreement no. 279611. The third author was also supported by Swedish Research Council grants 621-2010-4797 and 621-2012-5645.

References

[ABRW02] Michael Alekhnovich, Eli Ben-Sasson, Alexander A. Razborov, and Avi Wigderson. Space complexity in propositional calculus. *SIAM Journal on Computing*, 31(4):1184–1211, 2002. Preliminary version appeared in *STOC '00*. 2, 7

[AD08] Albert Atserias and Víctor Dalmau. A combinatorial characterization of resolution width. *Journal of Computer and System Sciences*, 74(3):323–334, May 2008. Preliminary version appeared in *CCC '03*. 2, 12

[AFT11] Albert Atserias, Johannes Klaus Fichte, and Marc Thurley. Clause-learning algorithms with many restarts and bounded-width resolution. *Journal of Artificial Intelligence Research*, 40:353–373, January 2011. Preliminary version appeared in *SAT '09*. 4, 23

[ALN14] Albert Atserias, Massimo Lauria, and Jakob Nordström. Narrow proofs may be maximally long. In *Proceedings of the 29th Annual IEEE Conference on Computational Complexity (CCC '14)*, pages 286–297, June 2014. 1

[AMO13] Albert Atserias, Moritz Müller, and Sergi Oliva. Lower bounds for DNF-refutations of a relativized weak pigeonhole principle. In *Proc. 28th Annual IEEE Conference on Computational Complexity (CCC '13)*, pages 109–120, June 2013. 4, 5

[AR03] Michael Alekhnovich and Alexander A. Razborov. Lower bounds for polynomial calculus: Non-binomial case. *Proc. Steklov Institute of Mathematics*, 242:18–35, 2003. Available at `http://people.cs.uchicago.edu/~razborov/files/misha.pdf`. Preliminary version appeared in *FOCS '01*. 2

[BBH+12] Boaz Barak, Fernando G. S. L. Brandão, Aram Wettroth Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proc. 44th Annual ACM Symposium on Theory of Computing (STOC '12)*, pages 307–326, May 2012. 3

[BBI12] Paul Beame, Chris Beck, and Russell Impagliazzo. Time-space tradeoffs in resolution: Superpolynomial lower bounds for superlinear space. In *Proc. 44th Annual ACM Symposium on Theory of Computing (STOC '12)*, pages 213–232, May 2012. 2, 24

[BD09] Michael Brickenstein and Alexander Dreyer. PolyBoRi: A framework for Gröbner-basis computations with Boolean polynomials. *Journal of Symbolic Computation*, 44(9):1326–1345, September 2009. 2

[BDG+09] Michael Brickenstein, Alexander Dreyer, Gert-Martin Greuel, Markus Wedler, and Oliver Wienand. New developments in the theory of Gröbner bases and applications to formal verification. *Journal of Pure and Applied Algebra*, 213(8):1612–1635, August 2009. 2

[Ben09] Eli Ben-Sasson. Size space tradeoffs for resolution. *SIAM Journal on Computing*, 38(6):2511–2525, May 2009. Preliminary version appeared in *STOC '02*. 2, 24

[Ber12] Christoph Berkholz. On the complexity of finding narrow proofs. In *Proc. 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS '12)*, pages 351–360, October 2012. 4

[BG03] Eli Ben-Sasson and Nicola Galesi. Space complexity of random formulae in resolution. *Random Structures and Algorithms*, 23(1):92–109, August 2003. Preliminary version appeared in *CCC '01*. 2

[BG13] Ilario Bonacina and Nicola Galesi. Pseudo-partitions, transversality and locality: A combinatorial characterization for the space measure in algebraic proof systems. In *Proc. 4th Innovations in Theoretical Computer Science Conference (ITCS '13)*, January 2013. 2

[BGL13] Olaf Beyersdorff, Nicola Galesi, and Massimo Lauria. Parameterized complexity of DPLL search procedures. *ACM Transactions on Computational Logic*, 14(3):20, August 2013. Preliminary version appeared in *SAT '11*. 23

[BGLR12] Olaf Beyersdorff, Nicola Galesi, Massimo Lauria, and Alexander A. Razborov. Parameterized bounded-depth Frege is not optimal. *ACM Transactions on Computation Theory*, 4:7:1–7:16, September 2012. Preliminary version appeared in *ICALP '11*. 23

[Bla37] Archie Blake. *Canonical Expressions in Boolean Algebra*. PhD thesis, University of Chicago, 1937. 1

[BN08] Eli Ben-Sasson and Jakob Nordström. Short proofs may be spacious: An optimal separation of space and length in resolution. In *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS '08)*, pages 709–718, October 2008. 2

[BN11] Eli Ben-Sasson and Jakob Nordström. Understanding space in proof complexity: Separations and trade-offs via substitutions. In *Proc. 2nd Symposium on Innovations in Computer Science (ICS '11)*, pages 401–416, January 2011. 2, 24

[BNT13] Chris Beck, Jakob Nordström, and Bangsheng Tang. Some trade-off results for polynomial calculus. In *Proc. 45th Annual ACM Symposium on Theory of Computing (STOC '13)*, pages 813–822, May 2013. 2, 24

[BPS07] Paul Beame, Toniann Pitassi, and Nathan Segerlind. Lower bounds for Lovász–Schrijver systems and beyond follow from multiparty communication complexity.

SIAM Journal on Computing, 37(3):845–869, 2007. Preliminary version appeared in *ICALP '05*. 3

[BS97] Roberto J. Bayardo Jr. and Robert Schrag. Using CSP look-back techniques to solve real-world SAT instances. In *Proc. 14th National Conference on Artificial Intelligence (AAAI '97)*, pages 203–208, July 1997. 1

[BW01] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. *Journal of the ACM*, 48(2):149–169, March 2001. Preliminary version appeared in *STOC '99*. 2, 4, 24

[CCT87] William Cook, Collette Rene Coullard, and Gyorgy Turán. On the complexity of cutting-plane proofs. *Discrete Applied Mathematics*, 18(1):25–38, November 1987. 3

[CEI96] Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proc. 28th Annual ACM Symposium on Theory of Computing (STOC '96)*, pages 174–183, May 1996. 2, 7

[Chv73] Vašek Chvátal. Edmond polytopes and a hierarchy of combinatorial problems. *Discrete Mathematics*, 4(1):305–337, 1973. 3

[CR79] Stephen A. Cook and Robert Reckhow. The relative efficiency of propositional proof systems. *Journal of Symbolic Logic*, 44(1):36–50, March 1979. 1

[CS88] Vašek Chvátal and Endre Szemerédi. Many hard examples for resolution. *Journal of the ACM*, 35(4):759–768, October 1988. 2

[CT12] Eden Chlamtáč and Madhur Tulsiani. Convex relaxations and integrality gaps. In Miguel F. Anjos and Jean B. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*, pages 139–169. Springer, 2012. 3

[DMR09] Stefan S. Dantchev, Barnaby Martin, and Martin Rhodes. Tight rank lower bounds for the Sherali-Adams proof system. *Theoretical Computer Science*, 410(21–23):2054–2063, May 2009. 14, 19

[ET01] Juan Luis Esteban and Jacobo Torán. Space bounds for resolution. *Information and Computation*, 171(1):84–97, 2001. Preliminary versions of these results appeared in *STACS '99* and *CSL '99*. 2

[FLM+13] Yuval Filmus, Massimo Lauria, Mladen Mikša, Jakob Nordström, and Marc Vinyals. Towards an understanding of polynomial calculus: New separations and lower bounds (extended abstract). In *Proc. 40th International Colloquium on Automata, Languages and Programming (ICALP '13)*, volume 7965 of *Lecture Notes in Computer Science*, pages 437–448. Springer, July 2013. 2

[FLN+12] Yuval Filmus, Massimo Lauria, Jakob Nordström, Neil Thapen, and Noga Ron-Zewi. Space complexity in polynomial calculus. In *Proc. 27th Annual IEEE Conference on Computational Complexity (CCC '12)*, pages 334–344, June 2012. 2

[FSS84] Merrick Furst, James B Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984. 5

[Gal77] Zvi Galil. On resolution with clauses of bounded size. *SIAM Journal on Computing*, 6(3):444–459, 1977. 2

[GHP02] Dima Grigoriev, Edward A. Hirsch, and Dmitrii V. Pasechnik. Complexity of semi-algebraic proofs. *Moscow Mathematical Journal*, 2(4):647–679, 2002. 3, 21

[Gom63] Ralph E. Gomory. An algorithm for integer solutions of linear programs. In R.L. Graves and P. Wolfe, editors, *Recent Advances in Mathematical Programming*, pages 269–302. McGraw-Hill, New York, 1963. 3

[GP14] Mika Göös and Toniann Pitassi. Communication lower bounds via critical block sensitivity. In *Proc. 46th Annual ACM Symposium on Theory of Computing (STOC '14)*, pages 847–856, May 2014. 3

[Gri01] Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1–2):613–622, May 2001. 3

[GV01] Dima Grigoriev and Nicolai Vorobjov. Complexity of null- and positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1–3):153–160, December 2001. 3

[Hak85] Armin Haken. The intractability of resolution. *Theoretical Computer Science*, 39(2-3):297–308, August 1985. 2

[Hås87] Johan Håstad. *Computational Limitations of Small-depth Circuits*. PhD thesis, Massachussetts Institute of Technology, 1987. 5

[IPS99] Russell Impagliazzo, Pavel Pudlák, and Jiri Sgall. Lower bounds for the polynomial calculus and the Gröbner basis algorithm. *Computational Complexity*, 8(2):127–144, 1999. 2, 13

[Las01] Jean B. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In *Proc. 8th International Conference on Integer Programming and Combinatorial Optimization*, volume 2081 of *Lecture Notes in Computer Science*, pages 293–303. Springer, June 2001. 3

[Lau01] Monique Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. *Mathematics of Operations Research*, 28:470–496, 2001. 3

[LS91] László Lovász and Alexander Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991. 3, 21

[MMZ+01] Matthew W. Moskewicz, Conor F. Madigan, Ying Zhao, Lintao Zhang, and Sharad Malik. Chaff: Engineering an efficient SAT solver. In *Proc. 38th Design Automation Conference (DAC '01)*, pages 530–535, June 2001. 1

[MS99] João P. Marques-Silva and Karem A. Sakallah. GRASP: A search algorithm for propositional satisfiability. *IEEE Transactions on Computers*, 48(5):506–521, May 1999. Preliminary version appeared in *ICCAD '96*. 1

[OZ13] Ryan O'Donnell and Yuan Zhou. Approximability and proof complexity. In *Proc. 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '13)*, pages 1537–1556, January 2013. 3

[Par00] Pablo A. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, May 2000. 3

[Pud99] Pavel Pudlák. On the complexity of propositional calculus. In S. Barry Cooper and John K. Truss, editors, *Sets and Proofs*, volume 258 of *London Mathematical Society Lecture Note Series*, pages 197–218. Cambridge University Press, 1999. 3

[Pud00] Pavel Pudlák. Proofs as games. *American Mathematical Monthly*, pages 541–550, 2000. 12

[Raz98] Alexander A. Razborov. Lower bounds for the polynomial calculus. *Computational Complexity*, 7(4):291–324, December 1998. 13

[SA90] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3:411–430, 1990. 3

[Sch08] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS '08)*, pages 593–602, October 2008. 3

[Urq87] Alasdair Urquhart. Hard examples for resolution. *Journal of the ACM*, 34(1):209–219, January 1987. 2