Electronic Colloquium on Computational Complexity, Report No. 52 (2019)

Polynomial calculus space and resolution width

Nicola Galesi*[1], Leszek A. Kołodziejczyk†[2], and Neil Thapen‡[3]

[1]Dipartimento di Informatica, *Sapienza Università di Roma*
[2]Institute of Mathematics, *University of Warsaw*
[3]Institute of Mathematics, *Czech Academy of Sciences*

Abstract

We show that if a k-CNF requires width w to refute in resolution, then it requires space $\sqrt{w}$ to refute in polynomial calculus, where the *space* of a polynomial calculus refutation is the number of monomials that must be kept in memory when working through the proof. This is the first analogue, in polynomial calculus, of Atserias and Dalmau's result lower-bounding clause space in resolution by resolution width.

As a by-product of our new approach to space lower bounds we give a simple proof of Bonacina's recent result that total space in resolution (the total number of variable occurrences that must be kept in memory) is lower-bounded by the width squared. As corollaries of the main result we obtain some new lower bounds on the PCR space needed to refute specific formulas, as well as partial answers to some open problems about relations between space, size, and degree for polynomial calculus.

*nicola.galesi@uniroma1.it

†Partially supported by grant 2017/27/B/ST1/01951 of the National Science Centre, Poland. lak@mimuw.edu.pl

‡Partially supported by GA ČR project 19-05497S, by ERC advanced grant 339691 (FEALORA) and by RVO:67985840. thapen@math.cas.cz

ISSN 1433-8092

1 Introduction

Propositional proof complexity studies the complexity of finding efficiently verifiable proofs, that is, polynomial-time checkable certificates that propositional formulas are unsatisfiable. Research in this area started with the work of Cook and Reckhow [12] and was originally viewed as a gradual advance towards showing that NP≠co-NP. The main focus was on proving upper and lower bounds on proof size. The most well-studied proof system in proof complexity is resolution, for which numerous exponential size lower bounds have been shown. By a result of Ben-Sasson and Wigderson [5], to show that a CNF requires large size in resolution it is usually enough to show that it requires large *width*, where the width of a proof is the size of its largest clause.

Naturally other complexity measures for proofs have also been investigated, often revealing interesting connections. A recent line of research has looked at the *space* measure, motivated by an analogy between proofs and boolean circuits or Turing machines, and more recently by applied SAT solving, where efficient memory access and management is a major concern. The study of space in resolution was initiated by Esteban and Torán [13], who defined the space of a resolution proof as the maximal number of clauses to be kept simultaneously in memory during verification of the proof. This definition was later generalized to other proof systems by Alekhnovich et al. [1]. As proved in [13], a CNF formula over n variables can be refuted in space $n + 1$, even in resolution. Tight lower bounds for resolution proof space were proved in a series of works [13, 4, 1], and Atserias and Dalmau [3] established the general result that for resolution, width is a lower bound on space.

Together with resolution, the main focus of this paper is *polynomial calculus resolution* (PCR), an algebraic proof system extending resolution by the capacity to reason about polynomial equations. Polynomial calculus (PC) was introduced by Clegg et al. [11] and was later extended by Alekhnovich et al. [1] to the more general system PCR. On the surface, PC and PCR are systems for proving membership in ideals of multivariate polynomials. However, they can also be viewed as refutational proof systems for CNF formulas: clauses can be efficiently translated to multilinear monomials over some (fixed) field $\mathbb{F}$, and a CNF formula F is shown to be unsatisfiable by proving that the constant 1 is in the ideal generated by polynomials representing clauses of F together with polynomials enforcing that variables take only boolean values. In PC and PCR the main proof complexity measure studied is *degree*, the maximal degree of a polynomial used in the proof. A connection between degree and the *size* of a proof (that is, the number of monomials used), was proved for PC in [11, 19], which inspired the similar connection between width and size for resolution of [5]. This result made it possible to lift most of the known degree lower bounds for PCR to size bounds [23, 19, 2, 18, 17, 21].

The *space* of a PCR proof is the maximal number of distinct monomials that must be simultaneously in memory during a verification of the proof. The study of PCR space started in [1], and grew in importance due to the fact that PCR underlies SAT-solvers based on Gröbner algorithms. Already in [1] it was shown that PCR is strictly more powerful than resolution in terms of space, though the separation proved there is relatively modest and witnessed by rather artificial formulas. Eventually, research on limitations of proof space in PCR led to several lower bounds [1, 8, 16, 6, 14] and to a framework to prove them [8].

An important open problem raised several times (see [22, 8, 16, 6, 14, 15]) is to determine whether the elegant relation between width and space for resolution given in [3] has an analogy in a relation between PCR degree and space, or even between resolution width and PCR space. This is relevant to the more fundamental issue: how far-reaching is the analogy between proof complexity for resolution and for PCR, two systems that have several common features but are of different computational nature?

1.1 Contributions

We give the less-expected answer to this open problem, by showing a connection between PCR space and resolution width for CNF formulas, although one that is quadratic, rather than linear. Our main result is the following theorem.

Theorem 20. *Let F be a k-CNF. If F has a PCR refutation in space s over some field $\mathbb{F}$, then F has a resolution refutation of width $s^2 - s + k$.*

Since width w resolution can easily be simulated by degree $w+1$ PCR, this also shows that PCR refutations in space s can be transformed into PCR refutations of degree $O(s^2)$.

Theorem 20 can be understood as a general lower bound on PCR space: as long as k is small, if a k-CNF requires width w to refute in resolution, then it requires space $\sqrt{w}$ to refute in PCR. The proof is quite different from previous PCR space lower bounds, which adapt a combinatorial argument from [1], and we outline our new approach in Section 1.2. Using this we also get a very simple proof of Bonacina's recent result [7] that, in resolution, total space is lower bounded by width squared. Our proof of that result (Theorem 4) does not use any technical notion such as that of *asymmetric width* required in [7].

As is typical for PCR space lower bounds, our main theorem depends very little on the particular rules of PCR. It only uses that the rules are sound, and that at each step we either add terms to the memory or delete them (but not both at once). To study term space in a general setting we describe a class of *configurational* proof systems, in which we are only guaranteed soundness, and show that there we get the weaker bound of $2s(s+1)+k$ on resolution width (Theorem 19). This class is similar in spirit to, and includes, the semantic *functional calculus* system of [1].

As a consequence of Theorem 20, we partially answer some other open questions about the relation between space, size, and degree in PCR. A brief discussion of these follows.

New space lower bounds for PCR. The framework developed in [8] can be used to derive all space lower bounds for PCR known until now. However, as observed in [14], there are CNF formulas for which PCR space lower bounds appear likely to hold, but this framework seems not to work. These include the *linear ordering principle* and *functional pigeonhole principle* formulas, as well of versions of them with constant initial width. Using well-known width lower bounds for these formulas [9, 18, 25, 27, 21] and Theorem 20 we are now able to prove PCR space lower bounds.

Simplification and generalization of a previous lower bound. When G is a bounded-degree connected graph with n nodes and expansion $\Omega(n)$, the well-known *Tseitin formula* $\mathrm{Ts}(G)$ requires width $\Omega(n)$ to refute in resolution [28, 5] and hence, by [3], also $\Omega(n)$ space in resolution. In PCR, $\Omega(\sqrt{n})$ space lower bounds for $\mathrm{Ts}(G)$ for random graphs were obtained in [14] using the framework of [8]. As a consequence of Theorem 20 and the width lower bound we also obtain a $\Omega(\sqrt{n})$ lower bound for space in PCR, but using no assumptions on G other than the expansion.

Separations independent of characteristic. It is left open in [14] whether there are formulas separating PCR size and degree from space for all fields at once, independently of the characteristic. Our space lower bounds for linear ordering principles give such an example separating PCR size from space.

A further example is provided by a variant of the *bijective* (*both functional and onto*) pigeonhole principle. A result of Riis ([26, 24]) shows that bijective pigeonhole principle formulas for $n+1$ pigeons and n holes have small PCR refutations in constant degree independently of the characteristic. Riis' result holds for a version of the principle where translations of wide clauses are replaced by certain sums. We introduce a constant degree version of the bijective pigeonhole principle that requires $\Omega(n)$ width to refute in resolution, but can be used to derive Riis' principle by means of small PCR proofs of constant degree. Theorem 20 hence gives us a separation of PCR size and degree from space independently of characteristic.

1.2 Outline of technique

Proof space lower bounds typically have the form: work down through the refutation and inductively show that at each step, if the amount of the proof that is currently loaded into memory is small, then there is some small object (such as a partial assignment) which semantically implies every formula in memory. This gives a contradiction when we get to the bottom of the refutation, where there is an unsatisfiable formula.

Our new idea is to pass up and down through the refutation possibly several times, satisfying part of the memory as we go down, and dually falsifying part as we go up. Our model is an argument of Buss in bounded arithmetic, showing that mathematical induction for NP properties is enough to prove induction for boolean combinations of NP properties [10, Corollary 4]; a propositional version of this might say that small-width resolution can simulate arguments in which each step is a boolean combination of constantly many small-width CNFs. Buss' proof uses the Hausdorff difference hierarchy, which we do not use explicitly but which, in our setting, tells us that at each step the contents of the memory can potentially be written in an alternating fashion, with positive and negative subformulas appearing in a controllable way.

We first apply this idea to give a simplified proof of Bonacina's lower bound on total space in resolution in terms of resolution width [7]. A key tool is an Atserias-Dalmau family $\mathcal{H}$ of partial assignments for a formula F, which is guaranteed to exist if F requires large resolution width [3]. Given a refutation of F in small total space, we find the first step j at which some assignment $\alpha \in \mathcal{H}$ falsifies some narrow clause in memory; then we find the last step $i < j$ at which some $\beta \supseteq \alpha$ in $\mathcal{H}$ satisfies all wide clauses in memory; then we reach a contradiction by considering the proof in the interval $[i, j]$ under β.

For the PCR bound we will repeat this step of moving to a subinterval of the proof several times. We first define what it means for a partial assignment α to *force* a memory configuration M to be true or false over a family $\mathcal{H}$. Then suppose that F requires large resolution width w, but has a PCR refutation in small term space s. We list the memory configurations in the refutation as $M_0, \ldots, M_t$. We then inductively find a decreasing sequence of intervals $[0, t] = [i_0, j_0] \supseteq [i_1, j_1] \supseteq \ldots$ in the proof and an increasing sequence of partial assignments $\emptyset = \alpha_0 \subseteq \alpha_1 \subseteq \ldots$ in $\mathcal{H}$, such that α_m forces M_{i_m} to be true and M_{j_m} to be false, and at each step the proof restricted to the interval $[i_m, j_m]$ under the assignment α_m becomes simpler in a certain technical sense. At each step α_m grows by at most $O(s)$ literals, and the restricted proof becomes trivial after s steps, so we reach a contradiction as long as w is bigger than $O(s^2)$, giving our bound.

1.3 Organization

Section 2 contains some preliminary definitions. In Section 3 we discuss width and space in resolution, introduce the Atserias-Dalmau characterization of width and prove our simple lower bound on total space in resolution. In Section 4 we define our forcing relation and prove some properties of it. In Section 5 we prove a simple version of our main theorem, with a $2s(s+1)+k$ bound on width (Theorem 18). In Section 6 we extend this argument to give our main results, a $2s(s+1)+k$ bound for any configurational system (Theorem 19) and a $s^2 - s + k$ bound for PCR (Theorem 20). Section 7 describes some consequences of our results for the relations between space, size and degree. In Section 8 we mention some open problems.

2 Preliminary definitions

A *literal* is either a boolean variable x or its negation $\bar{x}$. Boolean variables will take 0/1 values, identified with $\bot/\top$. A *clause* is a set of literals, treated as a disjunction. The *width* of a clause is the number of literals in it. A clause of width at most k is called a k-clause. A CNF formula

is a conjunction of clauses. A k-CNF formula is a CNF formula consisting of k-clauses. A *term* is a set of literals, treated as a conjunction.

A *partial assignment* is a partial function from the set of boolean variables to $\{0,1\}$. For us *assignment* will always mean partial assignment unless we specify otherwise. When convenient, we will identify an assignment with the set of literals which it makes true. We write $\mathrm{dom}(\alpha)$ for the domain of an assignment α and write $|\alpha|$ for $|\mathrm{dom}(\alpha)|$.

Resolution is a refutational propositional system for CNF formulas based on the *resolution* rule, which allows us to derive the clause $C \vee D$ from the clauses $C \vee x$ and $D \vee \bar{x}$. A resolution refutation of a CNF F is a sequence of clauses $C_0 \ldots, C_m$ ending with the empty clause and such that each C_i is either a clause in F or is obtained from earlier clauses by resolution. The *size* of a resolution refutation is the number of clauses in it. The *width* of a resolution refutation is the maximum width of a clause in it.

Polynomial calculus (PC) is an algebraic proof system defined in [11], which can be used to witness that a set of polynomials has no solution. A PC proof works over a fixed field $\mathbb{F}$ and proof lines in it are polynomials in $\mathbb{F}[x_1, \ldots, x_n]$. We will not work with PC but instead with a refinement of it, *polynomial calculus with resolution* (PCR), introduced in [1]. In PCR, proof lines are polynomials in $\mathbb{F}[x_1, \ldots, x_n, \bar{x}_1, \ldots, \bar{x}_n]$, with a formal algebraic variable for every boolean literal, not just for every boolean variable. This has the advantage that a term, even with negative literals, can be written as a single monomial rather than as a sum of possibly exponentially many monomials, as would happen if we had to write $1 - x$ to express $\bar{x}$. We will always have the axiom $\bar{x} = 1 - x$ available and will treat $\bar{x}$ semantically as the negation of x. That is, in any assignment α, if either $\alpha(x)$ or $\alpha(\bar{x})$ is defined then both are and $\alpha(\bar{x}) = 1 - \alpha(x)$.

A *monomial* m over $\mathbb{F}$ is a product of literals together with a coefficient from $\mathbb{F}$. The *term represented by* m is the conjunction of the literals appearing in m. The degree of a literal in m will never matter in this paper, so it is safe to think of a monomial as a term with a coefficient in front of it. A *polynomial* is a formal sum of monomials.

A PCR *refutation* of a set of polynomials P is a sequence $p_0, \ldots, p_t$ of polynomials, ending with the constant polynomial 1, where we interpret a proof-line p_i as asserting that $p_i = 0$. Each p_i either comes from P or is obtained by one of the rules of PCR applied to earlier lines. The rules are

$$\textbf{boolean axioms: } \frac{}{x^2 - x} \qquad \textbf{complementarity axioms: } \frac{}{x + \bar{x} - 1}$$

$$\textbf{linear combination: } \frac{p \quad q}{ap + bq} \qquad \textbf{multiplication: } \frac{p}{xp}$$

where p, q are any polynomials, x is any literal, and $a, b \in \mathbb{F}$. The *size* of a PCR refutation is the total number of monomials appearing in it, and the *degree* of a refutation is the maximum degree of any monomial in it.

We can translate a clause $\bigvee_i y_i$ in literals y_i into the semantically equivalent polynomial equation $\prod_i \bar{y}_i = 0$. Thus an unsatisfiable CNF translates into a set of polynomials with no solutions over $\{0,1\}$, and it makes sense to view PCR as a refutational system for CNFs. There is then a simple, direct simulation of resolution by PCR, and we see that degree in PCR is an analogous measure to width in resolution.

2.1 Space measures

As is usual when studying space in a refutational system, we require a refutation of a CNF F to be written in a special form, as a sequence of *configurations* $M_0, \ldots, M_t$.

In resolution, a configuration is a set of clauses and a refutation $M_0, \ldots, M_t$ is such that the first configuration is empty, the last one contains the empty clause, and for each $i < t$, M_{i+1} is obtained from M_i by one of the rules (1) *axiom download*: a clause of F is downloaded into M_{i+1}, (2) *deletion*: M_{i+1} is obtained from M_i deleting one or more clauses, or (3) *inference*: M_{i+1} is obtained from M_i by adding the conclusion of the resolution rule applied to two clauses in M_i.

The *clause space*, or simply *space*, of such a refutation is the maximum number of clauses appearing in any M_i. The *total space* of a configuration M_i is the total number of variable instances appearing in M, or equivalently the sum of the width of the clauses in M_i. The *total space* of a refutation is the maximum total space of any M_i.

In PCR a configuration is a set of polynomials, and a configurational PCR refutation of a CNF F is a sequence $M_0, \ldots, M_t$ where M_0 is empty, M_t contains the polynomial 1, and for each $i < t$, M_{i+1} is obtained from M_i by the rules (1)-(3) above, adapted to PCR. So in (1) the axioms we can download are polynomials translating the clauses of F and instances of the boolean and complementarity axioms, and in (3) we can infer new polynomials by linear combination or multiplication. The standard space measure for a PCR configuration M_i is *monomial space*, which counts the number of distinct monomials appearing in it [1]. However we will work with the *term space*, which is the number of distinct terms represented by the monomials in M_i and which lower bounds the monomial space. We define the *term space*, or simply *space*, of a PCR refutation is the maximum term space of any M_i.

It is natural to think of a configuration as a formula, namely a CNF in the case of resolution or a conjunction of polynomial equations in the case of PCR, and to think of rules (1)-(3) as rules for deriving a new formula. To state our most general results, let us use this idea and define a *configurational* proof system to be specified by a class of formulas and a set of sound unary or binary rules. A refutation of a CNF F in the system is a sequence $M_0, \ldots, M_t$ of formulas from the class, called *configurations*. M_0 is the constant $\top$, M_t is the constant $\bot$, and each M_{i+1} is obtained from M_i, possibly together with some initial clause C of F, by a rule. The limitation that we cannot use any configuration appearing earlier than M_i to derive M_{i+1} is a strengthening of the standard "treelike" restriction on proof structure. Configurational resolution and PCR, as described above, are examples of such systems, if we understand $\top$ as the empty conjunction and $\bot$ as the empty clause or the equation $1 = 0$.

We can study the complexity of such a system by studying the complexity of its configurations. Suppose that each configuration is labelled with a set of terms and is semantically equivalent to a boolean function of those terms. Then we can define the *term space* of a configuration to be the number of terms labelling it, and the term space of a refutation to be the maximum term space of its configurations. This measure (which could just as well be called "clause space") lower-bounds both clause space for resolution and monomial space for PCR, if we understand them as configurational systems and label configurations with the clauses or terms that appear in them. Our argument gives a lower bound for term space in *any* configurational system, even the "semantic" one in which configurations can be any formula and all sound rules are allowed – this is essentially the same as the *functional calculus* system defined in [1]. We prove a better bound, by a factor of two, in the specific case of PCR.

3 Width, space, and total space in resolution

We will make heavy use of a characterization of resolution width given by Atserias and Dalmau [3]. There, the family $\mathcal{H}$ defined below is referred to as a winning strategy for the Duplicator in a certain kind of pebble game.

Definition 1 ([3])**.** Let F be a k-CNF. A *width-w Atserias-Dalmau family* for F is a nonempty family $\mathcal{H}$ of partial assignments to the variables of F such that for each $\alpha \in \mathcal{H}$,

(i) $|\alpha| \leq w$

(ii) if $\beta \subseteq \alpha$ then $\beta \in \mathcal{H}$

(iii) if $|\alpha| < w$ and x is a variable of F, then there is $\beta \supseteq \alpha$ in $\mathcal{H}$ with $x \in \mathrm{dom}(\beta)$

(iv) α does not falsify any clause of F.

Lemma 2 ([3]). *If F is a k-CNF with no resolution refutation of width w, then there exists a width-$(w+1)$ Atserias-Dalmau family for F.*

One way to prove Lemma 2 is by considering the *Prover-Adversary game* on F. This is played between an Adversary, who claims she knows a total assignment satisfying F, and a Prover, who maintains a partial assignment α (his memory) and can in each round ask the Adversary the value of a variable, with the goal of extending α to falsify some clause of F, or can forget variables from α to save memory. By replacing each clause with the partial assignment negating it, and flipping the direction of the edges in the graph of the refutation, we can identify small-width resolution refutations of F with winning strategies for the Prover which use limited memory. If there is no such strategy, we can construct $\mathcal{H}$ from the positions α which are winning for the Adversary.

Theorem 3 ([3]). *Let F be a k-CNF. If F has a resolution refutation in space s, then it has a resolution refutation in width $s+k$.*

Proof. Let $M_0, \ldots, M_t$ be the sequence of configurations forming the space-s refutation. Suppose there is no refutation of F in width $s+k$. Let $\mathcal{H}$ be a width-$(s+k+1)$ Atserias-Dalmau family for F. We will inductively show that for each i there is $\alpha \in \mathcal{H}$ which satisfies every clause in M_i. This is trivial for M_0 and a contradiction for M_t.

Suppose it is true for M_i. Since it takes only one literal to satisfy a clause, we may assume $|\alpha| \leq s$. The only interesting case is axiom download, where M_{i+1} is $M_i \wedge C$ for some initial clause C from F. By part (iii) of Definition 1 we can extend α in k steps to some $\beta \in \mathcal{H}$ which sets all variables in C. By part (iv), β must satisfy C, so we are done. □

Notice that in the Prover strategy corresponding to a small-width refutation in Lemma 2 starts at the bottom of the proof and works up, trying to falsify clauses. An alternative proof of Theorem 3 would be to construct a small-width refutation directly as a Prover strategy, where this time the Prover starts at the top of the configurational proof and works down, trying to satisfy clauses. In the next theorem we combine both kinds of strategy, first going up and then down. We can think of the theorem as a lower bound on a space measure in which narrow clauses do not count towards the space of a configuration.

Theorem 4. *Let F be a k-CNF. Let $m, s \in \mathbb{N}$ with $m \geq k$. Suppose that F has a configurational resolution refutation in which each configuration contains at most s clauses of width greater than m. Then F has a resolution refutation of width $2m+s$.*

Proof. Let $M_0, \ldots, M_t$ be the configurational resolution refutation. Each M_i contains some number q of *narrow* clauses $C_1, \ldots, C_q$ of width at most m, and $r \leq s$ many *wide* clauses $D_1, \ldots, D_r$ of width greater than m. Suppose for a contradiction that F has no resolution refutation of width $2m+s$. Let $\mathcal{H}$ be a Atserias-Dalmau family for F of width $2m+s+1$.

The configuration M_t contains the empty clause, which is narrow and falsified by any assignment. Let j be least such that some narrow clause C in M_j is falsified by some assignment $\alpha \in \mathcal{H}$. Fix such a C and α. Without loss of generality, $|\alpha| \leq m$. Since C is falsified by α, it cannot have been introduced by axiom download. So we must have $C = E \vee F$ for clauses $E \vee x$ and $F \vee \bar{x}$ in M_{j-1}. Extend α to $\alpha' \in \mathcal{H}$ which gives a value to x, with $|\alpha'| \leq m+1$. Without loss of generality $\alpha'(x) = 1$. Hence α' falsifies $F \vee \bar{x}$, and by minimality of j, we know that $F \vee \bar{x}$ is a wide clause.

Now let $i < j$ be greatest such that there is some $\beta \supseteq \alpha'$ in $\mathcal{H}$ which satisfies every wide clause in M_i. Fix such a β. Without loss of generality, $|\beta| \leq |\alpha'| + s \leq m+s+1$. Since α' falsifies $F \vee \bar{x}$, we cannot have $i = j-1$. Therefore maximality of i implies that M_{i+1} extends M_i by adding a wide clause D which is not satisfied by any $\gamma \supseteq \beta$ in $\mathcal{H}$. Axioms are narrow, so D cannot be an axiom. Thus have $D = A \vee B$ for two clauses $A \vee y$ and $B \vee \bar{y}$ in M_i. Extend β to $\beta' \in \mathcal{H}$ which gives a value to y, with $|\beta'| \leq m+s+2$. Without loss of generality $\beta'(y) = 1$,

and we look at the clause $B \vee \bar{y}$. If this clause is wide, then β satisfies it, which means that β satisfies B and hence D, which is impossible. If it is narrow, then we can extend β' to $\gamma \in \mathcal{H}$ such that $|\gamma| \leq |\beta'| + m - 1 \leq 2m + s + 1$ and γ sets all variables in $B \vee \bar{y}$. The minimality of j implies that γ satisfies $B \vee \bar{y}$. We know that $\gamma(y) = 1$, so γ satisfies B and thus D, which is impossible. □

Theorem 4 has the following consequence, which is essentially the main result of [7] with an improved constant.

Corollary 5. *Let F be a k-CNF and $w \geq k$. Suppose F has no resolution refutation in width w. Then it has no resolution refutation in total space $w^2/8$.*

Proof. Suppose that there is a refutation Π in total space $w^2/8$. Then, if we set $m = w/4$ and $s = w/2$, no configuration in Π can contain more than s many clauses of width more than m. Hence we can apply the lemma to find a resolution refutation of width $2m + s = w$. □

4 Forcing with an Atserias-Dalmau family

In this section, we explain how to use the structure of an Atserias-Dalmau family $\mathcal{H}$ to define the relation "α forces the term t to a certain value". This is in fact a very simple version of a forcing relation as used in set theory and other areas of logic. For a recent application of a similar definition in proof complexity, see e.g. [20]. The idea is that no extension of α will ever give t a different value, as long as we only consider extensions within $\mathcal{H}$. We will use this in the next section to prove our main result. We present the constructions and proofs for PCR, but will explain in Section 6 how they can be generalized to an arbitrary configurational proof system.

Fix a k-CNF F and a width-w Atserias-Dalmau family $\mathcal{H}$ for F, for some $k, w \in \mathbb{N}$.

Definition 6. For an assignment $\alpha \in \mathcal{H}$ and a term t, we define

(i) α *forces* $t = 0$ if α sets some literal in t to 0

(ii) α *forces* $t = 1$ if no $\beta \in \mathcal{H}$ with $\beta \supseteq \alpha$ sets any literal in t to 0.

If either holds, we say that α *fixes* t.

We write these as $\alpha \Vdash t = 0$ and $\alpha \Vdash t = 1$. We now extend the definition to polynomials and configurations. We will treat polynomials as linear combinations of terms over our field $\mathbb{F}$.

Definition 7. For an assignment $\alpha \in \mathcal{H}$ and a polynomial $p = \sum_i a_i t_i$, we say that α *decides* p if it fixes every term t_i in p. We say that α *decides* a configuration M if it fixes every term in M or, equivalently, decides every polynomial in M.

If α decides p then, for each term t_i in p, there is a 0/1 value b_i such that $\alpha \Vdash t_i = b_i$; implicitly, α assigns value b_i to t_i. We say that α *forces* $p = 0$ if p, considered as a linear combination of terms, evaluates to 0 under this assignment. More formally, α forces $p = 0$ if α decides p and $\sum_i a_i b_i = 0$. We say that α forces $p \neq 0$ if α decides p and $\sum_i a_i b_i \neq 0$.

For a configuration M, we say that α *forces* M if α decides M and forces $p = 0$ for every polynomial p in M. We say that α forces $\neg M$ if α decides M and forces $p \neq 0$ for some p in M.

We write these relations as $\alpha \Vdash p = 0$, $\alpha \Vdash p \neq 0$ etc. Note that they are all preserved under extending α within the family $\mathcal{H}$.

The intuitive meaning of $\alpha \Vdash p = 0$ is that, if we consider only assignments in $\mathcal{H}$, then the equation $p = 0$ "holds" in every extension of α, and this is extended to negations and configurations in the natural way. Notice that whether a *term* is forced to some value depends on the structure of $\mathcal{H}$ in a potentially nontrivial way, but for polynomials and configurations, nothing new happens. This is because our application is to prove lower bounds on term space.

In this context terms can be very big, and the concept of forcing allows us to set their value without setting many variables. On the other hand, polynomials and configurations contain few terms, so they can be decided simply by fixing those few terms.

In the following lemmas we show that the $\Vdash$ relation usually behaves in an intuitive way, after first giving an example of how this can break down when α is very large.

Example. Assume that $\alpha \in \mathcal{H}$, $|\alpha| = w$, and that $x \notin \mathrm{dom}(\alpha)$. Then, since α has no proper extensions in $\mathcal{H}$, we have both $\alpha \Vdash x = 1$ and $\alpha \Vdash \bar{x} = 1$.

Lemma 8. *Let $\alpha \in \mathcal{H}$ and M be a configuration. We cannot have both $\alpha \Vdash M$ and $\alpha \Vdash \neg M$.*

Proof. This is immediate from the definitions. □

Lemma 9. *Let $\alpha \in \mathcal{H}$ and let $t_1, \ldots, t_s$ be terms. Then there is $\beta \supseteq \alpha$ in $\mathcal{H}$ such that β fixes $t_1, \ldots, t_s$ and $|\beta| \leq |\alpha| + s$.*

Proof. It is enough to show this for $s = 1$. If there is some $\gamma \supseteq \alpha$ in $\mathcal{H}$ which sets a literal x in t_1 to 0, we put $\beta = \alpha \cup \{x := 0\}$ so that $\beta \Vdash t_1 = 0$. We have $\beta \in \mathcal{H}$, since $\beta \subseteq \gamma$. If there is no such γ then by definition $\alpha \Vdash t_1 = 1$ and we put $\beta = \alpha$. □

Lemma 10. *Let $\alpha \in \mathcal{H}$ with $|\alpha| < w$. Let $t_1, \ldots, t_r$ be terms and $b_1, \ldots, b_r$ be boolean values such that $\alpha \Vdash t_i = b_i$ for each i. Then α can be extended to a total assignment A such that $A(t_i) = b_i$ for each i.*

Proof. To construct A, start with α and then, for each t_i forced to 1 by α, set all literals in t_i to 1. Set all remaining variables arbitrarily. The only way this construction can fail is if some variable x appears positively in a term t_i and negatively in a term t_j, where α forces both t_i and t_j to 1. But this cannot happen, since $|\alpha| < w$ implies that α has an extension in $\mathcal{H}$ setting either x or $\bar{x}$ to 0. □

Corollary 11. *Assume $k \geq 2$ and let $\alpha \in \mathcal{H}$ with $\alpha \leq w - k$. Let M and M' be successive configurations in a PCR refutation of F. Then it cannot be the case that $\alpha \Vdash M$ and $\alpha \Vdash \neg M'$.*

Proof. The configuration M' is semantically implied by M or by $M \wedge C$ for some clause C of F. We may assume that we are in the latter case. Let $\alpha \Vdash M$ and $\alpha \Vdash \neg M'$.

We first extend α in k steps to $\beta \in \mathcal{H}$ which sets all variables in C. By part (iv) of Definition 1, β satisfies a literal in C. We let $\alpha' \in \mathcal{H}$ be α plus this literal. Notice that $|\alpha'| < w$. List all terms in M and M' as $t_1, \ldots, t_r$. Since α' fixes all these terms, there exist boolean values $b_1, \ldots, b_r$ such that $\alpha' \Vdash t_j = b_j$ for each j. We use Lemma 10 to obtain a total assignment A extending α' which sets each t_j to b_j. Then A satisfies M since $\alpha' \Vdash M$ and falsifies M' since $\alpha' \Vdash \neg M'$. Also A satisfies C by construction of α'. This contradicts that $M \wedge C$ semantically implies M'. □

Corollary 12. *Assume $k \geq 2$ and let M and M' be successive configurations in a PCR refutation of F with term space s. Let $\alpha \in \mathcal{H}$ with $|\alpha| \leq w - k - s$. If $\alpha \Vdash M$, then there is $\beta \supseteq \alpha$ in $\mathcal{H}$ with $|\beta| \leq |\alpha| + s$ such that $\beta \Vdash M'$.*

Proof. This is immediate from Lemma 9 and Corollary 11. □

This suggests a possible approach to proving PCR space lower bounds. Given a refutation $M_0, \ldots, M_t$ with space s, use Corollary 12 to inductively find $\alpha_0, \ldots, \alpha_t$ in $\mathcal{H}$ such that $\alpha_i \Vdash M_i$, reaching a contradiction at M_t. However this does not work, since α_i may grow in size by s at each step, quickly reaching our limit $w - k$.

What is missing is a lemma saying that if $\alpha_i \Vdash M_i$, then we can find $\beta \subseteq \alpha_i$ such that $\beta \Vdash M_i$ and $|\beta|$ is bounded by a function of the space of M_i. This is called a *locality lemma* in the literature on space [1, 4, 8]. We do not expect a general lemma of this form to hold here,

because, for example, it is easy to envisage a large assignment α and a term t such that $\alpha \Vdash t = 1$ but this is not preserved in any smaller $\beta \subseteq \alpha$. Lemma 16 below is a kind of locality lemma, but has the limitation that it only controls the size of extensions of some fixed assignment α (α itself does not get smaller). We only apply it $O(s)$ times, and use it to control how fast our assignment grows.

5 Proof of main result

This section is devoted to a proof of an initial, somewhat simpler, version of our main result. We will adapt this proof to improve the bound on width from $2s(s+1)+k$ to $s(s-1)+k$ in the next section. We assume that F is a k-CNF (without loss of generality, $k \geq 2$) with a PCR refutation in space s over some fixed field $\mathbb{F}$. Let $M_0, \ldots, M_r$ be the sequence of configurations forming the refutation of F. For $0 \leq i \leq j \leq r$, the *proof interval* $[i, j]$ is the sequence of configurations $M_i, \ldots, M_j$.

We let $\mathcal{H}$ be a width-w Atserias-Dalmau family for F, with the value of w to be fixed later, and use the notion of forcing over $\mathcal{H}$ from the previous section. We will be interested in how many terms in a given configuration M are forced to 0 by an assignment from $\mathcal{H}$, or more precisely, in how many terms are not forced to 0. Given M and α, we write $\mathrm{Z}(M, \alpha)$ for the set of terms in M which are forced to 0 by α, and we write $\mathrm{NZ}(M, \alpha)$ for the remaining terms.

Definition 13. Let $m \geq 0$. An assignment $\alpha \in \mathcal{H}$ *guarantees m non-zeroes in M* if for all $\beta \supseteq \alpha$ in $\mathcal{H}$, we have $|\mathrm{NZ}(M, \beta)| \geq m$. We say that α *guarantees m non-zeroes in the proof interval* $[i, j]$ if for each $\ell \in [i, j]$, α guarantees m non-zeroes in M_ℓ.

Clearly the property of guaranteeing m non-zeroes is preserved under extending assignments within the family $\mathcal{H}$. The next lemma is a useful interaction of this property with forcing.

Lemma 14. *Suppose that $|\mathrm{NZ}(M, \alpha)| = m$ and that α guarantees m non-zeroes in M. Then α decides M.*

Proof. List $\mathrm{NZ}(M, \alpha)$ as $t_1, \ldots, t_m$. The remaining terms in M are forced to 0 by α, meaning that they each contain a literal set to 0 by α. Therefore, since α guarantees m non-zeroes in M, no $\beta \supseteq \alpha$ in $\mathcal{H}$ can force any t_i to 0, and so by definition α forces each t_i to 1. It follows that α fixes each term in M and thus decides M. □

We now prove two simple lemmas, allowing us to grow and shrink assignments, and then use these in the main lemma from which the space lower bound will follow.

Lemma 15. *Let M contain at most s terms and let $\alpha \in \mathcal{H}$ guarantee m non-zeroes in M. Then there is $\beta \supseteq \alpha$ in $\mathcal{H}$ such that β decides M and $|\beta| \leq |\alpha| + s - m$.*

Proof. This is a simple extension of the proof of Lemma 9. □

Lemma 16. *Let M contain at most s terms and let $\alpha \in \mathcal{H}$. Suppose there is $\gamma \supseteq \alpha$ in $\mathcal{H}$ with $|\mathrm{NZ}(M, \gamma)| = m$. Then there is β with $\alpha \subseteq \beta \subseteq \gamma$ such that $|\mathrm{NZ}(M, \beta)| = m$ and $|\beta| \leq |\alpha| + s - m$.*

Suppose furthermore that α guarantees m non-zeroes in M. Then $\gamma \Vdash M$ implies $\beta \Vdash M$ and $\gamma \Vdash \neg M$ implies $\beta \Vdash \neg M$.

Proof. List the terms in M as $t_1, \ldots, t_r$ with $r \leq s$. Suppose $\mathrm{NZ}(M, \gamma)$ is $t_1, \ldots, t_m$ and $\mathrm{Z}(M, \gamma)$ is $t_{m+1}, \ldots, t_r$. We define β by starting with α and adding, for each term t_i among $t_{m+1}, \ldots, t_r$, one literal from γ which sets t_i to 0. Then $|\mathrm{NZ}(M, \beta)| = |\mathrm{NZ}(M, \gamma)| = m$ and $|\beta| \leq |\alpha| + s - m$. In the "furthermore" part, β decides M by Lemma 14. The implications follow since $\beta \subseteq \gamma$. □

Lemma 17 (Main Lemma)**.** *Suppose $w \geq 2s(s+1)+k$. Then for each $m \leq s$ there is $\alpha \in \mathcal{H}$ and a proof interval $[i,j]$ such that*

(i) $\alpha \Vdash M_i$ *and* $\alpha \Vdash \neg M_j$

(ii) α *guarantees* m *non-zeroes in* $[i,j]$

(iii) $|\alpha| \leq 4 \sum_{r=0}^{m-1} (s-r)$.

Proof. We use induction on m. The base case for $m = 0$ is immediate, taking $\alpha = \emptyset$ and $[i,j]$ to be the whole refutation $[0,t]$. As M_0 has no terms and the last configuration M_t only contains the polynomial 1, the empty assignment $\emptyset$ forces M_0 and $\neg M_t$ and the other two conditions are trivial.

Now suppose that α and $[i,j]$ are such that conditions (i)–(iii) hold for m, where $m < s$. We will find a proof interval $[i',j'] \subseteq [i,j]$ and an assignment α'' satisfying (i)–(iii) for $m+1$. Note that (iii) implies $|\alpha| + 4(s-m) \leq 4[s + (s-1) + \cdots + 1] = 2s(s+1) \leq w - k$.

We work separately on the two ends of the proof interval. We first deal with the left end, distinguishing two cases:

(a) there is $\ell \in [i,j]$ such that for some $\beta \supseteq \alpha$ in $\mathcal{H}$ it holds that $|\mathrm{NZ}(M_\ell, \beta)| = m$ and $\beta \Vdash M_\ell$

(b) no such ℓ exists.

In case (a) we consider the largest such ℓ and a corresponding β; necessarily $\ell < j$. By condition (ii) and Lemma 16, we may assume without loss of generality that $|\beta| \leq |\alpha| + s - m$. By condition (ii) and Lemma 15, we may extend β to $\alpha' \in \mathcal{H}$ with $|\alpha'| \leq |\alpha| + 2(s-m)$ such that α' decides $M_{\ell+1}$. Since $\beta \Vdash M_\ell$, by Corollary 11, the soundness of PCR and the bound on $|\alpha'|$, it follows that $\alpha' \Vdash M_{\ell+1}$. We set $i' := \ell + 1$. In case (b) we set $\alpha' := \alpha$ and $i' := i$. In both cases, we have $|\alpha'| \leq |\alpha| + 2(s-m)$ and $\alpha' \Vdash M_{i'}$.

We now move to the right end of the interval and again distinguish two cases:

(c) there is $\ell \in [i',j]$ such that for some $\beta \supseteq \alpha'$ in $\mathcal{H}$ it holds that $|\mathrm{NZ}(M_\ell, \beta)| = m$

(d) no such ℓ exists.

In case (c) we consider the smallest such ℓ and a corresponding β. By Lemma 16 we may assume $|\beta| \leq |\alpha'| + s - m$. By condition (ii) and Lemma 14, β decides M_ℓ. Therefore $\beta \Vdash \neg M_\ell$, since if $\beta \Vdash M_\ell$ then ℓ and β satisfy the conditions of case (a), which is impossible by the choice of i'. It follows that $\ell > i'$. Using Lemma 15, we extend β to $\alpha'' \in \mathcal{H}$ with $|\alpha''| \leq |\alpha'| + 2(s-m) \leq w - k$ such that α'' decides $M_{\ell-1}$. We cannot have $\alpha'' \Vdash M_{\ell-1}$, by Corollary 11. Therefore $\alpha'' \Vdash \neg M_{\ell-1}$ and we set $j' := \ell - 1$. In case (d) we set $\alpha'' := \alpha'$ and $j' := j$. In both cases, $|\alpha''| \leq |\alpha| + 4(s-m)$ and $\alpha'' \Vdash \neg M_{j'}$.

This completes the construction. We have shown condition (i), and condition (iii) holds inductively. Finally, by condition (ii) for m we know that α'' guarantees m non-zeroes in $[i',j']$, since $\alpha'' \supseteq \alpha$. Furthermore, by the choice of j' we know that if $\gamma \supseteq \alpha''$ and $i' \leq \ell \leq j'$, then $|\mathrm{NZ}(M_\ell, \gamma)| \neq m$. Thus α'' in fact guarantees $m+1$ non-zeroes in $[i',j']$. □

Theorem 18. *Let F be a k-CNF. If F has a PCR refutation in monomial space s over some field $\mathbb{F}$, then F has a resolution refutation of width $2s(s+1)+k$.*

Proof. Suppose there is no such resolution refutation. Then we can choose our family $\mathcal{H}$ to have width $w = 2s(s+1)+k$, and it is enough to show that Lemma 17 leads to a contradiction for $m = s$. The lemma gives us a proof interval $[i,j]$ and $\alpha \in \mathcal{H}$ with $|\alpha| \leq w - k$ such that $\alpha \Vdash M_i$, $\alpha \Vdash \neg M_j$ and α guarantees s non-zeroes in $[i,j]$. For each $\ell \in [i,j]$ a (trivial) application of Lemma 15 shows that α decides M_ℓ. Using the fact that $\alpha \Vdash M_i$ and applying Corollary 11 to $M_{i+1}, \ldots, M_j$ in turn, we conclude that $\alpha \Vdash M_j$. But this is impossible. □

6 Improved bounds

In this section, we present two refined versions of our main result. First, we show that the bound from Theorem 18 works with respect to term space in any configurational proof system, not just PCR. Then, we improve the bound for PCR by roughly a factor of two.

6.1 A bound for general configurational systems

Recall from Section 2.1 that in general a configuration M with term space s is a formula φ labelled with a sequence of terms $t_1, \ldots, t_s$, such that φ is semantically equivalent to $g(t_1, \ldots, t_s)$ where g is a boolean function. Given $\alpha \in \mathcal{H}$, we say that α *decides* M if it fixes all terms, say to values $b_1, \ldots, b_s$ We say that α *forces* M or *forces* $\neg M$ if $g(b_1, \ldots, b_s)$ is respectively 1 or 0.

Using these definitions, all the arguments about PCR in Sections 4 and 5 go through for any configurational system, as we did not use any properties of PCR except for soundness of the rules. Therefore we have:

Theorem 19. *Let F be a k-CNF. If F has a refutation in term space s in any configurational proof system, then F has a resolution refutation of width $2s(s+1)+k$.*

We remark that, by counting terms more carefully, this can be improved to width $2s^2 + k$.

6.2 A stronger bound for PCR

We now show how to improve this bound in the case of PCR. The only specific property of PCR we use is that if M_ℓ and $M_{\ell+1}$ are successive configurations in a PCR refutation, then the terms in $M_{\ell+1}$ are either a subset or a superset of the terms in M_ℓ.

Theorem 20. *Let F be a k-CNF. If F has a PCR refutation in monomial space s over some field $\mathbb{F}$, then F has a resolution refutation of width $s^2 - s + k$.*

We use the following strengthening of Lemma 17.

Lemma 21. *Suppose $w \geq s(s-1)+k$. Then for each $m \leq s-1$ there is $\alpha \in \mathcal{H}$ and a proof interval $[i,j]$ in the PCR refutation such that*

(i) $\alpha \Vdash M_i$ *and* $\alpha \Vdash \neg M_j$

(ii) α *guarantees m non-zeroes in* $[i,j]$

(iii) $|\alpha| \leq 2\sum_{r=0}^{m-1}(s-1-r)$.

Proof. We use the same structure as the proof of Lemma 17, but with induction only up to $m = s-1$. In the induction, suppose we are in case (a) with $m < s-1$. We have $\ell \in [i,j]$ such that for some $\beta \supseteq \alpha$ in $\mathcal{H}$ it holds that $|\mathrm{NZ}(M_\ell, \beta)| = m$ and $\beta \Vdash M_\ell$, and we have chosen ℓ maximal, so that there is no such β for $M_{\ell+1}$. Furthermore α guarantees m non-zeroes at M_ℓ and $M_{\ell+1}$ and we have the bound $|\alpha| + 2(s-m-1) \leq 2\sum_{r=0}^{m}(s-1-r) \leq s(s-1) \leq w-k$. In Lemma 17, we used β to find $\alpha' \supseteq \alpha$ in $\mathcal{H}$ with $\alpha' \Vdash M_{\ell+1}$ and $|\alpha'| \leq |\alpha| + 2(s-m)$. We now want to improve this bound to $|\alpha'| \leq |\alpha| + s - m - 1$.

By the properties of PCR, we may list the terms in M_ℓ as $t_1, \ldots, t_p$ and the terms in $M_{\ell+1}$ as $t_1, \ldots, t_q$ with $p, q \leq s$. By Lemma 16 we may assume $|\beta| \leq |\alpha| + p - m$. If $q \leq p$, then already β decides $M_{\ell+1}$, so $\beta \Vdash M_{\ell+1}$ by Corollary 11; and also $|\mathrm{NZ}(M_{\ell+1}, \beta)| \leq |\mathrm{NZ}(M_\ell, \beta)| = m$. This contradicts the maximality of ℓ. So we must have $q > p$.

We apply the proof of Lemma 9 carefully to extend β to $\alpha' \in \mathcal{H}$ which fixes the remaining terms $t_{p+1}, \ldots, t_q$ in $M_{\ell+1}$. That is, for each of these terms t_i we add, if we can, a literal which sets t_i to 0, and otherwise do nothing. The resulting α' has size at most $|\alpha| + p - m + (q-p) \leq$

$w-k$, and thus $\alpha' \Vdash M_{\ell+1}$ by Corollary 11. Hence α' cannot set all of these terms to 0, or we would have $|\mathrm{NZ}(M_{\ell+1}, \alpha')| = |\mathrm{NZ}(M_\ell, \beta)| = m$, contradicting the maximality of ℓ. Therefore for at least one t_i we did not add a literal, giving $|\alpha'| \leq |\alpha| + p - m + (q - p - 1) \leq |\alpha| + s - m - 1$.

Now suppose we are in case (c) at the right end of the interval. We have $\ell \in [i', j]$ such that for some $\beta \supseteq \alpha'$ in $\mathcal{H}$ it holds that $|\mathrm{NZ}(M_\ell, \beta)| = m$ and we have chosen ℓ minimal, so that there is no such β for $M_{\ell-1}$. Again α' guarantees m non-zeroes at M_ℓ and $M_{\ell-1}$ and now we have the bound $|\alpha'| + s - m - 1 \leq w - k$. As in the proof of Lemma 17, we must have that $\beta \Vdash \neg M_\ell$ and $\ell > i'$. We list the terms in $M_{\ell-1}$ as $t_1, \ldots, t_p$ and the terms in M_ℓ as $t_1, \ldots, t_q$, and by Lemma 16 without loss of generality may assume $|\beta| \leq |\alpha'| + q - m$.

Similarly to before, we must have $p > q$ as $p \leq q$ implies $|\mathrm{NZ}(M_{\ell-1}, \beta)| \leq |\mathrm{NZ}(M_\ell, \beta)|$, contradicting the minimality of ℓ. By adding at most one literal for each term $t_{q+1}, \ldots, t_p$ we extend β to α'' which fixes all these terms; again this cannot set all of them to 0 or it would contradict the minimality of ℓ, so we have $|\alpha''| \leq |\beta| + p - q - 1 \leq |\alpha'| + p - m - 1 \leq w - k$. Hence $\alpha'' \Vdash \neg M_{\ell-1}$ by Corollary 11, since $\beta \Vdash \neg M_\ell$, and also $|\alpha''| \leq |\alpha| + 2(s - m - 1)$. □

Proof of Theorem 20. If there is no such resolution refutation, then F has an Asterias-Dalmau family $\mathcal{H}$ of width $w = s^2 - s + k + 1$, by Lemma 2. We apply Lemma 21 for $m = s - 1$. This gives us a proof interval $[i, j]$ and $\alpha \in \mathcal{H}$ with $|\alpha| \leq s(s-1) \leq w - k - 1$ such that $\alpha \Vdash M_i$, $\alpha \Vdash \neg M_j$ and α guarantees $s - 1$ non-zeroes in $[i, j]$. We will show inductively that for each ℓ in this interval there is $\beta \supseteq \alpha$ in $\mathcal{H}$ with $|\beta| \leq |\alpha| + 1$ such that $\beta \Vdash M_\ell$. This gives a contradiction for $\ell = j$.

Suppose this holds for ℓ. Necessarily every configuration in $[i, j]$ has either $s - 1$ or s terms. If M_ℓ has s terms, then the terms in $M_{\ell+1}$ are a subset of the terms in M_ℓ and thus $\beta \Vdash M_{\ell+1}$ by Corollary 11. If M_ℓ has $s - 1$ terms, then by Lemma 14, already $\alpha \Vdash M_\ell$. We can extend α to α' which decides $M_{\ell+1}$ by adding at most one literal, and then again apply Corollary 11. □

7 Consequences of the main result

In this section, we describe the consequences of our result outlined in Section 1.1.

7.1 New space lower bounds for PCR

As mentioned in the introduction, there are some CNF formulas for which it has seemed reasonable to expect PCR space lower bounds but, by [14], the general framework for proving such bounds developed in [8] either provably does not work or seems not to. Examples include the linear ordering principle and the functional pigeonhole principle.

7.1.1 Linear ordering principle

The *linear ordering principle* encodes the property that a finite linearly ordered set of n elements must have a minimal element. An unsatisfiable CNF formula expressing this principle, LOP_n, uses variables x_{ij}, for $i \neq j \in [n]$, and consists of the clauses:

$$\begin{cases} \bar{x}_{ij} \vee \bar{x}_{ji} & i, j \in [n] \quad i \neq j \\ \bar{x}_{ij} \vee \bar{x}_{jk} \vee x_{ik} & i, j, k \in [n] \quad i \neq j \neq k \neq i \\ \bigvee_{j \in [n], i \neq j} x_{ij} & i \in [n]. \end{cases}$$

First we consider the graph version of this principle, $\mathrm{GOP}(G)$, introduced in [27], in the encoding used to prove a degree lower bound for PCR in [18]. Let $G = (V, E)$ be a simple undirected graph over n nodes, that is, $V = [n]$. Let $\Gamma(i)$ be the set of neighbours of i in G.

The variables of $\mathrm{GOP}(G)$ are x_{ij} for $i < j \in [n]$. $\mathrm{GOP}(G)$ is defined as the conjunction of the following clauses:

$$\begin{cases} x_{ij} \vee x_{jk} \vee \bar{x}_{ik} & i,j,k \in [n] \quad i < j < k \\ \bar{x}_{ij} \vee \bar{x}_{jk} \vee x_{ik} & i,j,k \in [n] \quad i < j < k \\ \bigvee_{j \in \Gamma(i), i<j} x_{ij} \vee \bigvee_{j \in \Gamma(i), i>j} \bar{x}_{ij} & i \in [n]. \end{cases}$$

Definition 22. ([18]) The graph G is an (r,c)-*vertex expander* if for any set $U \subseteq V$ with $|U| \leq r$ it holds that $|\Gamma(U)| \geq c|U|$. The value c is the *vertex expansion* of G.

Theorem 23. *Let G be a simple undirected constant-degree graph which is an (r,c)-vertex expander. Then, over any field, refuting $\mathrm{GOP}(G)$ in PCR requires space $\Omega(\sqrt{cr}/2)$.*

Proof. It was proved in [18] that the polynomial translation of $\mathrm{GOP}(G)$ requires degree $\Omega(cr/4)$ to refute in PCR. Hence, $\mathrm{GOP}(G)$ requires width $\Omega(cr/4)$ in resolution. The result follows using our main Theorem 20. □

Since there are constant-degree graphs G with $c = \Omega(n)$ (see [18] for the precise G to use), we have the following.

Corollary 24. *There are simple undirected graphs G over n nodes such that refuting $\mathrm{GOP}(G)$ requires PCR space $\Omega(\sqrt{n})$.*

We can also lift the lower bound to LOP_n.

Corollary 25. *Over any field, refuting LOP_n requires PCR space $\Omega(\sqrt{n})$.*

Proof. (sketch) Let $G = ([n], E)$ be as in Corollary 24. Consider the substitution ρ_G to the variables of LOP_n defined by $\rho(x_{ij}) = \bar{x}_{ji}$ and $\rho(\bar{x}_{ij}) = x_{ji}$ for $i > j \in [n]$. It is not difficult to see that after applying ρ_G, the antisymmetry axioms of LOP_n become tautologies of the form $x_{ij}\bar{x}_{ij} = 0$, the transitivity axioms of LOP_n become transitivity axioms of $\mathrm{GOP}(G)$, and the monomials translating the wide clauses of LOP_n become derivable from the corresponding axioms of $\mathrm{GOP}(G)$ by a series of multiplications (effectively, by weakening).

Assume that LOP_n has a PCR refutation in space s. We obtain a PCR refutation of $\mathrm{GOP}(G)$ in space $s + O(1)$ as follows. First apply ρ_G to the whole refutation. To turn this into a valid refutation of $\mathrm{GOP}(G)$, whenever the original refutation downloaded an antisymmetry axiom of LOP_n, we now derive the monomial $x_{ij}\bar{x}_{ij}$ at the cost of a constant increase in space. Whenever the original refutation downloaded an LOP_n axiom of the form $\bigvee_{j \in [n], i \neq j} x_{ij}$, we download the corresponding axiom of $\mathrm{GOP}(G)$ and obtain the axiom of LOP_n by a series of steps in which we multiply a monomial by a single variable and immediately delete the old monomial, keeping only the result of multiplication; this increases space by 1. With transitivity axioms, there is nothing to do. It is not difficult to see that what remains is a valid proof of $\mathrm{GOP}(G)$ of space $s + O(1)$. □

7.1.2 Functional pigeonhole principle

The *functional pigeonhole principle* FPHP^m_n, for $m > n$, asserts that there cannot exist a total injective function mapping m pigeons into n holes. Its encoding as an unsatisfiable CNF, built using variables x_{ij} variables for $i \in [m]$ and $j \in [n]$, is the following:

$$\begin{cases} \bigvee_{j \in [n]} x_{ij} & i \in [m] \\ \bar{x}_{ij} \vee \bar{x}_{i'j} & i \neq i' \in [m], j \in [n] \\ \bar{x}_{ij} \vee \bar{x}_{ij'} & i \in [m], j \neq j' \in [n]. \end{cases}$$

PCR space lower bounds for FPHP^m_n were so far unknown, and, as proved in [14], the framework developed in [8] could not be used in this case.

We consider two constant-width versions of the functional pigeonhole principle. The *extended* version of FPHP_n^m, eFPHP_n^m, is obtained by replacing each large initial clause $\bigvee_{j\in[n]} x_{ij}$ for $i \in [m]$ with the CNF

$$(y_{i1} \vee x_{i1}) \quad \wedge \quad \bigwedge_{1\leq j\leq n-1} (\bar{y}_{ij} \vee x_{ij} \vee y_{i(j+1)}) \quad \wedge \quad (\bar{y}_{in} \vee x_{in})$$

which uses mn new variables y_{ij} for $i \in [m], j \in [n]$. Width lower bounds of $\Omega(n)$ for eFPHP_n^m in resolution can be easily obtained by modifying a Prover-Adversary argument proving a width lower bound for FPHP_n^m. Hence Theorem 20 implies lower bounds of $\Omega(\sqrt{n})$ on the space needed to refute eFPHP_n^m in PCR. The functional pigeonhole principle is an example of formula which is *weight-constrained* in the terminology of [16] (see Definition 7.1 in [16]). As such it was shown in [16, Theorem 1.5] that the PCR space needed to refute FPHP_n^m and eFPHP_n^m can differ by at most a constant factor. Hence Theorem 20 implies PCR space lower bounds for FPHP_n^m as well.

Corollary 26. *Over any field, refuting* FPHP_n^m *in PCR requires space* $\Omega(\sqrt{n})$.

A different constant-width version of the functional pigeonhole principle is the functional pigeonhole principle over bipartite graphs G, as defined in [21]. Using known width and degree lower bounds, we get a similar PCR space lower bound for this family of formulas when G is a suitable graph. Let $G = (U, V, E)$ be a bipartite graph. $\mathrm{FPHP}(G)$ is defined using variables x_{uv}, for $u \in U$, $v \in V$, as

$$\begin{cases} \bigvee_{v\in\Gamma(u)} x_{uv} & u \in U \\ \bar{x}_{uv} \vee \bar{x}_{u'v} & v \in V, u \neq u' \in \Gamma(v) \\ \bar{x}_{uv} \vee \bar{x}_{uv'} & u \in U, v \neq v' \in \Gamma(u). \end{cases}$$

Definition 27. ([21, Definition 5.1]) A bipartite graph $G = (U, V, E)$ is an (s, δ)-*boundary expander* if for each $U' \subseteq U$ with $|U'| \leq s$, it holds that $|\partial(U')| \geq \delta|U'|$, where the *boundary* $\partial(U)$ of U is $\{v \in V \ : \ |\Gamma(v) \cap U'| = 1\}$.

Theorem 28. *([21, Theorem 5.9]) Let $G = (U, V, E)$ be a bipartite graph which is an (s, δ)-boundary expander with left-degree bounded by d. Refuting* $\mathrm{FPHP}(G)$ *in PCR requires degree strictly greater than $\delta s/2d$.*

Hence $\mathrm{FPHP}(G)$ also requires width $\delta s/2d$ in resolution. From Theorem 20 we conclude:

Theorem 29. *Let $G = (U, V, E)$ be a bipartite graph which is an (s, δ)-boundary expander with left-degree bounded by d. Refuting* $\mathrm{FPHP}(G)$ *in PCR requires space* $\Omega(\sqrt{\delta s/2d})$.

Since, as mentioned in [21], there exist bipartite graphs with $|U| = n + 1$, $|V| = n$ and with left-degree 3 which are $(\gamma n, \delta)$-boundary expanders for $\gamma, \delta > 0$, we can conclude:

Corollary 30. *There exist bipartite graphs G with $|U| = n + 1$ and $|V| = n$ such that refuting* $\mathrm{FPHP}(G)$ *in PCR requires space* $\Omega(\sqrt{n})$.

7.2 Separations independent of characteristic

7.2.1 Separation of size from space

In [14], a separation of size and degree from space was proved for PCR: for each characteristic $p > 0$, there is a family of constant-width CNFs that have small low-degree refutations in PCR over characteristic p but require large PCR space over any field. However, it was left as an open problem whether there are formulas witnessing this sort of separation regardless of the characteristic of the field.

Theorem 20, together with the degree lower bound of [18] (which holds for any field) and the polynomial size resolution proofs for $\mathrm{GOP}(G)$ (see [18]) allow us to obtain a separation of PCR size and space independent of characteristic, using $\mathrm{GOP}(G)$.

Theorem 31. *Over any field, there are PCR refutations of size $O(n^3)$ of $\mathrm{GOP}(G)$ for any G. If G is the constant-degree vertex-expander graph with expansion $\Omega(n)$ of [18], then, over any field, refuting $\mathrm{GOP}(G)$ requires PCR space $\Omega(\sqrt{n})$.*

7.2.2 Separation of size and degree from space

To separate both size and degree from space in a way that works over any characteristic, we turn to a version of the bijective (functional onto) pigeonhole principle. Consider the following CNF, which we denote ex-bij-PHP$_n^{n+1}$. The variables are x_{ij}, y_{ij}, z_{ij} for $i \in [n+1], j \in [n]$. The idea is that the variables x_{ij} represent a bijection between $n+1$ pigeons and n holes, y_{ij} means that pigeon i goes to a hole with number at most j, and z_{ij} means that hole j is occupied by a pigeon with number at most i. The axioms are:

$$\begin{cases} \bar{x}_{ij} \vee \bar{x}_{i'j} & i \neq i' \in [n+1], j \in [n] \\ \bar{x}_{ij} \vee \bar{x}_{ij'} & i \in [n+1], j \neq j' \in [n] \\ y_{i1} \leftrightarrow x_{i1} & i \in [n+1] \\ y_{i(j+1)} \leftrightarrow (x_{i(j+1)} \vee y_{ij}) & i \in [n+1], j \in [n-1] \\ y_{in} & i \in [n+1] \\ z_{1j} \leftrightarrow x_{1j} & j \in [n] \\ z_{(i+1)j} \leftrightarrow (x_{(i+1)j} \vee z_{ij}) & i \in [n], j \in [n] \\ z_{(n+1)j} & j \in [n] \end{cases}$$

The equivalences $\leftrightarrow$ are written out as sets of clauses of width three or less, so ex-bij-PHP$_n^{n+1}$ is a 3-CNF of size $O(n^3)$. We have the following:

Theorem 32. *Over any field, the formula* ex-bij-PHP$_n^{n+1}$ *has a* poly(n)*-size, $O(1)$-degree PCR refutation, but requires space $\Omega(\sqrt{n})$ to refute in PCR.*

Proof. It is easy to verify, using for instance a routine Prover-Adversary argument, that refuting ex-bij-PHP$_n^{n+1}$ in resolution requires width $\Omega(n)$. Thus, our Theorem 18 gives the lower bound on PCR space.

To prove the existence of the polynomial size, constant-degree refutations of ex-bij-PHP$_n^{n+1}$, we show that, over any field, we can use ex-bij-PHP$_n^{n+1}$ to give a polynomial size, constant-degree derivation of the version of the bijective pigeonhole principle in which the statements that each pigeon goes to some hole and that each hole is occupied are expressed by means of sums rather than wide clauses:

$$\begin{cases} 1 - \sum_j x_{ij} & i \in [n+1] \\ 1 - \sum_i x_{ij} & j \in [n]. \end{cases}$$

It is well-known that over any field the sum version of the bijective pigeonhole principle between $n+1$ and n has a polynomial size, constant-degree PC refutation [26]. The idea is that adding up the axioms pigeon-by-pigeon gives $\sum_{ij} x_{ij} = n+1$, and adding them hole-by-hole gives $\sum_{ij} x_{ij} = n$; this implies $1 = 0$ over any field.

Now fix i. We sketch a derivation of $1 - \sum_j x_{ij}$ from the axioms of ex-bij-PHP$_n^{n+1}$, leaving the details to the reader (the derivation of $1 - \sum_i x_{ij}$ for fixed j is analogous). First, we replace x_{i1} by y_{i1} and then use the polynomials $x_{ij}x_{ij'}$ translating the axioms $\bar{x}_{ij} \vee \bar{x}_{ij'}$, together with the polynomials

$$x_{i(j+1)}(1 - y_{i(j+1)}) \qquad y_{ij}(1 - y_{i(j+1)}) \qquad y_{i(j+1)}(1 - y_{ij})(1 - x_{i(j+1)})$$

translating the axioms introducing $y_{i(j+1)}$, to derive, for each $j = 1, \ldots, n$ in turn, polynomials $y_{ij}x_{ik}$ for all $k > j$. This together with the axioms introducing y_{ij} makes it possible to derive $y_{i(j+1)} - (x_{i(j+1)} + y_{ij})$ for each j. Use that and the polynomial $1 - y_{in}$ translating the axiom clause y_{in} to derive, for each $j = n, \ldots, 1$ in turn, $1 - (y_{ij} + \sum_{k>j} x_{ik})$. For $j = 1$ this easily gives $1 - \sum_j x_{ij}$. □

7.3 Simplification of previous lower bounds

Let $G = (V, E)$ be an undirected graph. Let $\chi : V \to \{0, 1\}$ be a function, which we call an *odd-charging of* G if $\sum_{v \in V} \chi(v)$ is an odd number. Consider variables x_e for $e \in E$ and define $\mathrm{Par}(v, \chi)$ to be the CNF expansion of the formula encoding that the parity of edges incident with v is exactly $\chi(v)$, i.e. $\bigoplus_{v \in e} x_e = \chi(v)$. The Tseitin formula $\mathrm{Ts}(G, \chi)$ over G and an odd-charging χ of G is defined as

$$\mathrm{Ts}(G, \chi) := \bigwedge_{v \in V} \mathrm{Par}(v, \chi)$$

Notice that if the maximal degree of a vertex in G is d then the size of $\mathrm{Ts}(G, \chi)$ is $\leq |V| 2^{d-1}$.

Definition 33. (Connectivity expansion [14]) The *connectivity expansion* of $G = (V, E)$, $c(G)$, is the largest c such that for every $E' \subseteq E$, with $|E'| \leq c$, the graph $G' = (V, E \setminus E')$ has a connected component of size strictly greater than $|V|/2$.

A lower bound on the space to refute $\mathrm{Ts}(G, \chi)$ in PCR is given by the following Theorem.

Theorem 34. *([14]) Let $G = (V, E)$ be a connected graph of bounded degree d such that E can be partitioned into cycles of length at most b. Then, over any field, refuting $\mathrm{Ts}(G, \chi)$ in PCR requires space at least $c(G)/4b - d/8$.*

In [14], obtaining a PCR space lower bound for $\mathrm{Ts}(G)$ over a random graph involves showing that, for a suitable model of random constant-degree graphs, with high probability a random graph has both strong enough connectivity expansion and the property that the set of edges can be partitioned into small cycles. The authors of [14] raise the issue whether PCR space lower bounds for Tseitin formulas can be proved using expansion alone.

We can obtain asymptotically the same space lower bound using only expansion. We consider the expansion $e(G)$ of a graph G as defined in [5] and we use their resolution width lower bound of $\Omega(e(G))$ for resolution proofs of $\mathrm{Ts}(G, \chi)$. Hence using our Theorem 20 we can improve the result of [14] to:

Theorem 35. *Let $G = (V, E)$ be a connected graph of constant-degree d. Then it holds over any field that refuting $\mathrm{Ts}(G, \chi)$ in PCR requires space $\Omega(\sqrt{e(G)})$.*

Since there are graphs G over n nodes with $e(G) = \Omega(n)$ (see [5]), our result is asymptotically as good as that of [14].

8 Open problems

A natural question is whether older PCR space lower bounds, such as those in [8], can be reproved (or extended) in our framework. For example: use the methods of this paper to show that if F has a *m-winning strategy* in the sense of [8] then F requires PCR space linear in m. These bounds are typically linear in resolution width, so this could potentially be a route to strengthening our result to a general linear lower bound on PCR space in resolution width, matching the bound on resolution space in [3]. This would be consistent with what is known.

In the other direction, it is possible that the results here are tight up to a constant factor. Showing this means finding a formula F which requires width w in resolution but which has a PCR refutation in space $O(\sqrt{w})$. Plausible candidates for F are the Tseitin tautologies on random bounded degree graphs considered in [14].

The intriguing possibility that our bounds are essentially tight for general configurational systems but not for PC or PCR has also not been ruled out.

Acknowledgements. We are grateful to Ilario Bonacina for discussions about the relation between these results and older PCR space lower bound techniques.

References

[1] Michael Alekhnovich, Eli Ben-Sasson, Alexander A. Razborov, and Avi Wigderson. Space complexity in propositional calculus. *SIAM J. Comput.*, 31(4):1184–1211, 2002.

[2] Michael Alekhnovich and Alexander A. Razborov. Lower bounds for polynomial calculus: Non-binomial case. In *Proceedings of FOCS 2001*, pages 190–199. IEEE Computer Society, 2001.

[3] Albert Atserias and Víctor Dalmau. A combinatorial characterization of resolution width. *J. Comput. Syst. Sci.*, 74(3):323–334, 2008.

[4] Eli Ben-Sasson and Nicola Galesi. Space complexity of random formulae in resolution. *Random Struct. Algorithms*, 23(1):92–109, 2003.

[5] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow - resolution made simple. *J. ACM*, 48(2):149–169, 2001.

[6] Christoph Berkholz and Jakob Nordström. Supercritical space-width trade-offs for resolution. In *Proceedings of ICALP 2016*, pages 57:1–57:14, 2016.

[7] Ilario Bonacina. Total space in resolution is at least width squared. In *Proceedings of ICALP 2016*, pages 56:1–56:13, 2016.

[8] Ilario Bonacina and Nicola Galesi. A framework for space complexity in algebraic proof systems. *J. ACM*, 62(3):23:1–23:20, 2015.

[9] Maria Luisa Bonet and Nicola Galesi. Optimality of size-width tradeoffs for resolution. *Computational Complexity*, 10(4):261–276, 2001.

[10] Sam Buss. Axiomatizations and conservation results for fragments of bounded arithmetic. In American Mathematical Society, editor, *Proceedings of the Workshop Logic and Computation*, volume 106, pages 57–84, 1990.

[11] Matthew Clegg, Jeff Edmonds, and Russell Impagliazzo. Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proceedings of SToC 1996*, pages 174–183, 1996.

[12] Stephen A. Cook and Robert A. Reckhow. The relative efficiency of propositional proof systems. *J. Symb. Log.*, 44(1):36–50, 1979.

[13] Juan Luis Esteban and Jacobo Torán. Space bounds for resolution. *Inf. Comput.*, 171(1):84–97, 2001.

[14] Yuval Filmus, Massimo Lauria, Mladen Miksa, Jakob Nordström, and Marc Vinyals. Towards an understanding of polynomial calculus: New separations and lower bounds - (extended abstract). In *Proceedings of ICALP 2013*, pages 437–448, 2013.

[15] Yuval Filmus, Massimo Lauria, Mladen Miksa, Jakob Nordström, and Marc Vinyals. From small space to small width in resolution. *ACM Trans. Comput. Log.*, 16(4):28:1–28:15, 2015.

[16] Yuval Filmus, Massimo Lauria, Jakob Nordström, Noga Ron-Zewi, and Neil Thapen. Space complexity in polynomial calculus. *SIAM J. Comput.*, 44(4):1119–1153, 2015.

[17] Nicola Galesi and Massimo Lauria. On the automatizability of polynomial calculus. *Theory Comput. Syst.*, 47(2):491–506, 2010.

[18] Nicola Galesi and Massimo Lauria. Optimality of size-degree tradeoffs for polynomial calculus. *ACM Trans. Comput. Log.*, 12(1):4:1–4:22, 2010.

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

[20] Leszek Aleksander Kołodziejczyk and Neil Thapen. Approximate counting and NP search problems. *arXiv preprint arXiv:1812.10771*, 2018.

[21] Mladen Miksa and Jakob Nordström. A generalized method for proving polynomial calculus degree lower bounds. In *Proceedings of CCC 2015*, pages 467–487, 2015.

[22] Jakob Nordström. On the interplay between proof complexity and SAT solving. *ACM SIGLOG News*, 2(3):19–44, 2015.

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

[24] Alexander A. Razborov. Proof complexity of pigeonhole principles. In Werner Kuich, Grzegorz Rozenberg, and Arto Salomaa, editors, *Developments in Language Theory, 5th International Conference, DLT 2001, Vienna, Austria, July 16-21, 2001, Revised Papers*, volume 2295 of *Lecture Notes in Computer Science*, pages 100–116. Springer, 2001.

[25] Alexander A. Razborov. Resolution lower bounds for the weak functional pigeonhole principle. *Theor. Comput. Sci.*, 303(1):233–243, 2003.

[26] S. Riis. *Independence in Bounded Arithmetic*. PhD thesis, Oxford University, 1993.

[27] Nathan Segerlind, Samuel R. Buss, and Russell Impagliazzo. A switching lemma for small restrictions and lower bounds for k-DNF resolution. *SIAM J. Comput.*, 33(5):1171–1200, 2004.

[28] Alasdair Urquhart. Hard examples for resolution. *J. ACM*, 34(1):209–219, 1987.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il