

Space proof complexity for random 3-CNFs

Patrick Bennett* Ilario Bonacina† Nicola Galesi† Tony Huynh†‡
Mike Molloy* Paul Wollan†‡

Abstract

We investigate the space complexity of refuting 3-CNFs in Resolution and algebraic systems. We prove that every *Polynomial Calculus with Resolution* refutation of a random 3-CNF φ in n variables requires, with high probability, $\Omega(n)$ *distinct monomials* to be kept simultaneously in memory. The same construction also proves that every *Resolution* refutation φ requires, with high probability, $\Omega(n)$ clauses each of width $\Omega(n)$ to be kept at the same time in memory. This gives a $\Omega(n^2)$ lower bound for the *total space* needed in Resolution to refute φ . These results are best possible (up to a constant factor) and answer questions about space complexity of 3-CNFs posed in [FLN⁺12, FLM⁺13, BGT14, BG].

The main technical innovation is a variant of *Hall's Lemma*. We show that in bipartite graphs G with bipartition (L, R) and left-degree at most 3, L can be covered by certain families of disjoint paths, called *VW-matchings*, provided that L *expands* in R by a factor of $(2 - \epsilon)$, for $\epsilon < \frac{1}{23}$.

*Computer Science Department, University of Toronto, 10 Kings College Road, M5S 3G4 Toronto, Canada, {patrickb, molloy}@cs.toronto.edu.

†Computer Science Department, Sapienza University of Rome, via Salaria 113, 00198 Rome, Italy, {bonacina, galesi, huynh, wollan}@di.uniroma1.it.

‡Supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement no. 279558.

1 Introduction

During the last decade, an active line of research in proof complexity has been the space complexity of proofs and how space is related to other complexity measures (like size, length, width, degree) [ET01, ABSRW02, BSG03, Ben02, AD08, BN08, Nor09, BN11, FLN⁺12, FLM⁺13, BGT14, BG]. This investigation has raised several important foundational questions. Some of these have been solved, while several others are still open and challenging (see [Nor13] for a survey on this topic). Space of proofs concerns the minimal memory occupation of algorithms verifying the correctness of proofs in concrete propositional proof systems, and is thus also relevant in more applied algorithmic contexts. For instance, a major problem in state of the art SAT-solvers is memory consumption. In proof complexity, this resource is modeled by proof space. It is well-known that SAT-solvers used in practice (like CDCL) are based on low-level proof systems such as *Resolution*.

In this work we focus on two well known proof systems that play a central role in proof complexity: *Resolution* [Rob65, Bla37] and *Polynomial Calculus* [CEI96]. Resolution (RES) is a refutational proof system for unsatisfiable propositional CNF formulas using only one logical rule: $\frac{A \vee x \quad \neg x \vee B}{A \vee B}$. Polynomial calculus is an algebraic refutational proof system for unsatisfiable sets of polynomials (over $\{0, 1\}$ solutions) based on two rules: *linear combination* of polynomials and *multiplication* by variables. In this article, we consider the stronger system *Polynomial Calculus with Resolution* (PCR) which extends both Resolution and Polynomial Calculus [ABSRW02].

Several different measures for proof space were investigated for these two systems [ET01, ABSRW02, Ben02, AD08, BN08, Nor09, BN11, FLN⁺12, BGT14, BG]. In this work we focus on *total space* (for RES), which is the maximum number of variables (counted with repetitions) to be kept simultaneously in memory while verifying a proof; and *monomial space* (for PCR), which is the maximum number of distinct monomials to be kept simultaneously in memory while verifying a proof. Both measures were introduced in [ABSRW02], where some preliminary lower and upper bounds were given. In particular, for every unsatisfiable CNF in n variables, there is an easy upper bound of $O(n)$ for monomial space in PCR and $O(n^2)$ for total space in RES.

Major open problems about these two measures were solved only recently in [FLM⁺13, BG, BGT14]. In particular, [BG, BGT14] prove that, for $r \geq 4$, random r -CNFs over n variables require $\Theta(n^2)$ total space in resolution and $\Theta(n)$ monomial space in PCR. However, it is not at all obvious how to generalize the techniques in [BG, BGT14] to handle 3-CNFs. Indeed, it is an open problem whether there is any family of 3-CNFs requiring large *total space* (in RES) and *monomial space* (in PCR). In this work, we resolve this problem by proving that random 3-CNFs also require $\Theta(n^2)$ *total space* (in RES) and $\Theta(n)$ *monomial space* (in PCR).

Results. Let φ be a random 3-CNF in n variables. We prove that every PCR refutation of φ requires, with high probability, $\Omega(n)$ distinct monomials to be kept simultaneously in memory (Theorem 5.3). Moreover, every RES refutation of φ has, with high probability, $\Omega(n)$ clauses each of width $\Omega(n)$ to be kept at the same time in memory (Theorem 5.3). This gives a $\Omega(n^2)$ lower bound for the total space of every RES refutation of φ . These results resolve questions about space complexity of 3-CNFs mentioned in [FLM⁺13, BGT14, BG, FLN⁺12].

Both results follow using the framework proposed in [BG], where the construction of suitable families of assignments called *k-winning strategies* (Definition 2.1) leads to monomial space lower bounds in PCR (Theorem 2.2). This construction is made possible by a modification of Hall's Lemma [Hal35] for matchings to VW-matchings (Lemma 1.2).

Definition 1.1 (VW-matching). *Let G be a bipartite graph with bipartition (L, R) . A VW-matching in G is a subgraph F of G such that each connected component of F is a path with at most 4 edges and*

both endpoints in R . A VW-matching F covers a set of vertices S if $S \subseteq V(F)$. Define $L(F) = V(F) \cap L$ and $R(F) = V(F) \cap R$.

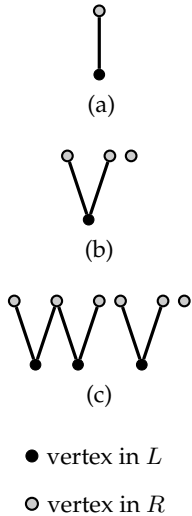


Figure 1.1

Figure 1.1 compare *matchings* (Figure 1.1.(a)), *2-matchings* as used in [BGT14, BG] (Figure 1.1.(b)) and *VW-matchings* (Figure 1.1.(c)). Note that for technical reasons, we allow 2-matchings and VW-matchings to contain isolated vertices from R . We can now state our variant of Hall’s Lemma. This lemma and its proof are independent from the proof complexity results and might be useful in other contexts.

Lemma 1.2 ($(2 - \epsilon)$ -Hall’s Lemma). *Let $\epsilon < \frac{1}{23}$. Let G be a bipartite graph with bipartition (L, R) such that each vertex in L has degree at most 3 and no pair of degree 3 vertices in L have the same set of neighbors. If $|N_G(L)| \geq (2 - \epsilon)|L|$, and each proper subset of L can be covered by a VW-matching, then L can be covered by a VW-matching.*

Note that the converse of Lemma 1.2 does not hold (unlike in Hall’s Lemma).

Outline of the paper. Section 2 contains some preliminary notions about proof complexity. In particular, the formal definitions of Resolution and Polynomial Calculus with Resolution, the model of space (based on [ET01, ABSRW02]) and the formal definition of total space and monomial space. We

present a simplified (but less general) version of the k -winning strategies of [BG] (Definition 2.1). These k -winning strategies were used in [BG] to prove monomial space lower bounds for PCR. Here we use the same k -winning strategies also to prove total space lower bounds for RES. For the connections with [BGT14], see Appendix A.

In Section 3, we present the proof of our version of the $(2 - \epsilon)$ -Hall’s Lemma (Lemma 1.2). This proof relies on a concentration result on the average right-degree and a discharging argument. We also prove a bound for the best possible value of ϵ for which Lemma 1.2 could hold and conjecture that this bound is in fact the optimal value of ϵ (Proposition 3.1).

In Section 4, we define a two player covering game *CoverGame*, whose aim is to dynamically build a VW-matching inside a fixed bipartite graph G (Definition 4.1). Informally, a player, *Choose*, queries nodes in the graph G and the other player, *Cover*, attempts to extend the current VW-matching to also cover the node queried (if not already covered). The main result of Section 4 is Theorem 4.3, where we prove that if the graph G has large left-expansion (i.e. large enough to apply Lemma 1.2 to sufficiently large subgraphs of G), then there is a winning strategy for *Cover* to force *Choose* to query a very large portion of the graph G . In the analysis of the game, we use the $(2 - \epsilon)$ -Hall’s Lemma and VW-matchings in a similar manner to how matchings and 2-matchings were used in RES and PCR [BSG03, Ats04, BGT14, BG]. A key difference is that we are looking for winning strategies of *Cover* for the *CoverGame* only on graphs G where the number of high degree vertices is suitably bounded (Theorem 4.3). This additional information allows us to identify a VW-matching covering all such high degree vertices in G but preserving expansion properties of the remaining graph. *Cover* will use this additional information to obtain a winning strategy. The full proofs of the technical Lemmas of this section are in Appendix B.

In Section 5, we prove (Lemma 5.1), that if *Cover* wins *CoverGame* on the adjacency graph of a CNF φ (see Section 2 for the definition of adjacency graph) guaranteeing VW-matchings of maximal size μ , then there exists a μ -winning strategy for the polynomial encoding of φ . Finally, the monomial space in PCR and the total space in RES for random 3-CNFs (Theorem 5.3) follow from

well-known results about expansion of its adjacency graph [CS88, BP96, BSW01, BSG03]. In order to get optimal lower bounds, we show in Lemma 5.2 (with proof in Appendix D) that the number of variables appearing in many clauses of a random CNF is w.h.p. suitably bounded as required in the conditions of Theorem 4.3.

2 Preliminaries

Let X be a set of variables. A **literal** is a boolean constant, 0 or 1, or a variable $x \in X$, or the negation $\neg x$ of a variable x . A **clause** is a disjunction of literals: $C = (\ell_1 \vee \dots \vee \ell_k)$. The **width** of a clause is the number of literals in it. A formula φ is in **Conjunctive Normal Form** (CNF) if $\varphi = C_1 \wedge \dots \wedge C_m$ where C_i are clauses. It is a k -CNF if each C_i contains at most k literals. Let φ be a CNF and X be the set of variables appearing in φ . The **adjacency graph** of φ is a bipartite graph G_φ with bipartition (L, R) such that L is the set of clauses of φ , $R = X$, and $(C, x) \in E$ if and only if x or $\neg x$ appears in C . If φ is a k -CNF, then G_φ has left-degree at most k .

Resolution (RES) [Bla37, Rob65] is a propositional proof system for refuting unsatisfiable CNFs. Starting from an unsatisfiable CNF φ , RES allows us to derive the empty clause \perp using the following inference rule:

$$\frac{C \vee x \quad D \vee \neg x}{C \vee D}.$$

Following [ABSRW02], we define $\overline{X} = \{\bar{x} : x \in X\}$, which we regard as a set of formal variables with the intended meaning of \bar{x} as $\neg x$. Given a field \mathbb{F} , the ring $\mathbb{F}[X, \overline{X}]$ is the ring of polynomials in the variables $X \cup \overline{X}$ with coefficients in \mathbb{F} . We use the following **standard encoding** (tr) of CNF formulas over X into a set of polynomials in $\mathbb{F}[X, \overline{X}]$: $tr(\varphi) = \{tr(C) : C \in \varphi\} \cup \{x^2 - x, x + \bar{x} - 1 : x \in X\}$, where

$$tr(x) = \bar{x}, \quad tr(\neg x) = x, \quad tr\left(\bigvee_{i=1}^n \ell_i\right) = \prod_{i=1}^n tr(\ell_i).$$

A set of polynomials P in $\mathbb{F}[X]$ is **contradictory** if and only if 1 is in the ideal generated by P . Notice that a CNF φ is unsatisfiable if and only if $tr(\varphi)$ is a contradictory set of polynomials.

Polynomial Calculus with Resolution (PCR) [ABSRW02] is an algebraic proof system for polynomials in $\mathbb{F}[X, \overline{X}]$. Starting from an initial set of contradictory polynomials P in $\mathbb{F}[X, \overline{X}]$, PCR allows us to derive the polynomial 1 using the following inference rules: for all $p, q \in \mathbb{F}[X, \overline{X}]$

$$\frac{p \quad q}{\alpha p + \beta q} \quad \forall \alpha, \beta \in \mathbb{F}, \quad \frac{p}{vp} \quad \forall v \in X \cup \overline{X}.$$

To force 0/1 solutions, we always include the **boolean axioms** $\{x^2 - x, x + \bar{x} - 1\}_{x \in X}$ among the initial polynomials, as in the case of the polynomial encoding of CNFs.

In order to study space of proofs we follow a model inspired by the definition of space complexity for Turing machines, where a machine is given a read-only input tape from which it can download parts of the input to the working memory as needed [ET01].

Given an unsatisfiable CNF formula φ , a RES (**resp.** PCR) **refutation** of φ is a sequence $\Pi = \langle \mathfrak{M}_0, \dots, \mathfrak{M}_\ell \rangle$ of sets of clauses (resp. polynomials), called **memory configurations**, such that: $\mathfrak{M}_0 = \emptyset$, $\perp \in \mathfrak{M}_\ell$ (resp. $1 \in \mathfrak{M}_\ell$), and for all $i \leq \ell$, \mathfrak{M}_i is obtained by \mathfrak{M}_{i-1} by applying one of the following rules:

(AXIOM DOWNLOAD) $\mathfrak{M}_i = \mathfrak{M}_{i-1} \cup \{C\}$, where C is a clause of φ (resp. a polynomial of $tr(\varphi)$);

(INFERENCE ADDING) $\mathfrak{M}_i = \mathfrak{M}_{i-1} \cup \{O\}$, where O is inferred by the RES inference rule (resp. PCR inference rules) from clauses (resp. polynomials) in \mathfrak{M}_i ;
(ERASURE) $\mathfrak{M}_i \subset \mathfrak{M}_{i-1}$.

If in the definition of PCR refutation we substitute the INFERENCE ADDING rule with:

(SEMANTICAL INFERENCE) \mathfrak{M}_i is contained in the ideal generated by \mathfrak{M}_{i-1} in $\mathbb{F}[X, \bar{X}]$,

we have what is called a *semantical PCR refutation* of φ [ABSRW02].

The **total space** of Π is the maximum over i of the number of variables (counted with repetitions) occurring in \mathfrak{M}_i .

The **monomial space** of a PCR refutation Π , denoted by $\text{MSpace}(\Pi)$, is the maximum over i of the number of *distinct* monomials appearing in \mathfrak{M}_i .

2.1 Space lower bounds and k -winning strategies

A **partial assignment** over a set of variables X is a map $\alpha : X \rightarrow \{0, 1, \star\}$. The **domain** of α is $\text{dom}(\alpha) = \alpha^{-1}(\{0, 1\})$. Given a partial assignment α and a CNF φ we can apply α to φ , obtaining a new formula $\alpha(\varphi)$ in the standard way, i.e. substituting each variable x of φ in $\text{dom}(\alpha)$ with the value $\alpha(x)$ and then simplifying the result. We say that α *satisfies* φ , and we write $\alpha \models \varphi$, if $\alpha(\varphi) = 1$. Similarly, for a family F of partial assignments, $F \models \varphi$ means that for each $\alpha \in F$, $\alpha \models \varphi$.

For each partial assignment α over $X \cup \bar{X}$ we assume that it respects the intended meaning of the variables; that is, $\alpha(\bar{x}) = 1 - \alpha(x)$ for each $x, \bar{x} \in \text{dom}(\alpha)$. Given a partial assignment α and a polynomial p in $\mathbb{F}[X, \bar{X}]$, we can apply α to p , obtaining a new polynomial $\alpha(p)$ in the standard way, similarly as before. The notation $\alpha \models p$ means that $\alpha(p) = 0$. If F is a family of partial assignments and P a set of polynomials, we write $F \models P$ if $\alpha \models p$ for each $\alpha \in F$ and $p \in P$. Notice that if φ is a CNF and α is a partial assignment then $\alpha \models \varphi$ if and only if $\alpha \models \text{tr}(\varphi)$.

Let A be a family of partial assignments, and let $\text{dom}(A)$ be the union of the domains of the assignments in A . We say that a set of partial assignments A is **flippable** if and only if for all $x \in \text{dom}(A)$ there exist $\alpha, \beta \in A$ such that $\alpha(x) = 1 - \beta(x)$. Two families of partial assignments A and A' are **domain-disjoint** if $\text{dom}(\alpha)$ and $\text{dom}(\alpha')$ are disjoint for all $\alpha \in A$ and $\alpha' \in A'$. Given non-empty and pairwise domain-disjoint sets of assignments¹ H_1, \dots, H_t , the **product-family** $\mathcal{H} = H_1 \otimes \dots \otimes H_t$ is the following set of assignments

$$\mathcal{H} = H_1 \otimes \dots \otimes H_t = \{\alpha_1 \cup \dots \cup \alpha_t : \alpha_i \in H_i\},$$

or, if $t = 0$, $\mathcal{H} = \{\lambda\}$, where λ is the partial assignment of the empty domain. Note $\text{dom}(\mathcal{H}) = \bigcup_i \text{dom}(H_i)$. We call the H_i the **factors** of \mathcal{H} . For a product-family $\mathcal{H} = H_1 \otimes \dots \otimes H_t$, the **rank** of \mathcal{H} , denoted $\|\mathcal{H}\|$, is the number of factors of \mathcal{H} different from $\{\lambda\}$. We do not count $\{\lambda\}$ in the rank since $\mathcal{H} \otimes \{\lambda\} = \mathcal{H}$. Given two product-families \mathcal{H} and \mathcal{H}' , we write $\mathcal{H}' \sqsubseteq \mathcal{H}$ if and only if each factor of \mathcal{H}' different from $\{\lambda\}$ is also a factor of \mathcal{H} . In particular, $\{\lambda\} \sqsubseteq \mathcal{H}$ for every \mathcal{H} .

A family of flippable product-families is called a **strategy** and denoted by \mathcal{L} . We now present a definition of suitable families of flippable products: the *k -winning strategies* [BG].

Definition 2.1 (*k -winning strategy* [BG]). *Let P be a set of polynomials in the ring $\mathbb{F}[X, \bar{X}]$. A non-empty strategy \mathcal{L} is a **k -winning strategy** if and only if for every $\mathcal{H} \in \mathcal{L}$ the following conditions hold:*

¹We always suppose that the partial assignments are respecting the intended meaning of the variables in \bar{X} . That is, if $x \in \text{dom}(\alpha)$, then $\alpha(\bar{x}) = 1 - \alpha(x)$; hence a variable x is in $\text{dom}(H_i)$ if and only if \bar{x} is in $\text{dom}(H_i)$.

(RESTRICTION) for each $\mathcal{H}' \sqsubseteq \mathcal{H}$, $\mathcal{H}' \in \mathcal{L}$;

(EXTENSION) if $\|\mathcal{H}\| < k$, then for each $p \in P$ there exists a flippable product-family $\mathcal{H}' \in \mathcal{L}$ such that $\mathcal{H}' \supseteq \mathcal{H}$ and $\mathcal{H}' \models p$.

Notice that, by the restriction property, $\{\lambda\}$ is in every k -winning strategy.

Theorem 2.2. Let φ be an unsatisfiable CNF and $k \geq 1$ an integer. If there exists a non-empty k -winning strategy \mathcal{L} for $\text{tr}(\varphi)$, then for every semantical PCR refutation Π of φ , $\text{MSpace}(\Pi) \geq \frac{k}{4}$. Moreover any resolution refutation of φ must pass through a memory configuration containing at least $\frac{k-1}{2}$ clauses each of width at least $\frac{k-1}{2}$. In particular, the RES refutation requires total space at least $\frac{(k-1)^2}{4}$.

The monomial space lower bound follows directly from the main theorem of [BG]. In Appendix A we show how to use k -winning strategies to construct the combinatorial objects used in [BGT14] to obtain total space lower bounds.

3 A $(2 - \epsilon)$ -Hall's Lemma for VW-matchings

We now prove our variant of Hall's Lemma. This lemma may be of independent interest.

Restated Lemma 1.2 ($(2 - \epsilon)$ -Hall's Lemma). Let $\epsilon < \frac{1}{23}$. Let G be a bipartite graph with bipartition (L, R) such that each vertex in L has degree at most 3 and no pair of degree 3 vertices in L have the same set of neighbors. If $|N_G(L)| \geq (2 - \epsilon)|L|$, and each proper subset of L can be covered by a VW-matching, then L can be covered by a VW-matching.

Proof. Observe that each vertex v in L has degree 2 or 3, otherwise v could not be covered by a VW-matching. Similarly, no degree 2 vertices in L have the same neighbourhood.

By assumption, no degree 3 vertices have the same neighbourhood. Define the hypergraph $\mathcal{H} = (V, E)$ with $V = N_G(L)$ and $E = \{N_G(x) : x \in L\}$.

By the above observations, $N_G : L \rightarrow E$ is a bijection and $|V| \geq (2 - \epsilon)|L| = (2 - \epsilon)|E|$. The degree of a vertex v in \mathcal{H} , denoted $\text{deg}_{\mathcal{H}}(v)$, is the number of distinct hyperedges which contain v . Let $L' \subseteq L$ and let $E' = \{N_G(x) : x \in L'\}$. The existence of a VW-matching in G covering $L' \subseteq L$ is equivalent to the existence of an injective function $f : E' \rightarrow \{\{x, y\} : x, y \in N_G(L')\}$, which we call a **2-path cover** of E' , such that

1. for every $e \in E'$, $f(e)$ is a subset of size 2 of e ;
2. for each triple of distinct hyperedges $e_1, e_2, e_3 \in E'$, it is not the case that $f(e_i) \cap f(e_{i+1}) \neq \emptyset$ for $i = 1, 2$.

Observe that all the configurations of hyperedges shown in Figure 3.1 have a 2-path cover using only degree 1 and 2 vertices of \mathcal{H} . If any of these configurations appear in \mathcal{H} , we can by assumption find a 2-path cover f of the remaining hyperedges, and then extend f to a 2-path cover of \mathcal{H} . Therefore, we may assume that no configuration from Figure 3.1 appears in \mathcal{H} and we show that this assumption leads to a contradiction.

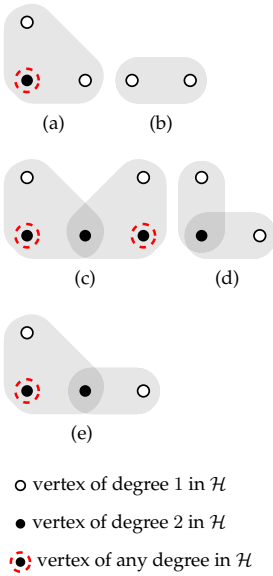


Figure 3.1: A set of reducible configurations for \mathcal{H}

Let d be the average degree in \mathcal{H} . Observe that $3|E| \geq \sum_{v \in V} \deg_{\mathcal{H}}(v) = d|V| \geq d(2 - \epsilon)|E|$, where the last inequality follows from the hypothesis that $|V| \geq (2 - \epsilon)|E|$. Thus, $d \leq \frac{3}{2 - \epsilon}$.

Let D be the set of vertices of \mathcal{H} of degree 1, and \mathcal{D} be the set of hyperedges which contain a vertex in D . Note that $|D| = |\mathcal{D}|$, since the configurations (a), (b) of Figure 3.1 do not appear in \mathcal{H} . We now prove a concentration result for $|D|$. An upper bound follows immediately from the following chain of inequalities

$$|D| = |\mathcal{D}| \leq |E| \leq \frac{1}{2 - \epsilon}|V|.$$

For the lower bound suppose $\frac{1 - 2\epsilon}{2 - \epsilon}|V| > |D|$. Then,

$$\begin{aligned} \frac{3}{2 - \epsilon}|V| &\geq d|V| = \sum_{v \in D} \deg_{\mathcal{H}}(v) + \sum_{v \in V \setminus D} \deg_{\mathcal{H}}(v) \\ &\geq |D| + 2|V \setminus D| > \frac{1 - 2\epsilon}{2 - \epsilon}|V| + 2\left(1 - \frac{1 - 2\epsilon}{2 - \epsilon}\right)|V| = \frac{3}{2 - \epsilon}|V|, \end{aligned}$$

which is a contradiction. Hence we have that $\frac{1 - 2\epsilon}{2 - \epsilon}|V| \leq |D| \leq \frac{1}{2 - \epsilon}|V|$.

We finish the proof with a discharging argument. Each vertex v of \mathcal{H} receives a charge of $\deg_{\mathcal{H}}(v)$. Let E_2 be the set of hyperedges in E of size 2 and \mathcal{D}_2 be $\mathcal{D} \cap E_2$. The edges in \mathcal{D}_2 receive a charge of -2, the edges in $\mathcal{D} \setminus \mathcal{D}_2$ receive charge -3. The edges not in \mathcal{D} receive no charge.

We now perform the following discharging rule. Each hyperedge e in \mathcal{D} gives a charge of -1 to each vertex in e . After discharging, every hyperedge has charge 0, and every vertex has non-negative charge. Let Z be the set of vertices with charge 0 after discharging. Observe that a vertex x is in Z if and only if every hyperedge containing x also contains a degree 1 vertex.

Let C denote the total charge. Then,

$$C = 3|E| - |E_2| - 3|\mathcal{D}| + |\mathcal{D}_2| \leq 3|E| - 3|D| \leq \frac{3}{2 - \epsilon}|V| - 3\frac{1 - 2\epsilon}{2 - \epsilon}|V| = \frac{6\epsilon}{2 - \epsilon}|V|.$$

It follows that $|Z| \geq \frac{2 - 7\epsilon}{2 - \epsilon}|V|$. Since $D \subseteq Z$ and $|D| \leq \frac{|V|}{2 - \epsilon}$, we conclude that $|Z \setminus D| \geq \frac{1 - 7\epsilon}{2 - \epsilon}|V|$. Because the configurations (c), (d), (e) of Figure 3.1 do not appear in \mathcal{H} , every vertex in $Z \setminus D$ has degree at least 3. Thus,

$$\begin{aligned} \frac{3}{2 - \epsilon}|V| &\geq d|V| \geq \sum_{v \in D} \deg_{\mathcal{H}}(v) + \sum_{v \in Z \setminus D} \deg_{\mathcal{H}}(v) \geq |D| + 3|Z \setminus D| \\ &\geq \frac{1 - 2\epsilon}{2 - \epsilon}|V| + 3\frac{1 - 7\epsilon}{2 - \epsilon}|V| = \frac{4 - 23\epsilon}{2 - \epsilon}|V|. \end{aligned}$$

Therefore, $3 \geq 4 - 23\epsilon$, which is a contradiction as $\epsilon < \frac{1}{23}$. \square

We end this section with a comment on the parameter ϵ . Since VW-matchings expand by a factor of at least $\frac{3}{2}$, we certainly require $\epsilon \leq \frac{1}{2}$ in the statement of Lemma 1.2. In Proposition 3.1, we show a stronger upper bound that $\epsilon \leq \frac{1}{3}$ is in fact necessary (see Appendix C for the proof).

Proposition 3.1. *For all $\epsilon > \frac{1}{3}$ there exists a bipartite graph G_ϵ with bipartition (L, R) such that each vertex in L has degree at most 3 and no pair of degree 3 vertices in L have the same set of neighbours. Moreover, $|N_{G_\epsilon}(L)| \geq (2 - \epsilon)|L|$ and each proper subset of L can be covered by a VW-matching but L cannot be covered by a VW-matching.*

We conjecture that Lemma 1.2 is true for $\epsilon \leq \frac{1}{3}$. Proposition 3.1 shows that this would be best possible.

4 A cover game over bipartite graphs

As an application, we use the previous result to build a winning strategy for a game played on bipartite graphs.

Definition 4.1 (Cover Game). *The **Cover Game** $\text{CoverGame}_{\text{VW}}(G, \mu)$ is a game between two players, Choose and Cover, on a bipartite graph G with bipartition (L, R) . At each step i of the game the players maintain a VW-matching F_i in G . At step $i + 1$ Choose can*

1. *remove a connected component from F_i , or*
2. *if the number of connected components of F_i is strictly less than μ , pick a vertex (either in L or R) and challenge Cover to find a VW-matching F_{i+1} in G such that*
 - (a) *F_{i+1} extends F_i . That is, each connected component of F_i is also a connected component of F_{i+1} ;*
 - (b) *F_{i+1} covers the vertex picked by Choose.*

Cover loses the game $\text{CoverGame}_{\text{VW}}(G, \mu)$ if at some point she cannot answer a challenge by Choose. Otherwise, Cover wins.

Definition 4.2 ((s, δ) -bipartite expander). *Let s be a positive integer and δ be a positive real number. A bipartite graph G with bipartition (L, R) is an (s, δ) -**bipartite expander** if all subsets $X \subseteq L$ of size at most s satisfy $|N_G(X)| \geq \delta|X|$.*

The next theorem shows that Cover has a winning strategy for the game $\text{CoverGame}_{\text{VW}}(G, \mu)$ for expander graphs G with appropriately chosen parameters.

Theorem 4.3. *Let G be a bipartite graph with bipartition (L, R) , s, D be integers, and $\epsilon < \frac{1}{23}$ be a real number. For every integer $d \geq D$ let $S_d \subseteq R$ be the set of vertices of R with degree bigger than d . Suppose that*

1. *each vertex in L has degree 3;*
2. *G is an $(s, 2 - \frac{\epsilon}{2})$ -bipartite expander;*
3. *for every $d \geq D$, $\frac{72d}{\epsilon}(|S_d| + d) + 1 \leq \frac{s}{2}$.*

Then Cover wins the cover game $\text{CoverGame}_{\text{VW}}(G, \mu)$ with $\mu = \frac{\epsilon s}{144D}$.

The proof of this result is similar to constructions that can be found for example in [BSG03, Ats04, BGT14].

For the rest of this section, fix a bipartite graph G with bipartition (L, R) , an integer s and a real number $\epsilon < \frac{1}{23}$ such that G is an $(s, 2 - \frac{\epsilon}{2})$ -bipartite expander and each vertex in L has degree 3. Given $A \subseteq L$ and $B \subseteq R$, we let $G_{A,B}$ be the subgraph of G induced by $(L \cup R) \setminus (A \cup B)$.

Definition 4.4 (VW-matching property). *Given two sets $A \subseteq L$ and $B \subseteq R$, we say that the pair (A, B) has the **VW-matching property**, if for every $C \subseteq L \setminus A$ with $|C| \leq s$, there exists a VW-matching F in $G_{A,B}$ covering C .*

Lemma 4.5. *Let $A \subseteq L$ and $B \subseteq R$ be such that the pair (A, B) does not have the VW-matching property. Then there exists a set $C \subseteq L \setminus A$ with $|C| < \frac{2}{\epsilon}|B|$, such that no VW-matching in $G_{A,B}$ covers C .*

Proof. Take $C \subseteq L \setminus A$ of minimal size such that no VW-matching in $G_{A,B}$ covers C . We have that $|C| \leq s$ and by minimality of C and Lemma 1.2 it follows that

$$|N_{G_{A,B}}(C)| < (2 - \epsilon)|C|.$$

But, by hypothesis G is an $(s, 2 - \frac{\epsilon}{2})$ -bipartite expander; hence $(2 - \frac{\epsilon}{2})|C| \leq |N_G(C)|$. Therefore,

$$(2 - \frac{\epsilon}{2})|C| \leq |N_G(C)| \leq |N_{G_{A,B}}(C)| + |B| < (2 - \epsilon)|C| + |B|.$$

Hence $|C| < \frac{2}{\epsilon}|B|$, as required. \square

Lemma 4.5 is the only place where we directly use the $(2 - \epsilon)$ -Hall's Lemma (Lemma 1.2) from the previous section. However, Lemma 4.5 itself plays a crucial role in proving the following Lemmas (see Appendix B for the proofs).

Lemma 4.6. *The pair (\emptyset, \emptyset) has the VW-matching property.*

Lemma 4.7 (component removal). *Let $A \subseteq L$ and $B \subseteq R$ be such that the pair (A, B) has the VW-matching property and $\frac{2}{\epsilon}|B| \leq s$. Then for each VW-matching F contained in the subgraph of G induced by $A \cup B$, $(A \setminus L(F), B \setminus R(F))$ has the VW-matching property.*

Lemma 4.8 (covering a vertex in L). *Let $A \subseteq L$ and $B \subseteq R$ be such that the pair (A, B) has the VW-matching property and let d be the maximum degree of a vertex in $R \setminus B$. If $\frac{24d}{\epsilon}(|B| + 3) + 1 \leq s$, then for each vertex v in $L \setminus A$, there is a VW-matching F in $G_{A,B}$ covering v and such that $(A \cup L(F), B \cup R(F))$ has the VW-matching property.*

Lemma 4.9 (covering a vertex in R). *Let $A \subseteq L$ and $B \subseteq R$ be such that the pair (A, B) has the VW-matching property and let d be the maximum degree of a vertex in $R \setminus B$. If $\frac{24d}{\epsilon}(|B| + 3d) + 1 \leq s$, then for each vertex v in $R \setminus B$, there is a VW-matching F in $G_{A,B}$ covering v and such that $(A \cup L(F), B \cup R(F))$ has the VW-matching property.*

Proof of Theorem 4.3. By the hypothesis on $|S_d|$, for each $d \geq D$, we can repeatedly apply Lemma 4.9 starting from (\emptyset, \emptyset) to cover vertexes in R of degree larger than D . By starting from vertices of R of maximum degree and proceeding in decreasing order until reaching the vertices of degree D , we can build a VW-matching M covering S_D such that $(L(M), R(M))$ has the VW-matching property. Moreover, by the choice of S_D , $G_{L(M), R(M)}$ (the subgraph induced by $(L \cup R) \setminus (L(M) \cup R(M))$) has degree at most D . We say that a VW-matching F is *compatible* with M if each connected component of F is either a connected component of M or disjoint from all connected components of M .

We describe a winning strategy \mathcal{L} for Cover to win $\text{CoverGame}_{\text{VW}}(G, \mu)$. Take \mathcal{L} to be the set of all VW-matchings F in G compatible with M such that

1. $(L(M) \cup L(F), R(M) \cup R(F))$ has the VW-matching property, and
2. $\frac{2}{\epsilon}|R(M) \cup R(F)| \leq s$.

This family is non-empty since the empty VW-matching is in \mathcal{L} . Moreover, \mathcal{L} is closed under removing connected components by Lemma 4.7. Suppose now that at step $i + 1$ of the game Choose picks a vertex v in $G_{L(M), R(M)}$ and that F_i has strictly less than $\mu = \frac{\epsilon s}{144D}$ components. Then,

$(L(M) \cup L(F_i), R(M) \cup R(F_i))$ satisfies the hypotheses of Lemma 4.8 and Lemma 4.9:

$$\begin{aligned}
\frac{24D}{\epsilon}(|R(M) \cup R(F_i)| + 3D) + 1 &\leq \frac{24D}{\epsilon}(|R(M)| + |R(F_i)| + 3D) + 1 \\
&\leq \frac{24D}{\epsilon}(|R(M)| + 3D) + 1 + \frac{24D}{\epsilon}|R(F_i)| \\
&\stackrel{(\star)}{\leq} \frac{24D}{\epsilon}(3|S_D| + 3D) + 1 + \frac{72D}{\epsilon}\mu \\
&\stackrel{(\star\star)}{\leq} \frac{s}{2} + \frac{72D}{\epsilon}\mu = \frac{s}{2} + \frac{72D}{\epsilon} \frac{\epsilon s}{144D} = s,
\end{aligned}$$

where the inequality (\star) follows from the fact that $|R(F_i)| \leq 3\mu$ and $|R(M)| \leq 3|S_D|$, where S_D is the set of vertices in R of degree bigger than D . The inequality $(\star\star)$ follows by the hypothesis on the size of S_D .

Hence, if v is covered by F_i we take $F_{i+1} = F_i$. If v is covered by M we take $F_{i+1} = F_i \cup M_v$, where M_v is the connected component of M covering v . Otherwise, by Lemma 4.8 and Lemma 4.9 applied to $(L(M) \cup L(F_i), R(M) \cup R(F_i))$, there exists a VW-matching F_{i+1} extending $F_i \cup M$ by a new connected component covering v such that $(L(F_{i+1}), R(F_{i+1}))$ has the VW-matching property. From the previous chain of inequalities, it follows easily that the pair $(L(F_{i+1}), R(F_{i+1}))$ satisfies the cardinality condition $\frac{2}{\epsilon}|R(M) \cup R(F_{i+1})| = \frac{2}{\epsilon}|R(F_{i+1})| \leq s$. \square

5 Space lower bounds for random 3CNFs

Lemma 5.1. *Let φ be an unsatisfiable 3-CNF and G_φ its adjacency graph. If Cover wins the cover game $\text{CoverGame}_{\text{VW}}(G_\varphi, \mu)$, then there is a μ -winning strategy \mathcal{L} for $\text{tr}(\varphi)$.*

Proof. First of all we prove that for every VW-matching F in G_φ , there exists a flippable product-family of assignments H_F such that $H_F \models L(F)$, $\text{dom}(H_F) = R(F)$, and $\|H_F\|$ is the number of connected components of F .

We prove the result by induction on the number of connected components of F . If F is the union of two disjoint VW-matchings F', F'' then by hypothesis $H_{F'} \models L(F')$, $\text{dom}(H_{F'}) = R(F')$ and $\|H_{F'}\|$ is the number of connected components of F' . And analogously for F'' . Then, since $R(F')$ and $R(F'')$ are disjoint, $H_F = H_{F'} \otimes H_{F''}$ is well-defined. We immediately see that $H_F \models L(F)$, $\text{dom}(H_F) = R(F)$ and $\|H_F\|$ is the number of connected components of F .

It remains to consider the case when the VW-matching F is just one connected component. It is easy to see that all the possibilities can be reduced to those in Table 5.1.

It is straightforward to check that a winning strategy for Cover in $\text{CoverGame}_{\text{VW}}(G_\varphi, \mu)$ defines, by previous observations, a family \mathcal{L} of flippable product-families such that for all $\mathcal{H} \in \mathcal{L}$

1. for each $\mathcal{H}' \sqsubseteq \mathcal{H}$, $\mathcal{H}' \in \mathcal{L}$;
2. if $\|\mathcal{H}\| < \mu$, then: (a) for each $C \in \varphi$, there exists a flippable product-family $\mathcal{H}' \in \mathcal{L}$ such that $\mathcal{H}' \models C$ and $\mathcal{H}' \supseteq \mathcal{H}$; and (b) for each variable $x \notin \text{dom}(\mathcal{H})$, there exists a flippable family $\mathcal{H}' \in \mathcal{L}$ such that $\mathcal{H}' \supseteq \mathcal{H}$ and $x \in \text{dom}(\mathcal{H}')$.

We claim that \mathcal{L} is a μ -winning strategy. The *restriction property* is immediate. For the *extension property* we use the properties in (2) above: if we have to extend to something in \mathcal{L} that satisfies a boolean axiom we use property 2.(b), otherwise for all other polynomials in $\text{tr}(\varphi)$ we use property 2.(a). \square

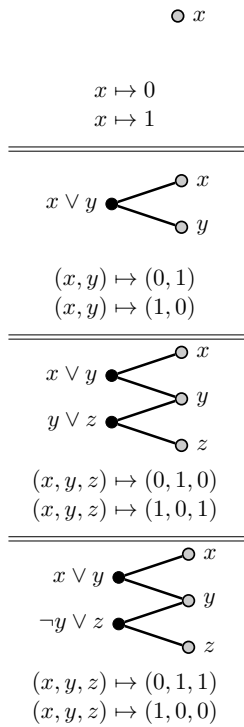


Table 5.1: Flippable assignments from VW-matchings

4.3. Thus, Cover wins the cover game $\text{CoverGame}_{\text{VW}}(G_\varphi, \mu)$ for $\mu = \Omega(n)$. Lemma 5.1 provides a $\Omega(n)$ -winning strategy and by Theorem 2.2 we have the monomial space lower bound in semantical PCR and the total space lower bound in RES. \square

References

- [ABSRW02] Michael Alekhnovich, Eli Ben-Sasson, Alexander A. Razborov, and Avi Wigderson. Space complexity in propositional calculus. *SIAM J. Comput.*, 31(4):1184–1211, 2002.
- [AD08] Albert Atserias and Víctor Dalmau. A combinatorial characterization of resolution width. *J. Comput. Syst. Sci.*, 74(3):323–334, 2008.
- [Ats04] Albert Atserias. On sufficient conditions for unsatisfiability of random formulas. *J. ACM*, 51(2):281–311, 2004.
- [Ben02] Eli Ben-Sasson. Size space tradeoffs for resolution. In John H. Reif, editor, *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 457–464. ACM, 2002.
- [BG] Ilario Bonacina and Nicola Galesi. A framework for space complexity in algebraic proof systems. *J. ACM*, to appear. Manuscript available at <http://wwwusers.di.uniroma1>.

Let $n, \Delta \in \mathbb{N}$ and let $X = \{x_1, \dots, x_n\}$ be a set of n variables. The probability distribution $\mathcal{R}(n, \Delta, 3)$ is obtained by the following experiment: choose independently uniformly at random Δn clauses from the set of all possible clauses with 3 literals over X . It is well-known that when Δ exceeds a certain constant θ_3 , φ is almost surely unsatisfiable [CS88, BP96, BSW01, BSG03]. Hence we always consider $\varphi \sim \mathcal{R}(n, \Delta, 3)$, where Δ is a constant bigger than θ_3 , which implies that φ is unsatisfiable with high probability. The proof of the next Lemma is in Appendix D.

Lemma 5.2. *Let $\Delta > \theta_3$ and $\varphi \sim \mathcal{R}(n, \Delta, 3)$ a random 3-CNF. For every integer d let S_d be the set of variables of φ appearing in at least d clauses of φ . Then for every constant $c > 0$ and $\epsilon > 0$, with high probability there exists a constant D such that for every $d \geq D$,*

$$\frac{72d}{\epsilon}(|S_d| + d) + 1 \leq cn.$$

Theorem 5.3. *If $\Delta > \theta_3$ and $\varphi \sim \mathcal{R}(n, \Delta, 3)$, then the following statements hold with high probability. For every semantical PCR refutation Π of $\text{tr}(\varphi)$, $\text{MSpace}(\Pi) \geq \Omega(n)$. Moreover, every RES refutation of φ must pass through a memory configuration containing $\Omega(n)$ clauses each of width $\Omega(n)$. In particular, each refutation of φ requires total space $\Omega(n^2)$.*

Proof. Let G_φ be the adjacency graph of φ . It is well known that G_φ is a $(\gamma n, 2 - \delta)$ -bipartite expander, for every $\delta > 0$ [CS88, BP96, BSW01, BSG03]. Hence in particular for $0 < \delta < \frac{1}{23}$ and using Lemma 5.2 with $c = \frac{\gamma}{2}$, we satisfy all the hypotheses of Theorem

it/~galesi/jacm.pdf. A preliminary version appeared as: *Pseudo-partitions, transversality and locality: a combinatorial characterization for the space measure in algebraic proof systems*. In ITCS, pages 455–472, 2013.

- [BGT14] Ilario Bonacina, Nicola Galesi, and Neil Thapen. Total space in resolution. In *55th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 641–650, 2014.
- [Bla37] Archie Blake. *Canonical Expressions in Boolean Algebra*. PhD thesis, 1937. University of Chicago.
- [BN08] Eli Ben-Sasson and Jakob Nordström. Short proofs may be spacious: An optimal separation of space and length in resolution. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 709–718. IEEE Computer Society, 2008.
- [BN11] Eli Ben-Sasson and Jakob Nordström. Understanding space in proof complexity: Separations and trade-offs via substitutions. In Bernard Chazelle, editor, *Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 7-9, 2011. Proceedings*, pages 401–416. Tsinghua University Press, 2011.
- [BP96] Paul Beame and Toniann Pitassi. Simplified and improved resolution lower bounds. In *FOCS*, pages 274–282. IEEE Computer Society, 1996.
- [BSG03] Eli Ben-Sasson and Nicola Galesi. Space complexity of random formulae in resolution. *Random Struct. Algorithms*, 23(1):92–109, 2003.
- [BSW01] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow - resolution made simple. *J. ACM*, 48(2):149–169, 2001.
- [CEI96] Matthew Clegg, Jeff Edmonds, and Russell Impagliazzo. Using the groebner basis algorithm to find proofs of unsatisfiability. In Gary L. Miller, editor, *STOC*, pages 174–183. ACM, 1996.
- [CS88] Vasek Chvátal and Endre Szemerédi. Many hard examples for resolution. *J. ACM*, 35(4):759–768, 1988.
- [ET01] Juan Luis Esteban and Jacobo Torán. Space bounds for resolution. *Inf. Comput.*, 171(1):84–97, 2001.
- [FLM⁺13] Yuval Filmus, Massimo Lauria, Mladen Mikša, Jakob Nordström, and Marc Vinyals. Towards an understanding of polynomial calculus: New separations and lower bounds - (extended abstract). In Fedor V. Fomin, Rusins Freivalds, Marta Z. Kwiatkowska, and David Peleg, editors, *ICALP (1)*, volume 7965 of *Lecture Notes in Computer Science*, pages 437–448. Springer, 2013.
- [FLN⁺12] Yuval Filmus, Massimo Lauria, Jakob Nordström, Neil Thapen, and Noga Ron-Zewi. Space complexity in polynomial calculus. In *Proceedings of the 27th Conference on Computational Complexity, CCC 2012, Porto, Portugal, June 26-29, 2012*, pages 334–344. IEEE, 2012.
- [Hal35] P. Hall. On representatives of subsets. *Journal of the London Mathematical Society*, s1-10(1):26–30, 1935.

- [Nor09] Jakob Nordström. Narrow proofs may be spacious: Separating space and width in resolution. *SIAM J. Comput.*, 39(1):59–121, 2009.
- [Nor13] Jakob Nordström. Pebble games, proof complexity, and time-space trade-offs. *Logical Methods in Computer Science*, 9(3), 2013.
- [Rob65] J. A. Robinson. A machine-oriented logic based on the resolution principle. *J. ACM*, 12(1):23–41, January 1965.

A Proof of Theorem 2.2

A *piecewise (p.w.) assignment* α of a set of variables X is a set of non-empty partial assignments to X with pairwise disjoint domains. We will sometimes call the elements of α the *pieces* of α . A piecewise assignment gives rise to a partial assignment $\bigcup \alpha$ to X together with a partition of the domain of $\bigcup \alpha$. For piecewise assignments α, β we will write $\alpha \sqsubseteq \beta$ to mean that every piece of α appears in β . We will write $\|\alpha\|$ to mean the number of pieces in α . Note that these are formally exactly the same as $\alpha \subseteq \beta$ and $|\alpha|$, if we regard α and β as sets.

Definition A.1 (*r*-free [BGT14]). A family \mathcal{F} of p.w. assignments is *r*-free for a CNF φ if it has the following properties:

(CONSISTENCY) No $\alpha \in \mathcal{F}$ falsifies any clause from φ ;

(RETRACTION) If $\alpha \in \mathcal{F}$, β is a p.w. assignment and $\alpha^* \sqsubseteq \beta \sqsubseteq \alpha$, then $\beta \in \mathcal{F}$;

(EXTENSION) If $\alpha \in \mathcal{F}$ and $\|\alpha\| < r$, then for every variable $x \notin \text{dom}(\alpha)$, there exist $\beta_0, \beta_1 \in \mathcal{F}$ with $\alpha \sqsubseteq \beta_0, \beta_1$ such that $\beta_0(x) = 0$ and $\beta_1(x) = 1$.

Theorem A.2 ([BGT14]). Let φ be an unsatisfiable CNF formula. If there is a family of p.w. assignments which is *r*-free for φ , then any resolution refutation of φ must pass through a memory configuration containing at least $\frac{r}{2}$ clauses each of width at least $\frac{r}{2}$. In particular, the refutation requires total space at least $\frac{r^2}{4}$.

By this theorem, in order to prove the total space lower bound of Theorem 2.2 we just have to prove that given a k -winning strategy for $\text{tr}(\varphi)$ we can build a $(k - 1)$ -free family for φ .

Let \mathcal{L} be the k -winning strategy. Define the $(k - 1)$ -free family \mathcal{F} as follows: $\alpha \in \mathcal{F}$ if and only if there exists $H_1 \otimes \dots \otimes H_t \in \mathcal{L}$ such that $\alpha = \alpha_1 \cup \dots \cup \alpha_t$ with $\alpha_i \in H_i$ and $t \leq k - 1$. The p.w. structure of α is inherited from the domain-disjointness of $H_1 \otimes \dots \otimes H_t$; in particular, $\|\alpha\| = \|H_1 \otimes \dots \otimes H_t\|$. The *retraction property* of \mathcal{F} is immediate from the corresponding property of \mathcal{L} .

To prove the *consistency property* of \mathcal{F} assume, by contradiction, that there is an $\alpha \in \mathcal{F}$ such that α falsifies some clause $C \in \varphi$. Since $\|\alpha\| \leq k - 1 < k$, there exists $\mathcal{H} = H_1 \otimes \dots \otimes H_t \in \mathcal{L}$ such that $\alpha \in \mathcal{H}$ and $\|\alpha\| = \|\mathcal{H}\|$. By the extension property of \mathcal{L} , there is an $\mathcal{H}' \supseteq \mathcal{H}$ such that $\mathcal{H}' \models \text{tr}(C)$. In particular there exists some partial assignment $\beta \supseteq \alpha$ such that $\beta \models \text{tr}(C)$. By construction, for every assignment γ , $\gamma \models \text{tr}(C)$ if and only if $\gamma \models C$. Thus $\beta \models C$, which is impossible since α falsifies C .

For the *extension property* let $\alpha \in \mathcal{F}$, with $\|\alpha\| < k - 1$ and let x be a variable of φ not in $\text{dom}(\alpha)$. By construction, there exists some $\mathcal{H} \in \mathcal{L}$ such that $\alpha \in \mathcal{H}$, $\|\alpha\| = \|\mathcal{H}\|$ and $\text{dom}(\alpha) = \text{dom}(\mathcal{H})$. By the extension property of \mathcal{F} there exists some flippable $\mathcal{H}' \in \mathcal{L}$ such that $\mathcal{H}' \supseteq \mathcal{H}$ and $\mathcal{H}' \models x^2 - x$. By taking restrictions in \mathcal{L} we can suppose that $\|\mathcal{H}'\| = \|\mathcal{H}\| + 1$. Hence there exist $\beta_0, \beta_1 \in \mathcal{F}$ extending α , setting x respectively to 0 and 1 and such that $\|\beta_0\| = \|\beta_1\| = \|\alpha\| + 1 \leq k - 1$. \square

B Proofs of the Lemmas from Section 4

For convenience we restate here also Lemma 4.5.

Restated Lemma 4.5. *Let $A \subseteq L$ and $B \subseteq R$ be such that the pair (A, B) does not have the VW-matching property. Then there exists a set $C \subseteq L \setminus A$ with $|C| < \frac{2}{\epsilon}|B|$, such that no VW-matching in $G_{A,B}$ covers C .*

Restated Lemma 4.6. *The pair (\emptyset, \emptyset) has the VW-matching property.*

Proof. Apply Lemma 4.5 with $A = \emptyset$ and $B = \emptyset$. □

Restated Lemma 4.7 (component removal). *Let $A \subseteq L$ and $B \subseteq R$ be such that the pair (A, B) has the VW-matching property and $\frac{2}{\epsilon}|B| \leq s$. Then for each VW-matching F contained in the subgraph of G induced by $A \cup B$, $(A \setminus L(F), B \setminus R(F))$ has the VW-matching property.*

Proof. Let $A' = A \setminus L(F)$ and $B' = B \setminus R(F)$ and suppose, by contradiction, that (A', B') does not have the VW-matching property. By Lemma 4.5, it is sufficient to prove that for each set $C \subseteq L \setminus A'$ with $|C| < \frac{2}{\epsilon}|B'|$, there is a VW-matching in $G_{A',B'}$ covering C . Let $C' = C \cap L(F)$ and $C'' = C \setminus C'$. By construction, F is a VW-matching such that $L(F) \subseteq A$, $R(F) \subseteq B$ and F covers C' . Moreover, we have that

$$|C''| \leq |C| < \frac{2}{\epsilon}|B'| < \frac{2}{\epsilon}|B| \stackrel{(*)}{\leq} s,$$

where the inequality $(*)$ is by hypothesis. Hence there exists a VW-matching F'' of C'' in $G_{A,B}$, and so $F \cup F''$ is a VW-matching covering C in $G_{A',B'}$. □

Restated Lemma 4.8 (covering a vertex in L). *Let $A \subseteq L$ and $B \subseteq R$ be such that the pair (A, B) has the VW-matching property and let d be the maximum degree of a vertex in $R \setminus B$. If $\frac{24d}{\epsilon}(|B| + 3) + 1 \leq s$, then for each vertex v in $L \setminus A$, there is a VW-matching F in $G_{A,B}$ covering v and such that $(A \cup L(F), B \cup R(F))$ has the VW-matching property.*

Proof. Fix $v \in L \setminus A$ and let Π be the set of all VW-matchings F in $G_{A,B}$, covering v and such that F is connected.

Since $1 \leq s$ and (A, B) has the VW-matching property, we know that Π is non-empty. For every $F \in \Pi$, let (A_F, B_F) be the pair $(A \cup L(F), B \cup R(F))$, and suppose for a contradiction that for every $F \in \Pi$, (A_F, B_F) does not have the VW-matching property. By Lemma 4.5, for every $F \in \Pi$ there is a set $C_F \subseteq L \setminus A_F$ with $|C_F| < \frac{2}{\epsilon}|B_F|$ and such that there is no VW-matching of C_F in G_{A_F, B_F} .

Let $C = \bigcup_{F \in \Pi} C_F$. Then

$$|C| \leq \sum_{F \in \Pi} |C_F| < |\Pi| \frac{2}{\epsilon}(|B| + 3) \leq 12d \frac{2}{\epsilon}(|B| + 3),$$

since $|\Pi| \leq 3 + 3 \cdot 2 \cdot (d - 1) \cdot 2 \leq 12d$ and $|B_F| \leq |B| + 3$. Hence, by our assumption about the size of $|B|$, we have that $|C \cup \{v\}| \leq s$. Furthermore, $C \cup \{v\} \subseteq L \setminus A$, so by the fact that (A, B) has the VW-matching property, there is a VW-matching F' covering $C \cup \{v\}$ in $G_{A,B}$.

There must be some $F \in \Pi$ such that F is a connected component of F' . Let F'' be F' with the component F removed. Then F'' is a VW-matching in G_{A_F, B_F} and F'' covers C_F , contradicting the choice of C_F . □

Restated Lemma 4.9 (covering a vertex in R). Let $A \subseteq L$ and $B \subseteq R$ be such that the pair (A, B) has the VW-matching property and let d be the maximum degree of a vertex in $R \setminus B$. If $\frac{24d}{\epsilon}(|B| + 3d) + 1 \leq s$, then for each vertex v in $R \setminus B$, there is a VW-matching F in $G_{A,B}$ covering v and such that $(A \cup L(F), B \cup R(F))$ has the VW-matching property.

Proof. Fix $v \in R \setminus B$ and let D be $N_G(v) \setminus A$. By hypothesis $|D| \leq d$. If $|D| = 0$, then $N_G(v) \subseteq A$, and so we can cover v by taking F to be the VW-matching consisting only of the vertex v . This is a valid VW-matching covering v and clearly $(A \cup L(F), B \cup R(F))$ has the VW-matching property.

If $|D| > 0$, by the cardinality condition on B , we can apply Lemma 4.8 $|D|$ times obtaining a VW-matching F in $G_{A,B}$ covering D and such that $(A \cup L(F), B \cup R(F))$ has the VW-matching property.

Now, since $N_G(v) \subseteq A \cup L(F)$, it follows that $(A \cup L(F), B \cup R(F) \cup \{v\})$ has the VW-matching property. Either v is covered by F , or it is possible to add $\{v\}$ as a new connected component to F while still maintaining the property of being a VW-matching in $G_{A,B}$. \square

C Proof of Proposition 3.1

As promised, we now prove Proposition 3.1, here rephrased in terms of hypergraphs.

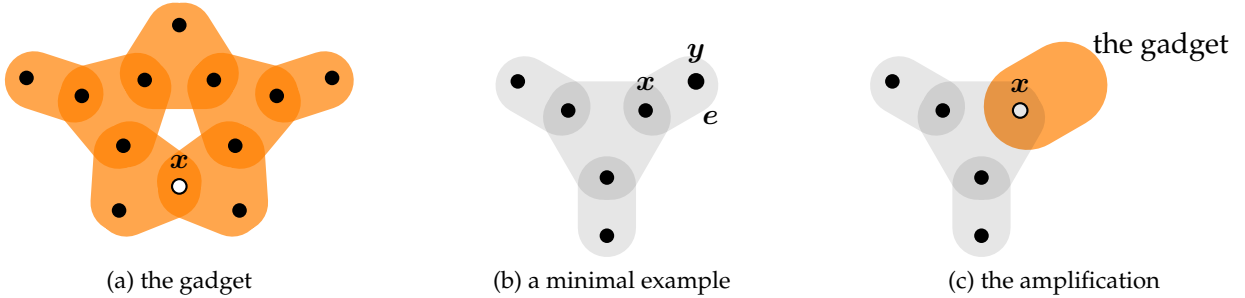


Figure C.1: The construction

Proposition C.1. For every $\epsilon > \frac{1}{3}$, there exists a hypergraph \mathcal{H}_ϵ such that \mathcal{H}_ϵ has no isolated vertices, each hyperedge of \mathcal{H}_ϵ has size 2 or 3, $|V(\mathcal{H})| \geq (2 - \epsilon)|E(\mathcal{H})|$, every proper subset of $E(\mathcal{H}_\epsilon)$ has a 2-path cover, but \mathcal{H}_ϵ does not have a 2-path cover.

Proof. Let $\epsilon > \frac{1}{3}$ and consider the gadget \mathcal{G} shown in Figure C.1.(a). It is easy to verify that every 2-path cover of \mathcal{G} must cover the vertex x . Next note that the hypergraph \mathcal{H} shown in Figure C.1.(b) is obviously not 2-path coverable, but every proper subset of $E(\mathcal{H})$ is 2-path coverable. We have $\frac{|V(\mathcal{H})|}{|E(\mathcal{H})|} = \frac{6}{4}$. However, we can increase this ratio via the amplification trick shown in Figure C.1.(c).

That is, let e be a hyperedge of \mathcal{H} of size 2. Label the vertices of e as x and y , where y has degree 1. Let \mathcal{H}_1 be the hypergraph obtained from \mathcal{H} by deleting y and then gluing \mathcal{G} to $\mathcal{H} - y$ along x . Since every 2-path cover of \mathcal{G} must use the vertex x , \mathcal{H}_1 does not have a 2-path cover. On the other hand, since every proper subset of $E(\mathcal{G})$ has a 2-path cover avoiding x , it follows that every proper subset of $E(\mathcal{H}_1)$ has a 2-path cover. Note that this amplification trick increases the number of vertices of \mathcal{H} by 10 and the number of edges of \mathcal{H} by 6. Moreover, we can repeat this amplification trick arbitrarily many times since \mathcal{G} also has pendent edges of size 2. So, choose n such that $\frac{6+10n}{4+6n} \geq 2 - \epsilon$ and take \mathcal{H}_ϵ to be the graph obtained from \mathcal{H} by performing the amplification trick n times. \square

D Proof of Lemma 5.2

Restated Lemma 5.2. Let $\Delta > \theta_3$ and $\varphi \sim \mathcal{R}(n, \Delta, 3)$ a random 3-CNF. For every integer d let S_d be the set of variables of φ appearing in at least d clauses of φ . Then for every constant $c > 0$ and $\epsilon > 0$, with high probability there exists a constant D such that for every $d \geq D$,

$$\frac{72d}{\epsilon}(|S_d| + d) + 1 \leq cn.$$

Proof. Let G_φ be the adjacency graph of φ . First of all we show that w.h.p. there are at most $\frac{en}{2^d}$ many variable nodes of degree d for every $d \geq 24e\Delta$ and that w.h.p. there is no variable node of degree bigger than $\log n$. First note that the expected number of variable nodes of degree at least $\log n$ is

$$n \binom{\Delta n}{\log n} \left(\frac{3}{n-2}\right)^{\log n} \leq n \left(\frac{e\Delta n}{\log n}\right)^{\log n} \left(\frac{3}{n-2}\right)^{\log n} = o(1).$$

So w.h.p. there are no such nodes. Let $d \geq 24e\Delta$. The probability that there are $\frac{en}{2^d}$ many variable nodes of degree d is at most

$$\binom{n}{\frac{en}{2^d}} \left[\binom{\Delta n}{d} \left(\frac{3}{n-2}\right)^d \right]^{\frac{en}{2^d}} \leq \left(\frac{en}{2^d}\right)^{\frac{en}{2^d}} \left[\left(\frac{e\Delta n}{d}\right)^d \left(\frac{3}{n-2}\right)^d \right]^{\frac{en}{2^d}} \leq \left(\frac{12e\Delta}{d}\right)^{\frac{edn}{2^d}} \leq \left(\frac{1}{2}\right)^{\frac{edn}{2^d}},$$

so, by the union bound, the probability that there exists any d between $24e\Delta$ and $\log n$ such that there are $\frac{en}{2^d}$ many variable nodes of degree d is at most $\sum_{24e\Delta \leq d \leq \log n} \left(\frac{1}{2}\right)^{\frac{edn}{2^d}}$. To bound this sum, note that the ratio of consecutive terms is

$$2^{\frac{edn}{2^d} - \frac{e(d+1)n}{2^{d+1}}} = 2^{\frac{e(d-1)n}{2^{d+1}}} \geq 2$$

for d in this range, and so the sum is of the order of its last term, which is $\left(\frac{1}{2}\right)^{\frac{en \log n}{2^{\log n}}} = o(1)$.

So we have that w.h.p.

$$|S_d| \leq \sum_{d' \geq d} \frac{en}{2^{d'}} \leq \frac{2en}{2^d}$$

and, for (a not yet chosen constant) $D \geq 24e\Delta$, we have that for each d such that $D \leq d \leq \log n$:

$$\frac{72d}{\epsilon}(|S_d| + d) + 1 \leq \frac{72d}{\epsilon} \left(\frac{2en}{2^d} + d\right) + 1 \leq \frac{72D}{\epsilon} \frac{2en}{2^D} + O(\log^2 n) \leq cn,$$

where the last inequality holds if D is a sufficiently large constant. \square