THEORY OF COMPUTING
www.theoryofcomputing.org

Polynomial calculus space and resolution width

Nicola Galesi* Leszek A. Kołodziejczyk† Neil Thapen‡

May 16, 2022

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.

A conference version of this paper appeared in the Proceedings of the 60th Annual Symposium on Foundations of Computer Science (FOCS), 2019 [20].

*Partially supported by the MIUR under grant "Dipartimenti di eccellenza 2018-2022" of the Department of Computer Science of Sapienza University.

†Partially supported by grant 2017/27/B/ST1/01951 of the National Science Centre, Poland.

‡Partially supported by GA ČR project 19-05497S and by RVO:67985840.

ACM Classification: F.2.2, F.4.1

AMS Classification: 03F20, 68Q17

Key words and phrases: proof complexity, resolution, polynomial calculus, space, width

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 [15] 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 [7], 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 [16], 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 [16], 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 [16, 6, 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. [14] 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 are 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*, that is, 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 [14, 23], which inspired the similar connection between width and size for resolution of [7]. This result made it possible to lift most of the known degree lower bounds for PCR to size bounds [29, 23, 2, 22, 21, 26].

We define the *space* of a PCR proof to be the maximum number of *distinct* monomials that must be simultaneously in memory during a verification of the proof. It is also common in the literature to define space by counting the *total* number of monomials in memory, including repetitions; clearly any lower bound on our notion of PCR space will also hold for this measure. 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. 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 (and it is open whether there is a separation if we count PCR space with repetitions). Eventually, research on limitations of proof space in PCR led to several lower bounds [1, 10, 19, 8, 17] and to a framework to prove them [10].

An important open problem raised several times (see [27, 10, 19, 8, 17, 18]) 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. The optimal result, consistent with present knowledge about PCR space bounds and resolution width bounds, would be that a CNF that can be refuted in PCR space s can also be refuted in resolution using width linear in s. We are not able to prove this, but we show a weaker, quadratic bound. Our main result is the following theorem.

Theorem 6.4. *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 6.4 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. An earlier result in this direction appeared in [17], building on the framework of [1, 10], showing a relation between the resolution width of a formula F and the PCR space of a lifted version of F. Precisely, if F requires resolution width w then its XOR-ified version requires PCR space $\Omega(w)$.

The previous PCR space lower bounds of which we are aware all ultimately rely on a combinatorial argument from [1]. Our approach, which we outline in the next subsection, is quite different. Using it, we also get a very simple proof of Bonacina's recent result [9] that, in resolution, total space is lower bounded by width squared. Our proof of that result (Theorem 3.4) in particular does not use any technical notion such as that of *asymmetric width* required in [9].

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 in such systems we get the weaker bound of $2s^2 + k$ on resolution width (Theorem 6.1). This class is similar in spirit to, and includes, the semantic *functional calculus* system of [1].

As a consequence of Theorem 6.4 we answer some open questions about the relation between space, size, and degree in PCR. Since our bound is quadratic, in some cases the answers are not tight. A brief discussion of these follows.

New space lower bounds for PCR. The framework developed in [10] can be used to derive all space lower bounds for PCR known until now. However, as observed in [17], 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 as versions of them with constant initial width. Using well-known width lower bounds for these formulas [12, 22, 31, 33, 26] and Theorem 6.4 we are now able to prove PCR space lower bounds.

Simplification and generalization of a previous lower bound. Space lower bounds of the order $\Omega(\sqrt{n})$ for the well-known *Tseitin formulas* $\mathrm{Ts}(G)$ are shown in [17]. These bounds are for families of random graphs G over n nodes admitting two properties: good expansion and that the edges of G can be partitioned into small cycles. Applying Theorem 6.4 and the linear width lower bound for $\mathrm{Ts}(G)$ proved for expander graphs in [7], we simplify and asymptotically match the space lower bound in [17] using only an expansion property.

Separations independent of characteristic. It is left open in [17] whether there are formulas separating PCR size and degree from space for all fields at once, independently of the characteristic. We obtain some such examples, though due to the quadratic term in Theorem 6.4 the separations are not as strong as the characteristic-dependent ones from [17].

Our space lower bounds for linear ordering principles give a characteristic-independent example separating PCR size from space. A further example is provided by a variant of the *bijective* (*both functional and onto*) pigeonhole principle. Riis ([32, 30]) proved that the bijective pigeonhole principle formulas for $n+1$ pigeons and n holes have small PCR refutations in constant degree, over any field. Riis' result concerned a version of the principle where translations of wide clauses are replaced by certain sums, but we check that it also holds for the usual formulation of bijective PHP restricted to bounded-degree graphs. On the other hand, it is known that bijective PHP restricted to certain bounded-degree expanders requires $\Omega(n)$ width to refute in resolution. Hence, Theorem 6.4 gives us a separation of PCR size and degree from space independent of characteristic.

1.2 Outline of technique

Consider presenting a refutation of a CNF F on a blackboard. At each step we either write a clause of F on the board ("upload it to memory"), or do some logical manipulation of formulas already on the board, or erase a formula to make room. The presentation ends when we are able to write down a predetermined contradiction, such as $1=0$ (see Section 2.1). With this in mind, our model of a refutation is a sequence of *memory configurations* $M_0,\dots,M_t$, where M_i describes the contents of the blackboard at time i. The space required by the refutation is the size of the largest configuration, measured in some appropriate way.

Proof space lower bounds typically have the form: under the assumption that a refutation uses small memory, work forward through the refutation, at each step building a small partial assignment which semantically implies every formula in memory; but this is impossible, because the last step of the refutation contains an unsatisfiable formula. A dual argument also appears in proof complexity, in proofs of resolution width lower bounds: work backwards through the refutation from the end, maintaining a small assignment which falsifies one of the clauses in memory (a related construction was used in [18] for an alternate proof of the Atserias-Dalmau result [3] that space is lower bounded by width).

Our new idea for proving space lower bounds is to combine these forms of argument and pass backwards and forwards through the refutation possibly several times, satisfying part of the memory as we go down, and dually falsifying part as we go up. Our method is inspired by a propositional version of 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 [13, Corollary 4]. 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 [9, 11], where total space counts the total number of symbols simultaneously on the blackboard. Given a formula F, we let $\mathcal{H}$ be an *Atserias-Dalmau family* for F. This is a family of partial assignments "locally" satisfying F, and 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 steps in the interval $[i, j]$ under β.

A key point in the argument for resolution is that we can satisfy high-width clauses in memory using a restricted-size assignment from the class $\mathcal{H}$. To apply a similar argument for PCR space we have to understand how to determine the value of a high-degree monomial using a small assignment $\alpha \in \mathcal{H}$. We use a very simple version of the forcing method known e.g. from set theory, which has already appeared in various guises in proof complexity. The idea is that α forces a monomial to a value if no extension of α will ever give the monomial a different value, as long as we only consider extensions within $\mathcal{H}$. In the case of PCR, the simple one-interval construction used in our proof of Bonacina's result sketched above is not enough to obtain a contradiction. Instead, we have to iterate the construction, refining the interval and extending the assignment α some number of times bounded by the space s used in the PCR refutation. Each time, α grows by at most $O(s)$ literals, and after at most s iterations the restricted refutation becomes trivial. We reach a contradiction as long as the resolution width required to prove F is larger than $O(s^2)$; this gives 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 5.6). In Section 6 we extend this argument to give our main results, a $2s^2+k$ bound for any configurational system (Theorem 6.1) and an s^2-s+k bound for PCR (Theorem 6.4). 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 $\perp/\top$. A *term* is a set of literals, treated as a conjunction. 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 *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 $C_0 \dots, C_m$ 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 [14], 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 $\mathbb{F}[x_1,\dots,x_n]$. 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 $\mathbb{F}[x_1,\dots,x_n,\bar{x}_1,\dots,\bar{x}_n]$, 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 $p_0,\dots,p_t$ 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

boolean axioms: $\dfrac{}{x^2 - x}$ **complementarity axioms:** $\dfrac{}{x+\bar{x}-1}$

linear combination: $\dfrac{p \quad q}{ap+bq}$ **multiplication:** $\dfrac{p}{xp}$

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* $M_0,\dots,M_t$.

In resolution, a configuration is a set of clauses and a refutation $M_0,\dots,M_t$ is such that the first configuration is empty, the last one contains the empty clause, and for each $i<t$, configuration 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,

(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.

Definition 2.1. The *clause space*, or simply *space*, of such a resolution 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_i, 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 $M_0,\dots,M_t$ where M_0 is empty, M_t contains the polynomial 1, and for each $i<t$, configuration M_{i+1} is obtained from M_i by the rules (1)-(3) above, adapted to PCR. So in rule (1) the axioms we can download are polynomials translating the clauses of F and instances of the boolean and complementarity axioms, and in rule (3) we can infer new polynomials by linear combination or multiplication. There are several possible definitions of the "monomial space" of a PCR configuration. We could count monomials or just count terms (that is, ignore coefficients), and we could count them with or without repetitions. We choose to ignore coefficients and count without repetitions, that is, to work with what we call *term space*, as defined below. In particular this is always less than or equal to the other measures, so our lower bounds will carry across.

Definition 2.2. The *term space* of a PCR configuration M_i is the number of distinct terms represented by the monomials in M_i. The *term space*, or simply *space*, of a PCR refutation is the maximum term space of any configuration M_i in the refutation.

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 Γ of formulas and a set of rules. Each rule is sound (over $0/1$ assignments) and takes as premises either a *single* formula from Γ, or a formula from Γ together with a clause; its conclusion is a formula from Γ. A simple example of such a rule is "from φ and a clause C derive $\varphi\wedge C$", but if we replaced $\varphi\wedge C$ with any logical consequence of $\varphi\wedge C$ in Γ, this would also be a valid rule. A refutation of a CNF F in the system is a sequence $M_0,\dots,M_t$ of formulas from Γ, called *configurations*. M_0 is the constant $\top$, M_t is the constant $\bot$, and each M_{i+1} is obtained from applying a rule to the previous configuration M_i, possibly together with some initial clause C of F. 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 $1=0$.

Notice that each formula in such a refutation (not counting initial clauses) is used at most once, so in this sense the refutation is treelike. In fact it is "pathlike", since every formula is derived from exactly one premise (again if we do not count premises which are initial clauses).

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, over $0/1$ assignments, 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 3.1 ([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 3.2 ([3]). *Let $w \geq k$. 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.*

In fact most of the time (except for Section 6.2) we prefer to use a weaker version of this lemma, which gives only a family of width w. Using the full version would involve improving many bounds by 1, which would be messy to write. This weaker version has a simple and intuitive proof which we now sketch.

The *width-w Prover-Adversary game* on F 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) of size at most w and who in each round either asks the Adversary the value of a variable and adds the answer to α, or forgets variables from α to free some memory. The Prover wins when α falsifies some clause from F.

Let us say that the *starting position* of the game is the initial content α of the Prover's memory. By replacing each clause in a refutation with the partial assignment negating it, and flipping the direction of the edges in the underlying graph, we can identify width-w resolution refutations of F with winning strategies for the Prover in the game whose starting position is the empty partial assignment. If there is no such Prover strategy (equivalently, if there is no width-w refutation of F), then it is not hard to show that the set of starting positions for which the Adversary has a winning strategy satisfies (i)-(iv) above.

Theorem 3.3 ([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 $M_0,\dots,M_t$ 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 $|\alpha| \le s$. 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 3.1 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 the Prover strategy corresponding to a small-width refutation in Lemma 3.2 starts at the bottom of the proof and works up, trying to falsify clauses. An alternative proof of Theorem 3.3 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 3.4. *Let F be a k-CNF. Let $m,s \in \mathbb{N}$ with $m \ge k$. 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 $M_0,\dots,M_t$ be the configurational resolution refutation. Each M_i contains some number q of *narrow* clauses $C_1,\dots,C_q$ of width at most m, and $r \le s$ many *wide* clauses $D_1,\dots,D_r$ of width greater than m. Suppose for a contradiction that F has no resolution refutation of width $2m+s$. Let $\mathcal{H}$ be an 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, $|\alpha| \le m$. Since C is falsified by α, it cannot have been introduced by axiom download. So we must have $C = E \vee G$ for clauses $E \vee x$ and $G \vee \bar{x}$ in M_{j-1}. Extend α to $\alpha' \in \mathcal{H}$ which gives a value to x, with $|\alpha'| \le m+1$. Without loss of generality $\alpha'(x) = 1$. Hence α' falsifies $G \vee \bar{x}$, and by minimality of j, we know that $G \vee \bar{x}$ 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, $|\beta| \le |\alpha'| + s \le m+s+1$. Since α' falsifies $G \vee \bar{x}$, 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 we 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 $|\beta'| \le m+s+2$. 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 $|\gamma| \le |\beta'| + m - 1 \le 2m+s+1$ 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 3.4 has the following consequence, which is essentially the main result of [9] with the constant improved by a factor of two.

Corollary 3.5. *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", which we will use in the next section to prove our main result. This is in fact a very simple version of a forcing relation as used in set theory and other areas of logic. Definitions in a similar spirit are common in proof complexity, where we often want to "evaluate" complex formulas over families of partial assignments. See for example the evaluation of formulas as decision trees in lower-bound proofs for constant-depth Frege [5], and see [28, 25] for a recent application of essentially Definition 4.1 below. We present the constructions and proofs here 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 4.1. 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 *decides* t. (In [20, 28] this relation is called "α *fixes* t".)

We write (i) as $\alpha \Vdash t = 0$ and (ii) as $\alpha \Vdash t = 1$. Note that although (i) and (ii) appear rather different from each other, they both have the effect that no extension of α will ever give t a different value, as long as we only consider extensions within $\mathcal{H}$. We now extend the definition to polynomials and configurations. We will treat polynomials as linear combinations of terms over our field $\mathbb{F}$.

Definition 4.2. 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 decides every term t_i in p. We say that α *decides* a configuration M if it decides 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 $\alpha \Vdash p = 0$, $\alpha \Vdash p \neq 0$, $\alpha \Vdash M$, $\alpha \Vdash \neg M$. 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 deciding 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 4.3. 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 4.4. *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 4.5. *Let $\alpha \in \mathcal{H}$ and let $t_1, \dots, t_s$ be terms. Then there is $\beta \supseteq \alpha$ in $\mathcal{H}$ such that β decides $t_1, \dots, t_s$ 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 4.6. *Let $\alpha \in \mathcal{H}$ with $|\alpha| < w$. Let $t_1, \dots, t_r$ be terms and $b_1, \dots, b_r$ 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. □

Lemma 4.7. *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 3.1, β satisfies a literal in C. We let $\alpha' \in \mathcal{H}$ be α plus this literal. Notice that $|\alpha'| < w$, due to the assumption that $k \geq 2$. List all terms in M and M' as $t_1, \dots, t_r$. Since α' decides all these terms, there exist boolean values $b_1, \dots, b_r$ such that $\alpha' \Vdash t_j = b_j$ for each j. We use Lemma 4.6 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 the fact that $M \wedge C$ semantically implies M'. □

Corollary 4.8. *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 Lemmas 4.5 and 4.7. □

This suggests a possible approach to proving PCR space lower bounds. Given a refutation $M_0,\dots,M_t$ with space s, use Corollary 4.8 to inductively find $\alpha_0,\dots,\alpha_t$ 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, 6, 10]. 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 5.4 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 – initial version

This section is devoted to a proof of an initial, somewhat simpler, version of our main result giving a width bound of $2s(s+1)+k$. In the next section we will use essentially the same proof, but with more careful counting, to get improved bounds of $2s^2+k$ for a general configurational system and $s(s-1)+k$ for PCR in particular.

Here we assume that F is a k-CNF with a PCR refutation in space s over some fixed field $\mathbb{F}$. Let $M_0,\dots,M_t$ be the sequence of configurations forming the refutation of F. For $0 \leq i \leq j \leq t$, the *proof interval* $[i,j]$ is the sequence of configurations $M_i,\dots,M_j$. We may assume without loss of generality that $k \geq 3$, since if k is 1 or 2 then F always has a width-k refutation.

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 5.1. 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 5.2. *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 decides 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 5.3. *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 $|\beta| \le |\alpha| + s - m$.*

Proof. Repeat the proof of Lemma 4.5, and observe that at most $s-m$ terms can be made zero in this process. □

Lemma 5.4. *Let M contain at most s terms and let $\alpha \in \mathcal{H}$. Suppose there is $\gamma \supseteq \alpha$ in $\mathcal{H}$ such that $|\mathrm{NZ}(M,\gamma)| = m$. Then there is β with $\alpha \subseteq \beta \subseteq \gamma$ such that $|\mathrm{NZ}(M,\beta)| = m$ and $|\beta| \le |\alpha| + s - m$.*

Suppose furthermore that α guarantees m non-zeroes in M. Then β decides M, and either both β and γ force M or both β and γ force $\neg M$.

Proof. List the terms in M as $t_1,\ldots,t_r$ with $r \le s$. 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 $|\beta| \le |\alpha| + s - m$. In the "furthermore" part, β, and hence also γ, decides M by Lemma 5.2. The reason why β and γ force the same value for M is that $\beta \subseteq \gamma$ and forcing is preserved under extensions within $\mathcal{H}$. □

Lemma 5.5 (Main Lemma)**.** *Let F be a k-CNF with a PCR refutation $M_0,\ldots,M_t$ in term space s, and let $\mathcal{H}$ be a width-w Atserias-Dalmau family for F. Suppose $w \ge 2s(s+1)+k$. Then for each $m \le s$ 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) $|\alpha| \le 4\sum_{r=0}^{m-1}(s-r)$.

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 condition (iii) implies

$$|\alpha| + 4(s-m) \le w - k$$

since $|\alpha| + 4(s-m) \le 4[s+(s-1)+\cdots+1] = 2s(s+1) \le w-k$. We will extend α first to an assignment α' with $|\alpha'| + 2(s-m) \le w-k$, then to the required α'', using the size bounds to make sure that the assignments we construct are well-behaved.

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; by Lemma 4.4 and condition (i), it must be the case that $\ell < j$. By condition (ii) and Lemma 5.4, we may assume without loss of generality that $|\beta| \le |\alpha| + s - m$. By condition (ii) and Lemma 5.3, we may extend β to $\alpha' \in \mathcal{H}$ with $|\alpha'| \le |\alpha| + 2(s-m)$ such that α' decides $M_{\ell+1}$. Since $\beta \Vdash M_\ell$, it follows from Lemma 4.7 and the bound on $|\alpha'|$ 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 $|\alpha'| \le |\alpha| + 2(s-m)$ 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 5.4 we may assume $|\beta| \le |\alpha'| + s - m$. By condition (ii) and Lemma 5.2, 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 5.3, we extend β to $\alpha'' \in \mathcal{H}$ with $|\alpha''| \le |\alpha'| + 2(s-m) \le w - k$ such that α'' decides $M_{\ell-1}$. We cannot have $\alpha'' \Vdash M_{\ell-1}$, by Lemma 4.7. 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, $|\alpha''| \le |\alpha| + 4(s-m)$ 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 $\mathcal{H} \ni \gamma \supseteq \alpha''$ and $i' \le \ell \le j'$, then $|\mathrm{NZ}(M_\ell, \gamma)| \ne m$. Thus α'' in fact guarantees $m+1$ non-zeroes in $[i', j']$. □

Theorem 5.6. *Let F be a k-CNF. If F has a PCR refutation in term 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 5.5 leads to a contradiction for $m = s$. The lemma gives us a proof interval $[i, j]$ and $\alpha \in \mathcal{H}$ with $|\alpha| \le w - k$ such that $\alpha \Vdash M_i$, $\alpha \Vdash \neg M_j$ and α guarantees s non-zeroes in $[i, j]$. For each $\ell \in [i, j]$, Lemma 5.2 shows that α decides M_ℓ. Using the fact that $\alpha \Vdash M_i$ and applying Lemma 4.7 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 5.6 works, even in a slightly stronger form, in any configurational proof system, not just in 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 decides all terms, fixing their values to say $b_1,\ldots,b_s$. 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 Section 4 and Section 5 go through for any configurational system, as we did not use any properties of PCR except for soundness of the rules. Thus, the bound of $2s^2+2s+k$ from Theorem 5.6 holds for all configurational systems, with the same proof. In fact, we now show that the $2s$ term can be removed from the bound by means of a more careful argument (readers who are not interested in this improvement can skip ahead to Section 6.2).

Theorem 6.1. *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 $2s^2+k$.*

The rest of this subsection is devoted to a proof of Theorem 6.1. We argue as follows. Suppose F is a k-CNF with a space s refutation in a configurational system, but without a resolution refutation of width $2s^2+k$. Then there is an Atserias-Dalmau family $\mathcal{H}$ for F of width $w=2s^2+k$. The theorem is now proved using the argument from Section 5 and the following strengthened version of Lemma 5.5.

Lemma 6.2. *Let F be a k-CNF with a refutation $M_0,\ldots,M_t$ in term space s in a configurational system, and let $\mathcal{H}$ be a width-w Atserias-Dalmau family for F. Suppose $w \geq 2s^2+k$. Then for each $m \leq s$ there is $\alpha \in \mathcal{H}$ and a proof interval $[i,j]$ in the configurational refutation such that*

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

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

(iii) $|\alpha| \leq 4\sum_{r=0}^{m-1}(s-r)-2m$.

To prove Lemma 6.2, we need the following more precise version of Lemma 5.3.

Lemma 6.3. *Let $\alpha \in \mathcal{H}$ and let M be a configuration. Then there is $\beta \supseteq \alpha$ in $\mathcal{H}$ which decides M with* $|\beta| \leq |\alpha| + |\mathrm{NZ}(M,\alpha)| - |\mathrm{NZ}(M,\beta)|$.

Proof. List $\mathrm{NZ}(M,\alpha)$ as $t_1,\ldots,t_p$. Consider each t_i in turn and, if possible in $\mathcal{H}$, add one literal to α to make t_i zero. As in Lemma 4.5, this gives $\beta \supseteq \alpha$ in $\mathcal{H}$ which decides M. The number of literals added is bounded above by the number of terms made 0, which is precisely $|\mathrm{NZ}(M,\alpha)| - |\mathrm{NZ}(M,\beta)|$. □

Proof of Lemma 6.2. In the proof of Lemma 5.5, at each induction step from m to $m+1$ we grew our assignment α in four stages. That is, it gained up to $s-m$ bits twice in case (a), and again up to $s-m$ bits twice in case (c). Thus it increased by at most $4(s-m)$ bits in total, giving the bound $|\alpha| \leq 4\sum_{r=0}^{m-1}(s-r)$ in item (iii) at stage m of Lemma 5.5. We will show here that we can save two bits in each induction step, leading to the bound in the current lemma. Precisely, in case (a) we will grow α by first $s-m$ and then $s-m-1$ bits, and then the same in case (c). As in Lemma 5.5, in case (b) or case (d) we do not need to grow α and the requirement on its size becomes more relaxed.

So let $m < s$ and suppose we have the inductive hypothesis for m and want to prove it for $m+1$. We plan to grow α by at most $4(s-m)-2$ bits, and we observe

$$\begin{aligned}|\alpha|+4(s-m)-2 \le 4\sum_{r=0}^{m}(s-r)-2m-2 &= 2\sum_{r=0}^{m}(s-r)+2\sum_{r=0}^{m}(s-r-1)\\ &\le 2\sum_{r=0}^{s}(s-r)+2\sum_{r=0}^{s-1}(s-r-1)\\ &= s(s+1)+s(s-1)=2s^2 \le w-k.\end{aligned}$$

Imitating the proof of Lemma 5.5, suppose we are in case (a) at the left end of our current proof 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 $\beta \Vdash M_\ell$, and we have chosen ℓ maximal, so that there is no such β for $M_{\ell+1}$. Furthermore, α guarantees m non-zeroes at $M_{\ell+1}$, and $|\beta| \le |\alpha|+s-m$.

In Lemma 5.5, we extended β to $\alpha' \in \mathcal{H}$ with $\alpha' \Vdash M_{\ell+1}$ and $|\alpha'| \le |\beta|+s-m$. We want to improve this bound to $|\alpha'| \le |\beta|+s-m-1$. By Lemma 6.3 there is $\alpha' \supseteq \beta$ in $\mathcal{H}$ which decides $M_{\ell+1}$ with

$$|\alpha'| \le |\beta| + |\mathrm{NZ}(M_{\ell+1},\beta)| - |\mathrm{NZ}(M_{\ell+1},\alpha')|. \qquad (\star)$$

We have $|\mathrm{NZ}(M_{\ell+1},\beta)| \le s$ and $|\mathrm{NZ}(M_{\ell+1},\alpha')| \ge m$ so $|\alpha'| \le |\beta|+s-m$. Thus $|\alpha'| \le |\alpha|+2(s-m) \le |\alpha|+4(s-m)-2 \le w-k$, using the assumption that $m < s$. Hence we can apply Lemma 4.7 to get that $\alpha' \Vdash M_{\ell+1}$. This in turn implies that $|\mathrm{NZ}(M_{\ell+1},\alpha')| \ge m+1$ by maximality of ℓ. Putting this improved bound on $|\mathrm{NZ}(M_{\ell+1},\alpha')|$ back into ($\star$) gives us the stronger bound on $|\alpha'|$.

Now suppose we are in case (c) at the right end of the proof 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 $M_{\ell-1}$ and now we have the bounds $|\alpha'|+2(s-m)-1 \le w-k$ and $|\beta| \le |\alpha'|+s-m$. We know that $\beta \Vdash \neg M_\ell$ and thus that $\ell > i'$. By Lemma 6.3 there is $\alpha'' \supseteq \beta$ in $\mathcal{H}$ deciding $M_{\ell-1}$ with $|\alpha''| \le |\beta| + |\mathrm{NZ}(M_{\ell-1},\beta)| - |\mathrm{NZ}(M_{\ell-1},\alpha'')|$. By the minimality of ℓ, we have $|\mathrm{NZ}(M_{\ell-1},\alpha'')| \ge m+1$. As before, we also have $|\mathrm{NZ}(M_{\ell-1},\beta)| \le s$, so $|\alpha''| \le |\beta|+s-m-1$. Thus $|\alpha''| \le w-k$, so we can apply Lemma 4.7 and the fact that $\beta \Vdash \neg M_\ell$ to conclude that $\alpha'' \Vdash \neg M_{\ell-1}$. This completes the induction step. □

6.2 A stronger bound for PCR

We now show how to improve the bound by a factor of two in the case of PCR.

Theorem 6.4. *Let F be a k-CNF. If F has a PCR refutation in term space s over some field $\mathbb{F}$, then F has a resolution refutation of width s^2-s+k.*

The only specific property of PCR used in the proof is that if M_ℓ and $M_{\ell+1}$ are successive configurations in a PCR refutation, then either all the terms in $M_{\ell+1}$ appear in M_ℓ or all the terms in M_ℓ appear in $M_{\ell+1}$. Thus, if the PCR refutation has space s, we can always list the terms in M_ℓ as $t_1,\dots,t_p$ and the terms in $M_{\ell+1}$ as $t_1,\dots,t_q$ for some $p,q \le s$; it will depend on the rule used to derive $M_{\ell+1}$ whether $p \le q$ or vice versa. The reason why this can be helpful is that when we are trying to build an assignment deciding both M_ℓ and $M_{\ell+1}$, and we have first built one deciding the terms $t_1,\dots,t_{\min(p,q)}$ from the smaller of the two configurations, then we only have to decide the remaining terms in the larger configuration. The whole process will therefore require deciding at most $\max(p,q) \le s$ terms, and not up to $2s$ terms as would be the case if the configurations were more loosely related. This lets us obtain a version of Lemma 5.5 with better bounds.

The strengthening of Lemma 5.5 that we use is as follows.

Lemma 6.5. *Let F be a k-CNF with a PCR refutation $M_0,\ldots,M_t$ in term space s, and let $\mathcal{H}$ be a width-w Atserias-Dalmau family for F. Suppose $w \ge s(s-1)+k$. Then for each $m \le s-1$ 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) $|\alpha| \le 2\sum_{r=0}^{m-1}(s-1-r)$.

Proof. We use the same structure as the proofs of Lemmas 5.5 and 6.2, but with induction only up to $m = s-1$. As before, in the induction step we only have to consider cases (a) and (c), because (b) and (d) are trivial.

So, suppose that $m < s-1$ and that we are in case (a) at the left end of the proof 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 $\beta \Vdash M_\ell$, 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 $|\alpha| + 2(s-m-1) \le 2\sum_{r=0}^{m}(s-1-r) \le s(s-1) \le w-k$. In Lemma 5.5, we used β to find $\alpha' \supseteq \alpha$ in $\mathcal{H}$ with $\alpha' \Vdash M_{\ell+1}$ and $|\alpha'| \le |\alpha| + 2(s-m)$. We now want to improve this bound to $|\alpha'| \le |\alpha| + s - m - 1$.

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 $t_1,\ldots,t_q$ for some $p,q \le s$. By Lemma 5.4 we may assume $|\beta| \le |\alpha| + p - m$. If $q \le p$, then all terms in $M_{\ell+1}$ appear in M_ℓ, so already β decides $M_{\ell+1}$, and thus $\beta \Vdash M_{\ell+1}$ by Lemma 4.7. Moreover, in this case $|\mathrm{NZ}(M_{\ell+1},\beta)| \le |\mathrm{NZ}(M_\ell,\beta)| = m$, which contradicts the maximality of ℓ. So we must have $q > p$.

We apply the proof of Lemma 4.5 carefully to extend β to $\alpha' \in \mathcal{H}$ which decides 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 $|\beta| + (q-p) \le |\alpha| + p - m + (q-p) \le w-k$, and thus $\alpha' \Vdash M_{\ell+1}$ by Lemma 4.7. Hence α' cannot set all of the terms $t_{p+1},\ldots,t_q$ 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, which gives $|\alpha'| \le |\alpha| + p - m + (q-p-1) \le |\alpha| + s - m - 1$.

Now suppose we are in case (c) at the right end of the proof 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 $|\alpha'| + s - m - 1 \le w-k$. As in the proof of Lemma 5.5, 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 5.4 without loss of generality may assume $|\beta| \le |\alpha'| + q - m$.

Arguing as before, we see that now we it must be the case that $p > q$, since $p \le q$ would imply $|\mathrm{NZ}(M_{\ell-1},\beta)| \le |\mathrm{NZ}(M_\ell,\beta)| \le m$, contradicting the minimality of ℓ. By adding at most one literal to β

for each term $t_{q+1},\ldots,t_p$, we extend β to α'' which decides all these terms. Again, α'' cannot set all of the terms $t_{q+1},\ldots,t_p$ to 0 or it would contradict the minimality of ℓ. So, we we have $|\alpha''| \le |\beta| + p - q - 1 \le |\alpha'| + p - m - 1 \le w - k$. This means that we can apply Lemma 4.7 and the fact that $\beta \Vdash \neg M_\ell$ to conclude that $\alpha'' \Vdash \neg M_{\ell-1}$. Also, $|\alpha''| \le |\alpha| + 2(s-m-1)$. This completes the induction step. □

Proof of Theorem 6.4. If there is no such resolution refutation, then F has an Atserias-Dalmau family $\mathcal{H}$ of width $w = s^2 - s + k + 1$, by Lemma 3.2. We apply Lemma 6.5 for $m = s-1$. This gives us a proof interval $[i,j]$ and $\alpha \in \mathcal{H}$ with $|\alpha| \le s(s-1) \le w-k-1$ 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 $|\beta| \le |\alpha| + 1$ 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 Lemma 4.7. If M_ℓ has $s-1$ terms, then by Lemma 5.2, 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 Lemma 4.7. □

7 Consequences of the main result

In this section we describe some consequences of our result, as 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 [17], the general framework for proving such bounds developed in [10] 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 maximal element. An unsatisfiable CNF formula expressing this principle, LOP_n, uses variables x_{ij}, for $i \ne j \in [n]$, and consists of the clauses:

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

The idea is that the variables describe an ordering of $[n]$ and that x_{ij} holds when i is below j in the ordering. Thus, the first three groups of clauses correspond respectively to linearity, antisymmetry, and transitivity. The final group consists of wide clauses expressing that there is no maximal element.

First we consider the graph version of this principle, $\mathrm{GOP}(G)$, introduced in [33]. For this we use a slightly different encoding of an ordering into propositional variables, since we will use a degree lower bound for PCR from [22] that works with this new encoding. But we will show that our result transfers back to LOP_n as written above. 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 $i < j \in [n]$, with the role of x_{ji} played by $\bar{x}_{ij}$. $\mathrm{GOP}(G)$ is defined as the conjunction of the following clauses:

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

Note that because of the different choice of variables, the linearity and antisymmetry axioms are not needed, and the transitivity axioms have turned into two groups of axioms together asserting that there is no 3-cycle in the relation described by the variables.

Theorem 7.1. *There are simple undirected constant-degree graphs G over n nodes such that refuting* $\mathrm{GOP}(G)$ *requires PCR space $\Omega(\sqrt{n})$ over any field.*

Proof. It was proved in [22] that there is a family $\mathcal{G}_n$ of simple constant-degree graphs over n nodes such that for any $G \in \mathcal{G}_n$ refuting (the polynomial translation of) $\mathrm{GOP}(G)$ in PCR over any field requires degree $\Omega(n)$. The result follows using our main Theorem 6.4. $\square$

We can also lift the lower bound to LOP_n.

Corollary 7.2. *Over any field, refuting* LOP_n *requires PCR space $\Omega(\sqrt{n})$.*

Proof. (sketch) Let $G = ([n], E)$ be as in Theorem 7.1. Consider the following substitution ρ, which maps literals of LOP_n to literals of $\mathrm{GOP}(G)$. For $i < j$, ρ maps $x_{ji} \mapsto \bar{x}_{ij}$ and $\bar{x}_{ji} \mapsto x_{ij}$, and ρ is otherwise the identity. It is not difficult to see that after applying ρ, the linearity and antisymmetry axioms of LOP_n become tautologies of the form $x_{ij} \vee \bar{x}_{ij}$, the transitivity axioms of LOP_n become transitivity axioms of $\mathrm{GOP}(G)$, and the wide clauses of LOP_n become derivable from the corresponding axioms of $\mathrm{GOP}(G)$ by weakening the disjunction "some neighbour of i in G is above i in the ordering" to "some element of G is above i in the ordering".

Now 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 to the whole refutation. To turn this into a valid PCR refutation of $\mathrm{GOP}(G)$, whenever the original refutation downloaded a linearity or antisymmetry axiom of LOP_n, we now need to derive the monomial $\bar{x}_{ij}x_{ij}$ (recall that $\bar{x}_{ij}x_{ij}$ is the translation of the tautology $x_{ij} \vee \bar{x}_{ij}$ into the algebraic syntax of PCR). This derivation is possible in a constant amount of space, which we can re-use for every such axiom. 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 version of weakening appropriate for PCR – we repeatedly multiply the monomial by a single variable and immediately delete the old monomial, keeping only the result of multiplication. Again this process takes a constant amount of space, which can be re-used. For transitivity axioms, there is nothing to change.

It is not difficult to see that the result is a valid proof of $\mathrm{GOP}(G)$ of space $s + O(1)$. By Theorem 7.1, this completes the argument. $\square$

7.1.2 Functional pigeonhole principle

The *functional pigeonhole principle* FPHP^m_n, for $m > n$, expresses 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} for $i \in [m]$ and $j \in [n]$, is the following:

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

The variable x_{ij} stands for "pigeon i goes to hole j". The first group of clauses asserts that the map taking pigeons to holes is total, while the last two groups assert respectively that it is injective and well-defined.

No nontrivial PCR space lower bounds for FPHP^m_n were previously known, and, as proved in [17], the framework for obtaining lower bounds developed in [10] could not be used in this case.

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

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

Width lower bounds of $\Omega(n)$ for eFPHP^m_n in resolution can be easily obtained by modifying a routine Prover-Adversary argument proving a width lower bound for FPHP^m_n [3]. Hence Theorem 6.4 implies lower bounds of $\Omega(\sqrt{n})$ on the space needed to refute eFPHP^m_n in PCR. The functional pigeonhole principle is an example of formula which is *weight-constrained* in the terminology of [19] (see Definition 7.1 in [19]). As such it was shown in [19, Theorem 1.5] that the PCR space needed to refute FPHP^m_n and eFPHP^m_n can differ by at most a constant factor. Hence Theorem 6.4 implies PCR space lower bounds for FPHP^m_n as well.

Corollary 7.3. *Over any field, refuting* FPHP^m_n *in PCR requires space* $\Omega(\sqrt{n})$.

A different constant-width version of the functional pigeonhole principle is the functional pigeonhole principle over constant-degree bipartite graphs G, as defined in [26]. 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 $u \in U$, $v \in \Gamma(u)$, 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 7.4. ([26, Definition 4.1]) A bipartite graph $G = (U,V,E)$ is an (r,c)-*boundary expander* if for each $U' \subseteq U$ with $|U'| \le r$, it holds that $|\partial(U')| \ge c|U'|$, where the *boundary* $\partial(U')$ of U' is $\{v \in V \,:\, |\Gamma(v) \cap U'| = 1\}$.

Theorem 7.5. *([26, Theorem 4.9]) Let $G = (U,V,E)$ be a bipartite graph which is an (r,c)-boundary expander with left-degree bounded by d. Refuting* $\mathrm{FPHP}(G)$ *in PCR over any field requires degree strictly greater than $cr/2d$.*

Hence $\mathrm{FPHP}(G)$ also requires width $cr/2d$ in resolution. From Theorem 6.4 we conclude:

Theorem 7.6. *Let $G = (U,V,E)$ be a bipartite graph which is an (r,c)-boundary expander with left-degree bounded by d. Refuting $\mathrm{FPHP}(G)$ in PCR requires space $\Omega(\sqrt{cr/d})$.*

Since, as mentioned in [26], there exist bipartite graphs with $|U| = n+1$, $|V| = n$ and with left-degree 3 which are $(\gamma n, c)$-boundary expanders for $\gamma, c > 0$, we can conclude:

Corollary 7.7. *There exist constant left-degree bipartite graphs G with $|U| = n+1$ and $|V| = n$ such that refuting $\mathrm{FPHP}(G)$ in PCR requires space $\Omega(\sqrt{n})$.*

Note that this gives an alternative proof of the $m = n+1$ case of Corollary 7.3 above, since this $\mathrm{FPHP}(G)$ is a restriction of FPHP_n^{n+1}.

7.2 Separations independent of characteristic

Showing a *separation* between two measures means finding a family of formulas which has small proofs by one measure but requires large proofs by the other. We use this notion rather informally, not least because "small" and "large" mean different things for different measures.

In [17], a separation of size and degree from space was proved for PCR, namely that 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 independently of the characteristic of the field.

Theorem 6.4, together with some earlier results, makes it possible to prove characteristic-independent separations of PCR space from other measures of proof complexity. However, it has to be noted that, due to the quadratic term in the statement of Theorem 6.4, the lower bounds on space we obtain can be no better than $\Omega(\sqrt{n})$, where n is the size of the formula; they are not as strong as the $\Omega(n)$ lower bounds obtained in the separations of size and degree from space from [17].

7.2.1 Separation of size from space

Theorem 6.4, the degree lower bound of [22] (which holds for any field), and the polynomial size resolution proofs for $\mathrm{GOP}(G)$ (see [22]) immediately give a separation of PCR size and space independent of characteristic for $\mathrm{GOP}(G)$. We write n for the number of vertices in G.

Theorem 7.8. *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 [22], 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 (or functional onto) pigeonhole principle*, which asserts that there cannot exist a *bijection* between m pigeons and n holes (assuming $m \neq n$). The formula bij-PHP_n^m itself is obtained from the functional pigeonhole principle FPHP^m_n by adding clauses saying that each hole is occupied by a pigeon. Thus, bij-PHP^m_n consists of the clauses:

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

For a bipartite graph $G = (U,V,E)$, the formula bij-PHP(G) is, as in previous examples, obtained by restricting bij-PHP^m_n for the appropriate m,n to variables x_{uv} for $u \in U, v \in \Gamma(u)$. In other words, bij-PHP(G) contains the clauses:

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

This is sometimes called the *perfect matching principle*, PMP(G).

Theorem 7.9. *For every n, there exists a bipartite graph G with $|U| = n+1$, $|V| = n$ such that the formula* bij-PHP(G) *has size $O(n)$ and has a polynomial-size, constant-degree PCR refutation over any field, but requires space $\Omega(\sqrt{n})$ to refute in PCR.*

Proof. Using suitable boundary expanders, it is shown in [24, Section 4] that for every n, there exists a bounded-degree bipartite graph $G = (U,V,E)$ with $|U| = n+1$, $|V| = n$ such that refuting bij-PHP(G) in resolution requires width $\Omega(n)$. Fix such a graph G. Due to the fixed bound on the degree, bij-PHP(G) is a constant-width CNF of size $O(n)$. It follows from Theorem 6.4 and the width lower bound that refuting bij-PHP(G) in PCR requires space $\Omega(\sqrt{n})$.

To prove the existence of the polynomial-size, constant-degree refutations of bij-PHP(G), consider the version of the bijective pigeonhole principle in which the variables are x_{uv} for all $u \in U, v \in V$, but 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:

$$\begin{cases} 1 - \sum_{u\in V} x_{uv} & u \in U \\ 1 - \sum_{v\in U} x_{uv} & v \in V. \end{cases}$$

It is well-known that over any field this sum version of the bijective pigeonhole principle has a polynomial-size, constant-degree PC refutation [32]. The idea is that adding up the axioms pigeon-by-pigeon gives $\sum_{uv} x_{uv} = n+1$, and adding them hole-by-hole gives $\sum_{uv} x_{uv} = n$. This implies $1 = 0$ over any field. Of course, this refutation still works if we substitute 0 for each x_{uv} with $(u,v) \notin E$, which gives us a polynomial-size, constant-degree refutation of the "sum version" of bij-PHP(G).

It remains to argue that there is a polynomial-size, constant-degree PCR derivation of the "sum version" of bij-PHP(G) from bij-PHP(G) itself. In fact, for each fixed u we can derive $1 - \sum_{v\in\Gamma(u)} x_{uv}$ from bij-PHP(G) in constant size. First, use the axiom $\prod_{v\in\Gamma(u)} \bar{x}_{uv}$ to derive $\prod_{v\in\Gamma(u)}(1 - x_{uv})$. Then, kill off each term of degree at least 2 using the axioms $x_{uv}x_{uv'}$ and, if necessary, multiplications. This leaves $1 - \sum_{v\in\Gamma(u)} x_{uv}$. An analogous argument works to derive $1 - \sum_{u\in\Gamma(v)} x_{uv}$, for fixed v. □

7.3 Space lower bounds for Tseitin formulas over expanders

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 expressing that the parity of edges incident with v is exactly $\chi(v)$, that is, that $\bigoplus_{v \in e} x_e = \chi(v)$. 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 at most $|V| 2^{d-1}$.

In [17], lower bounds on the PCR space needed to refute $\mathrm{Ts}(G, \chi)$ for some G are proved using the following notion of graph expansion, introduced in [1].

Definition 7.10. The *connectivity expansion* $c(G)$ of a graph $G = (V, E)$ is the largest c such that for every $E' \subseteq E$ with $|E'| \le c$, the graph $G' = (V, E \setminus E')$ has a connected component of size strictly greater than $|V|/2$.

Theorem 7.11. ([17]) *Let $G = (V, E)$ be a connected graph of degree bounded by d such that E can be partitioned into cycles of length at most b. Let χ be an odd-charging of G. Then, over any field, refuting* $\mathrm{Ts}(G, \chi)$ *in PCR requires space at least $c(G)/4b - d/8$.*

In [17], Theorem 7.11 is used to show that if $d \ge 4$, then with high probability refuting $\mathrm{Ts}(G, \chi)$ for a random d-regular G with n nodes requires PCR space $\Omega(\sqrt{n})$. This involves showing that, for a suitable model of random bounded-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 [17] raise the question of whether PCR space lower bounds for Tseitin formulas can be proved using expansion alone.

We are able to answer this question positively, albeit using a different notion of expansion than in [17]. We use a theorem from [7]: for a suitable notion of expansion $e(G)$ (see [7] for a precise definition), refuting $\mathrm{Ts}(G, \chi)$ for a connected graph G requires resolution width $e(G)$. Theorem 6.4 thus gives:

Theorem 7.12. *Let $G = (V, E)$ be connected, and let χ be an odd-charging of G. Then, over any field, refuting* $\mathrm{Ts}(G, \chi)$ *in PCR requires space* $\Omega(\sqrt{e(G)})$.

It is known that for arbitrarily large n, there are constant-degree graphs G with n nodes and expansion $e(G) = \Omega(n)$ (for a discussion, refer e.g. to [34]).

Finally we remark that recently P. Austrin and K. Risse in [4] proved a $\Omega(n/\log n)$ lower bound for the space of refuting $\mathrm{Ts}(G, \chi)$ in PCR, for certain random constant-degree graphs G.

8 Open problems

A natural question is whether older PCR space lower bounds can be reproved (or extended) in our framework. For example, [10] defines an *m-winning strategy*, which is something like a more elaborate Atserias-Dalmau family, and shows that a CNF with such a strategy requires PCR space linear in m; can this be reproved using the methods of this paper? These older 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 proved in this paper are already 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 $O(\sqrt{w})$.

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 our results and older PCR space lower bound techniques, and to Jakob Nordström for helpful comments on an earlier version of this paper, in particular for suggesting bij-PHP(G) as a replacement for the more complicated formula we originally used to give a characteristic-independent separation of size and degree from space. We also thank Jakob for very kindly helping us with the conference presentation of [20], which we missed due to last-minute health problems of the speaker.

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. [doi:10.1137/S0097539700366735] 2, 3, 6, 8, 12, 23

[2] MICHAEL ALEKHNOVICH AND ALEXANDER A. RAZBOROV: Lower bounds for polynomial calculus: Non-binomial case. In *Proceedings of FOCS 2001*, pp. 190–199, 2001. [doi:10.1109/SFCS.2001.959893] 2

[3] ALBERT ATSERIAS AND VÍCTOR DALMAU: A combinatorial characterization of resolution width. *J. Comput. Syst. Sci.*, 74(3):323–334, 2008. [doi:10.1016/j.jcss.2007.06.025] 2, 3, 4, 5, 8, 9, 20, 24

[4] PER AUSTRIN AND KILIAN RISSE: Perfect matching in random graphs is as hard as Tseitin. In *Proceedings of SODA 2022*, pp. 979–1012, 2022. [doi:10.1137/1.9781611977073.43] 23

[5] PAUL BEAME, RUSSELL IMPAGLIAZZO, JAN KRAJÍČEK, TONIANN PITASSI, PAVEL PUDLÁK, AND ALAN WOODS: Exponential lower bounds for the pigeonhole principle. In *Proceedings of STOC 1992*, pp. 200–220, 1992. [doi:10.1145/129712.129733] 10

[6] ELI BEN-SASSON AND NICOLA GALESI: Space complexity of random formulae in resolution. *Random Struct. Algorithms*, 23(1):92–109, 2003. [doi:10.1002/rsa.10089] 2, 12

[7] ELI BEN-SASSON AND AVI WIGDERSON: Short proofs are narrow – resolution made simple. *J. ACM*, 48(2):149–169, 2001. [doi:10.1145/375827.375835] 2, 4, 23

[8] CHRISTOPH BERKHOLZ AND JAKOB NORDSTRÖM: Supercritical space-width trade-offs for resolution. In *Proceedings of ICALP 2016*, pp. 57:1–57:14, 2016. [doi:10.4230/LIPIcs.ICALP.2016.57] 2, 3

[9] ILARIO BONACINA: Total space in resolution is at least width squared. In *Proceedings of ICALP 2016*, pp. 56:1–56:13, 2016. [doi:10.4230/LIPIcs.ICALP.2016.56] 3, 5, 10

[10] ILARIO BONACINA AND NICOLA GALESI: A framework for space complexity in algebraic proof systems. *J. ACM*, 62(3):23:1–23:20, 2015. [doi:10.1145/2699438] 2, 3, 12, 18, 20, 23

[11] ILARIO BONACINA, NICOLA GALESI, AND NEIL THAPEN: Total space in resolution. *SIAM J. Comput.*, 45(5):1894–1909, 2016. [doi:10.1137/15M1023269] 5

[12] MARIA LUISA BONET AND NICOLA GALESI: Optimality of size-width tradeoffs for resolution. *Computational Complexity*, 10(4):261–276, 2001. [doi:10.1007/s000370100000] 4

[13] SAM BUSS: Axiomatizations and conservation results for fragments of bounded arithmetic. In WILFRIED SIEG, editor, *Logic and Computation*, volume 106 of *Contemporary Mathematics*, pp. 57–84, 1990. [doi:10.1090/conm/106/1057816] 5

[14] MATTHEW CLEGG, JEFF EDMONDS, AND RUSSELL IMPAGLIAZZO: Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proceedings of STOC 1996*, pp. 174–183, 1996. [doi:10.1145/237814.237860] 2, 6

[15] STEPHEN A. COOK AND ROBERT A. RECKHOW: The relative efficiency of propositional proof systems. *J. Symb. Log.*, 44(1):36–50, 1979. [doi:10.2307/2273702] 2

[16] JUAN LUIS ESTEBAN AND JACOBO TORÁN: Space bounds for resolution. *Inf. Comput.*, 171(1):84–97, 2001. [doi:10.1006/inco.2001.2921] 2

[17] 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 *Proceedings of ICALP 2013*, pp. 437–448, 2013. [doi:10.1007/978-3-642-39206-1_37] 2, 3, 4, 18, 20, 21, 23

[18] YUVAL FILMUS, MASSIMO LAURIA, MLADEN MIKŠA, JAKOB NORDSTRÖM, AND MARC VINYALS: From small space to small width in resolution. *ACM Trans. Comput. Log.*, 16(4):28:1–28:15, 2015. [doi:10.1145/2746339] 3, 4

[19] 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. [doi:10.1137/120895950] 2, 3, 20

[20] NICOLA GALESI, LESZEK ALEKSANDER KOLODZIEJCZYK, AND NEIL THAPEN: Polynomial calculus space and resolution width. In *Proceedings of FOCS 2019*, pp. 1325–1337, 2019. [doi:10.1109/FOCS.2019.00081] 1, 10, 24

[21] NICOLA GALESI AND MASSIMO LAURIA: On the automatizability of polynomial calculus. *Theory Comput. Syst.*, 47(2):491–506, 2010. [doi:10.1007/s00224-009-9195-5] 2

[22] NICOLA GALESI AND MASSIMO LAURIA: Optimality of size-degree tradeoffs for polynomial calculus. *ACM Trans. Comput. Log.*, 12(1):4:1–4:22, 2010. [doi:10.1145/1838552.1838556] 2, 4, 18, 19, 21

[23] 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. [doi:10.1007/s000370050024] 2

[24] DMITRY ITSYKSON, VSEVOLOD OPARIN, MIKHAIL SLABODKIN, AND DMITRY SOKOLOV: Tight lower bounds on the resolution complexity of perfect matching principles. *Fundam. Inform.*, 145(3):229–242, 2016. [doi:10.3233/FI-2016-1358] 22

[25] LESZEK ALEKSANDER KOŁODZIEJCZYK AND NEIL THAPEN: Approximate counting and NP search problems. *J. Math. Log.*, 2022. To appear. arXiv:1812.10771. 10

[26] MLADEN MIKŠA AND JAKOB NORDSTRÖM: A generalized method for proving polynomial calculus degree lower bounds. In *Proceedings of CCC 2015*, pp. 467–487, 2015. [doi:10.4230/LIPIcs.CCC.2015.467] 2, 4, 20, 21

[27] JAKOB NORDSTRÖM: On the interplay between proof complexity and SAT solving. *ACM SIGLOG News*, 2(3):19–44, 2015. [doi:doi.org/10.1145/2815493.2815497] 3

[28] PAVEL PUDLÁK AND NEIL THAPEN: Random resolution refutations. *Computational Complexity*, 28(2):185–239, 2019. [doi:10.1007/s00037-019-00182-7] 10

[29] ALEXANDER A. RAZBOROV: Lower bounds for the polynomial calculus. *Computational Complexity*, 7(4):291–324, 1998. [doi:10.1007/s000370050013] 2

[30] ALEXANDER A. RAZBOROV: Proof complexity of pigeonhole principles. In *Developments in Language Theory, 5th International Conference, DLT 2001*, pp. 100–116, 2001. [doi:10.1007/3-540-46011-X_8] 4

[31] ALEXANDER A. RAZBOROV: Resolution lower bounds for the weak functional pigeonhole principle. *Theor. Comput. Sci.*, 303(1):233–243, 2003. [doi:10.1016/S0304-3975(02)00453-X] 4

[32] SØREN RIIS: *Independence in Bounded Arithmetic*. Ph. D. thesis, Oxford University, 1993. 4, 22

[33] 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. [doi:10.1137/S0097539703428555] 4, 18

[34] ALASDAIR URQUHART: Hard examples for resolution. *J. ACM*, 34(1):209–219, 1987. [doi:10.1145/7531.8928] 23

AUTHORS

Nicola Galesi
Department of Computer, Control and Management Engineering
Sapienza University Rome
nicola.galesi@uniroma1.it
http://wwwusers.di.uniroma1.it/~galesi/

Leszek A. Kołodziejczyk
Institute of Mathematics
University of Warsaw
lak@mimuw.edu.pl
https://www.mimuw.edu.pl/~lak/

Neil Thapen
Institute of Mathematics
Czech Academy of Sciences
thapen@math.cas.cz
https://users.math.cas.cz/~thapen/

ABOUT THE AUTHORS

NICOLA GALESI was born in Bari, the capital of Puglia, a beautiful region in the south of Italy. He has one daughter, Eva. He holds a PhD in Computer Science with focus on Complexity Theory, obtained in 2000 under the supervision of Maria Luisa Bonet at the Universitat Politecnica de Catalunya, in Barcelona. Nicola's main research interests were always in proof complexity, studying the limits of feasible proofs and their relation with the limits of feasible computations. In 2022 he started a new academic position focused on mathematics with the Department of Computer, Control and Management Engineering "Antonio Ruberti" in Sapienza. Nicola loves reading novels, listening to all kinds of live music from classical to rock and practicing tai chi and table tennis.

LESZEK KOŁODZIEJCZYK received his PhD in philosophy in 2005, based on a thesis in finite model theory written under the supervision of Marcin Mostowski. Since then, he has been working at the Institute of Mathematics of the University of Warsaw, with breaks for postdocs in Prague and San Diego. He is a mathematical logician with eclectic research interests that include bounded arithmetic, nonstandard models of arithmetic, and reverse mathematics. When not working, he enjoys reading, walking in the woods, and spending time with his family.

NEIL THAPEN received his doctorate in 2002 from the University of Oxford, where his supervisor was Alex Wilkie. He works in mathematical logic, in particular on bounded arithmetic and related things in proof complexity, and sometimes on games. He has been at the Institute of Mathematics in Prague since 2005.