Resolution Lower Bounds for Refutation Statements

Michal Garlík
Dept. Ciències de la Computació, Universitat Politècnica de Catalunya,
C. Jordi Girona, 1-3, 08034 Barcelona, Spain
mgarlik@cs.upc.edu

Abstract

For any unsatisfiable CNF formula we give an exponential lower bound on the size of resolution refutations of a propositional statement that the formula has a resolution refutation. We describe three applications. (1) An open question in [2] asks whether a certain natural propositional encoding of the above statement is hard for Resolution. We answer by giving an exponential size lower bound. (2) We show exponential resolution size lower bounds for reflection principles, thereby improving a result in [1]. (3) We provide new examples of CNFs that exponentially separate Res(2) from Resolution (an exponential separation of these two proof systems was originally proved in [10]).

2012 ACM Subject Classification Theory of computation → Proof complexity

Keywords and phrases reflection principles, refutation statements, Resolution, proof complexity

Digital Object Identifier 10.4230/LIPIcs.MFCS.2019.37

Related Version `https://arxiv.org/abs/1905.12372v1`

Funding Funded by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement ERC-2014-CoG 648276 (AUTAR).

1 Introduction

Proving lower bounds on the size of propositional proofs is the central task of proof complexity theory. After Cook and Reckhow [4] motivated this line of research as an approach towards establishing NP $\neq$ coNP, some initial success for weak proof systems followed, e.g., the first exponential size lower bound for Resolution was proved by Haken [6]. Nevertheless, many important open problems from the 1980s and 1990s remain unsolved, and it seems that proving nontrivial lower bounds on the size of propositional proofs is hard. If it is hard for people, it is natural to ask if it is also hard for the proof systems themselves. In trying to formalize this question so that it makes sense to a proof system, we must say what we mean by "proving is hard". It can be "there are no short proofs", a statement which appears in a propositional formalization of reflection principles. By "short" we mean polynomial in the size of the formula being proven or refuted. The negation of the *reflection principle* for a proof system P is a conjunction of the statement "y is a P-refutation of length s of formula x of length n" and the statement "z is a satisfying assignment of formula x". In a propositional formulation of the principle, P, s, n are fixed parameters and x, y, z are disjoint sets of variables. A possible way to formalize the above question is then to take the first conjunct of the negation of the reflection principle and plug in for the x-variables some formula F of length n. The resulting formula was discussed and utilized by Pudlák [9]; we denote it by $\mathrm{REF}^F_{P,s}$ and call it a *refutation statement* for P. We may now ask whether some proof system Q can shortly refute $\mathrm{REF}^F_{P,s}$, and if it can not, we can interpret this to mean that lower bounds for P-refutations of F are hard for Q.

Pudlák [9] found connections between the reflection principles and automatizability, and these were elaborated on in [1]. Following [3], a proof system P is *automatizable* if there is a deterministic algorithm that when given as input an unsatisfiable CNF formula F outputs a P-refutation of F in time polynomial in the size of the shortest P-refutation of F. Recently, Atserias and Müller [2] showed that Resolution is not automatizable unless P = NP. Refutation statements for Resolution play a prominent role in their proof. They show that strong enough resolution size lower bounds for $\mathrm{REF}^F_{\mathrm{Res},s}$ with an unsatisfiable F imply their result. However, they leave the lower bound problem for $\mathrm{REF}^F_{\mathrm{Res},s}$ as an open question, and in place of $\mathrm{REF}^F_{\mathrm{Res},s}$ they use in the proof a different formulation of the refutation statement, obtained by a *relativization* of $\mathrm{REF}^F_{\mathrm{Res},s}$, for which lower bounds are easier to get. In this paper we focus mainly on giving an answer to the question.

1.1 Results in This Paper

The result that requires the most work is the following lower bound.

▶ **Theorem 1.** *For each $\epsilon > 0$ there is $\delta > 0$ and an integer t_0 such that if n, r, s, t are integers satisfying $t \geq s \geq n+1$, $r \geq n \geq 2$, $t \geq r^{3+\epsilon}$, $t \geq t_0$, and F is an unsatisfiable CNF consisting of r clauses $C_1, \ldots, C_r$ in n variables $x_1, \ldots, x_n$, then any resolution refutation of $\mathrm{REF}^F_{s,t}$ has length greater than 2^{t^δ}.*

We then show that this theorem implies an exponential resolution size lower bound for the encoding of the refutation statement for which the lower bound question in [2] is originally asked.

The formula $\mathrm{REF}^F_{s,t}$ in the theorem is a variant of the refutation statement insisting that the resolution refutation it describes has the form of a levelled graph. A similar simplifying assumption, making it more practical to design random restrictions, is used in [11] for a propositional version of the coloured polynomial local search principle. Our proof proceeds with defining a random restriction tailored to $\mathrm{REF}^F_{s,t}$ and to an adversary argument. The nature of the refutation statement and the fact that the relations between refutation lines are encoded in unary, rather than in binary, necessitate a more complicated adversary argument than in [8] or [11], and this in turn poses more requirements on the random restriction. We discuss these details after the proof, in Remarks 20 and 21.

We see two reasons for working with the unary encoding of $\mathrm{REF}^F_{s,t}$. First, $\mathrm{REF}^F_{s,t}$ is weaker than refutation statements encoded in binary or relativized refutation statements. Hence lower bounds for $\mathrm{REF}^F_{s,t}$ imply lower bounds for the other encodings. Second, researchers who dealt with propositional encodings of reflection principles or refutation statements opted for the unary encoding [1, 7, 9].

Besides these two reasons, we need to work with the unary encoding to give an answer to the above mentioned lower bound question from [2]. Our answer is stated in the following theorem, the proof of which (an easy reduction to Theorem 1) is in the full version of this paper [5]. Below, $\mathrm{REF}(F, \tilde{s})$ denotes the encoding of the resolution refutation statement in [2]. We can assume that the formula $\mathrm{REF}^F_{\mathrm{Res},s}$ is the same as $\mathrm{REF}(F, s)$.

▶ **Theorem 2.** *For each $\epsilon > 0$ there is $\delta > 0$ and an integer t_0 such that if $n, r, \tilde{s}$ are integers satisfying $r \geq n \geq 2$, $\left\lfloor \frac{\tilde{s}}{n+1} \right\rfloor \geq r^{3+\epsilon}$, $\left\lfloor \frac{\tilde{s}}{n+1} \right\rfloor \geq t_0$, and F is an unsatisfiable CNF consisting of r clauses $C_1, \ldots, C_r$ in n variables $x_1, \ldots, x_n$, then any resolution refutation of $\mathrm{REF}(F, \tilde{s})$ has length greater than $2^{\lfloor \frac{\tilde{s}}{n+1} \rfloor^\delta}$.*

Our next result is that the negation of the reflection principle for Resolution, expressed by the formula $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{s,t}$, exponentially separates the system Res(2) from Resolution. It was shown by Atserias and Bonet [1] that a similar encoding of the negation of the reflection

principle separates the two theories almost-exponentially (giving a $2^{\Omega(2^{\log^{\epsilon} n})}$ resolution lower bound and a polynomial Res(2) upper bound). The exponential separation of Res(2) from Resolution was originally proved in [10] using a variation of the graph ordering principle. Our lower bound is stated in Theorem 3 below.

▶ **Theorem 3.** *For every $c > 4$ there is $\delta > 0$ and an integer n_0 such that if n, r, s, t are integers satisfying $t \geq s \geq n + 1$, $r \geq n \geq n_0$, $n^c \geq t \geq r^4$, then any resolution refutation of $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{s,t}$ has length greater than 2^{n^δ}.*

The proof of the theorem also yields new examples of CNFs exponentially separating Res(2) from Resolution.

▶ **Theorem 4.** *Let $\delta_1 > 0$ and let $\{A_n\}_{n\geq 1}$ be a family of unsatisfiable CNFs such that A_n is in n variables, has the number of clauses polynomial in n, and has no resolution refutations of length at most $2^{n^{\delta_1}}$. Then there is $\delta > 0$ and a polynomial p such that $A_n \wedge \mathrm{REF}^{A_n}_{n+1,p(n)}$ has no resolution refutations of length at most 2^{n^δ} and has polynomial size* Res(2) *refutations.*

A Res(2) upper bound for $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{s,t}$, needed for completing the separation by this formula as well as by the formulas in Theorem 4, is stated in the following theorem.

▶ **Theorem 5.** *The negation of the reflection principle for Resolution expressed by the formula $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{s,t}$ has Res(2) refutations of size $O(trn^2 + tr^2 + st^2n^3 + st^3n)$.*

A polynomial size Res(2) upper bound on a similar encoding of the negation of the reflection principle for Resolution was proved in [1]. We simplify the proof and adapt it to $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{s,t}$ (see the full version [5]).

1.2 Outline of This Paper

The rest of the paper is organized as follows.

In Section 2 we give the necessary preliminaries.

In Section 3, Resolution of s levels of t clauses is introduced, and the clauses of the refutation statement $\mathrm{REF}^F_{s,t}$ for this refutation system are listed. We also state here a quadratic simulation of Resolution by this system.

In Section 4 we prove Theorem 1.

In Section 5 we define the formula $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{s,t}$ and we prove Theorems 3 and 4.

2 Preliminaries

For an integer s, the set $\{1, \ldots, s\}$ is denoted by $[s]$. We write $\mathrm{dom}(f), \mathrm{im}(f)$ for the domain and image of a funciton f. If x is a propositional variable, the *positive literal of* x, denoted by x^1, is x, and the *negative literal of* x, denoted by x^0, is $\neg x$. A *clause* is a set of literals. A clause is written as a disjunction of its elements. A *term* is a set of literals, and is written as a conjunction of the literals. A *CNF* is a set of clauses, written as a conjunction of the clauses. A *k-CNF* is a CNF whose every clause has at most k literals. A *DNF* is a set of terms, written as a disjunction of the terms. A *k-DNF* is a DNF whose every term has at most k literals. We will identify 1-DNFs with clauses. A clause is *non-tautological* if it does not contain both the positive and negative literal of the same variable. A clause C is a *weakening* of a clause D if $D \subseteq C$. A clause D is the *resolvent of* clauses C_1 and C_2 *on* a variable x if $x \in C_1, \neg x \in C_2$ and $D = (C_1 \setminus \{x\}) \cup (C_2 \setminus \{\neg x\})$. If E is a weakening of the resolvent of C_1 and C_2 on x, we say that E is obtained by the *resolution rule* from C_1 and C_2, and we call C_1 and C_2 the *premises* of the rule.

Let F be a CNF and C a clause. A *resolution derivation of C from F* is a sequence of clauses $\Pi = (C_1, \ldots, C_s)$ such that $C_s = C$ and for all $u \in [s]$, C_u is a weakening of a clause in F, or there are $v, w \in [u-1]$ such that C_u is obtained by the resolution rule from C_v and C_w. The *length* of the derivation Π is s. For $u \in [s]$, the *height of u in* Π is the maximum h such that there is a subsequence $(C_{u_1}, \ldots, C_{u_h})$ of Π in which $u_h = u$ and for each $i \in [h-1]$, C_{u_i} is a premise of a resolution rule by which $C_{u_{i+1}}$ is obtained in Π. The *height* of Π is the maximum height of u in Π for $u \in [s]$. A *resolution refutation of F* is a resolution derivation of the empty set from F.

A *partial assignment* to the variables $x_1, \ldots, x_n$ is a partial map from $\{x_1, \ldots, x_n\}$ to $\{0, 1\}$. Let σ be a partial assignment. The CNF $F \restriction \sigma$ is formed from F by removing every clause containing a literal satisfied by σ, and removing every literal falsified by σ from the remaining clauses. If $\Pi = (C_1, \ldots, C_s)$ is a sequence of clauses, $\Pi \restriction \sigma$ is formed from Π by the same operations. Note that if Π is a resolution refutation of F, then $\Pi \restriction \sigma$ is a resolution refutation of $F \restriction \sigma$.

The Res(k) refutation system is a generalization of Resolution. Its lines are k-DNFs and it has the following inference rules (A, B are k-DNFs, $j \in [k]$, and $l, l_1, \ldots, l_j$ are literals):

$$\frac{A \vee l_1 \qquad B \vee (l_2 \wedge \cdots \wedge l_j)}{A \vee B \vee (l_1 \wedge \cdots \wedge l_j)} \wedge\text{-introduction} \qquad \frac{}{x \vee \neg x} \text{ Axiom}$$

$$\frac{A \vee (l_1 \wedge \cdots \wedge l_j) \qquad B \vee \neg l_1 \vee \cdots \vee \neg l_j}{A \vee B} \text{ Cut} \qquad \frac{A}{A \vee B} \text{ Weakening}$$

Let F be a CNF. A *Res(k) derivation from F* is a sequence of k-DNFs $(D_1, \ldots, D_s)$ so that each D_i either belongs to F or follows from the preceding lines by an application of one of the inference rules. The *size* of a Res(k) derivation is the number of symbols in it.

3 Resolution Refutations of s Levels of t Clauses

We introduce a variant of Resolution in which the clauses forming a refutation are arranged in layers.

▶ **Definition 6.** Let F be a CNF of r clauses in n variables $x_1, \ldots, x_n$. We say that F has a *resolution refutation of s levels of t clauses* if there is a sequence of clauses $C_{i,j}$ indexed by all pairs $(i, j) \in [s] \times [t]$, such that each clause $C_{1,j}$ on the first level is a weakening of a clause in F, each clause $C_{i,j}$ on level $i \in \{2, \ldots, s\}$ is a weakening of the resolvent of two clauses from level $i-1$ on a variable, and the clause $C_{s,t}$ is empty.

The following proposition shows that this system quadratically simulates Resolution and preserves the refutation height. The proof (in the full version [5] of the paper) uses a simple self-replicating pattern both to transport a premise of the resolution rule to the required level and to fill in all clauses $C_{i,j}$ that do not directly participate in the simulation.

▶ **Proposition 7.** *If a $(n-1)$-CNF F in n variables has a resolution refutation of height h and length s, then F has a resolution refutation of h levels of $3s$ clauses.*

We proceed to our formalization of the refutation statement for this refutation system. Let n, r, s, t be integers. Let F be a CNF consisting of r clauses $C_1, \ldots, C_r$ in n variables $x_1, \ldots, x_n$. We define a propositional formula $\mathrm{REF}^F_{s,t}$ expressing that F has a resolution refutation of s levels of t clauses.

We first list the variables of $\mathrm{REF}^F_{s,t}$. *D-variables* $D(i, j, k, b)$, $i \in [s]$, $j \in [t], k \in [n], b \in \{0, 1\}$, encode clauses $C_{i,j}$ as follows: $D(i, j, k, 1)$ (resp. $D(i, j, k, 0)$) means that the literal x_k (resp. $\neg x_k$) is in $C_{i,j}$. *L-variables* $L(i, j, j')$ (resp. *R-variables* $R(i, j, j')$),

$i \in \{2,\dots,s\}, j, j' \in [t]$, say that $C_{i-1,j'}$ is a premise of the resolution rule by which $C_{i,j}$ is obtained, and it is the premise containing the positive (resp. negative) literal of the resolved variable. *V-variables* $V(i,j,k)$, $i \in \{2,\dots,s\}, j \in [t], k \in [n]$, say that $C_{i,j}$ is obtained by resolving on x_k. *I-variables* $I(j,m)$, $j \in [t], m \in [r]$, say that $C_{1,j}$ is a weakening of C_m.

$\mathrm{REF}^F_{s,t}$ is the union of the following fifteen sets of clauses:

$$\neg I(j,m) \vee D(1,j,k,b) \qquad j \in [t], m \in [r], b \in \{0,1\}, x_k^b \in C_m, \tag{1}$$

clause $C_{1,j}$ contains the literals of C_m assigned to it by $I(j,m)$,

$$\neg D(i,j,k,1) \vee \neg D(i,j,k,0) \qquad i \in [s], j \in [t], k \in [n], \tag{2}$$

no clause $C_{i,j}$ contains x_k and $\neg x_k$ at the same time,

$$\neg L(i,j,j') \vee \neg V(i,j,k) \vee D(i-1,j',k,1) \qquad i \in \{2,\dots,s\}, j, j' \in [t], k \in [n], \tag{3}$$

$$\neg R(i,j,j') \vee \neg V(i,j,k) \vee D(i-1,j',k,0) \qquad i \in \{2,\dots,s\}, j, j' \in [t], k \in [n], \tag{4}$$

clause $C_{i-1,j'}$ used as the premise given by $L(i,j,j')$ (resp. $R(i,j,j')$) in resolving on x_k must contain x_k (resp. $\neg x_k$),

$$\neg L(i,j,j') \vee \neg V(i,j,k) \vee \neg D(i-1,j',k',b) \vee D(i,j,k',b) \qquad i \in \{2,\dots,s\}, j, j' \in [t], k, k' \in [n], b \in \{0,1\}, (k',b) \neq (k,1), \tag{5}$$

$$\neg R(i,j,j') \vee \neg V(i,j,k) \vee \neg D(i-1,j',k',b) \vee D(i,j,k',b) \qquad i \in \{2,\dots,s\}, j, j' \in [t], k, k' \in [n], b \in \{0,1\}, (k',b) \neq (k,0), \tag{6}$$

clause $C_{i,j}$ derived by resolving on x_k must contain each literal different from x_k (resp. $\neg x_k$) from the premise given by $L(i,j,j')$ (resp. $R(i,j,j')$),

$$\neg D(s,t,k,b) \qquad k \in [n], b \in \{0,1\}, \tag{7}$$

clause $C_{s,t}$ is empty,

$$V(i,j,1) \vee V(i,j,2) \vee \ldots \vee V(i,j,n) \qquad i \in \{2,\dots,s\}, j \in [t], \tag{8}$$

$$I(j,1) \vee I(j,2) \vee \ldots \vee I(j,r) \qquad j \in [t], \tag{9}$$

$$L(i,j,1) \vee L(i,j,2) \vee \ldots \vee L(i,j,t) \qquad i \in \{2,\dots,s\}, j \in [t], \tag{10}$$

$$R(i,j,1) \vee R(i,j,2) \vee \ldots \vee R(i,j,t) \qquad i \in \{2,\dots,s\}, j \in [t], \tag{11}$$

$$\neg V(i,j,k) \vee \neg V(i,j,k') \qquad i \in \{2,\dots,s\}, j \in [t], k, k' \in [n], k \neq k', \tag{12}$$

$$\neg I(j,m) \vee \neg I(j,m') \qquad j \in [t], m, m' \in [r], m \neq m', \tag{13}$$

$$\neg L(i,j,j') \vee \neg L(i,j,j'') \qquad i \in \{2,\dots,s\}, j, j', j'' \in [t], j' \neq j'', \tag{14}$$

$$\neg R(i,j,j') \vee \neg R(i,j,j'') \qquad i \in \{2,\dots,s\}, j, j', j'' \in [t], j' \neq j'', \tag{15}$$

the V, I, L, R-variables define functions with the required domains and ranges.

4 A Lower Bound on Lengths of Resolution Refutations of $\mathrm{REF}^F_{s,t}$

We restate Theorem 1 from the Introduction.

▶ **Theorem 8.** *For each $\epsilon > 0$ there is $\delta > 0$ and an integer t_0 such that if n, r, s, t are integers satisfying*

$$t \geq s \geq n+1, \qquad r \geq n \geq 2, \qquad t \geq r^{3+\epsilon}, \qquad t \geq t_0, \tag{16}$$

and F is an unsatisfiable CNF consisting of r clauses $C_1, \dots, C_r$ in n variables $x_1, \dots, x_n$, then any resolution refutation of $\mathrm{REF}^F_{s,t}$ has length greater than 2^{t^δ}.

The rest of this section is devoted to a proof of the theorem. We argue by contradiction. Fix $\epsilon > 0$ and assume that for each $\delta > 0$ and t_0 there are integers n, r, s, t satisfying (16), an unsatisfiable CNF F, and a resolution refutation Π of $\mathrm{REF}^F_{s,t}$, such that F consists of r clauses $C_1, \ldots, C_r$ in n variables $x_1, \ldots, x_n$, and Π has length at most 2^{t^δ}.

The forthcoming distribution on partial assignments to the variables of $\mathrm{REF}^F_{s,t}$ employs in its definition and analysis two important parameters, p and w. We choose them as function of t and ϵ as follows:

$$p = t^{-a} \text{ with } a = \min\left\{\frac{2+\epsilon/2}{3+\epsilon/2}, \frac{3}{4}\right\}, \qquad w = t^{4/5}.$$

We now fix values of t_0, δ for which we will get the desired contradiction. Take t_0 so large and $\delta > 0$ so small that the inequalities

$$\max\left\{e^{-\frac{pw}{3}} + 2s \cdot e^{-\frac{pt}{3}}, e^{-\frac{pt}{8r}}\right\} \cdot 2^{t^\delta} + 3s \cdot e^{-\frac{pt}{3}} + 3p + 67p^3st < 1, \tag{17}$$

$$10pt + 4w < \frac{t}{4}, \tag{18}$$

$$e^{e^{\ln(t) - \frac{pt}{3}}} < 2, \tag{19}$$

hold for any n, r, s, t satisfying (16).

▶ **Definition 9.** For $i \in [s], j, j' \in [t], k \in [n], b \in \{0,1\}, m \in [r]$, we say that (i,j) is the *home pair* of the variable $D(i,j,k,b)$ (resp. $R(i,j,j')$; $L(i,j,j')$; $V(i,j,k')$; $I(j,m)$ if $i = 1$).

We write $V(i,j,\cdot)$ to stand for the set $\{V(i,j,k) : k \in [n]\}$. Similarly, we write $I(j,\cdot), L(i,j,\cdot), R(i,j,\cdot)$ to stand for the corresponding sets of variables, and we denote by $D(i,j,\cdot,\cdot)$ the set of variables $\{D(i,j,k,b) : k \in [n], b \in \{0,1\}\}$.

Let σ be a partial assignment. We say that $V(i,j,\cdot)$ is *set to* k *by* σ if $\sigma(V(i,j,k)) = 1$ and $\sigma(V(i,j,k')) = 0$ for all $k' \in [n], k' \neq k$. Similarly for $I(j,\cdot), L(i,j,\cdot), R(i,j,\cdot)$. We say that $D(i,j,\cdot,\cdot)$ is *set to* a clause $C_{i,j}$ *by* σ if for all $k \in [n], b \in \{0,1\}$, $\sigma(D(i,j,k,b)) = 1$ if $x^b_k \in C_{i,j}$ and $\sigma(D(i,j,k,b)) = 0$ if $x^b_k \notin C_{i,j}$.

For $Y \in \{D(i,j,\cdot,\cdot), V(i,j,\cdot), I(j,\cdot), R(i,j,\cdot), L(i,j,\cdot),\}$, we say that Y is *set by* σ if Y is set to v by σ for some value v. We will often omit saying "by σ" if σ is clear from the context.

▶ **Definition 10.** A *random restriction* ρ is a partial assignment to the variables of $\mathrm{REF}^F_{s,t}$ given by the following experiment:

1. For each pair $(i,j) \in [s] \times [t]$, with independent probability p include (i,j) in a set A_D. Then for each $(i,j) \in A_D$ and for each $k \in [n]$, independently, with probability $1/2$ choose between including the literal x_k or $\neg x_k$ in a clause $C_{i,j}$. Set $D(i,j,\cdot,\cdot)$ to $C_{i,j}$.
2. For each $j \in [t]$, with independent probability p include the pair $(1,j)$ in a set A_I. Then for each $(1,j) \in A_I \setminus A_D$, independently, choose at random $m \in [r]$ and set $I(j,\cdot)$ to m.
3. For each pair $(i,j) \in \{2,\ldots,s\} \times [t]$, with independent probability p include (i,j) in a set A_V. Then for each $(i,j) \in A_V$, independently, choose at random $k \in [n]$ and set $V(i,j,\cdot)$ to k.
4. For each pair $(i,j) \in \{2,\ldots,s\} \times [t]$, with independent probability p include the pair (i,j) in a set A_{RL}. Then, for each $i \in \{2,\ldots,s\}$, define $A_i := A_{RL} \cap (\{i\} \times [t])$ and do the following. If $|A_i| > 2pt$, define $h_i := \emptyset$, $B_{i-1} := \emptyset$. Otherwise, choose at random an injection h_i from $\{L(i,j,\cdot) : (i,j) \in A_i\} \cup \{R(i,j,\cdot) : (i,j) \in A_i\}$ to $[t]$. Define $B_{i-1} := \{(i-1,j) : j \in \mathrm{im}(h_i)\}$. Set $L(i,j,\cdot)$ to $h_i(L(i,j,\cdot))$ and $R(i,j,\cdot)$ to $h_i(R(i,j,\cdot))$ for all $(i,j) \in A_i$.

▶ **Lemma 11.** *With probability at least* $1 - 3s \cdot e^{-pt/3}$*, all of the following are satisfied.*

(i) *For each* $i \in [s]$*, the cardinality of* $A_D \cap (\{i\} \times [t])$ *is at most* $2pt$*.*

(ii) *For each* $i \in \{2, \dots, s\}$*, the cardinality of* A_i *is at most* $2pt$ *and the cardinality of* $A_V \cap (\{i\} \times [t])$ *is at most* $2pt$*.*

(iii) *The cardinality of* A_I *is at most* $2pt$*.*

Proof. By the Chernoff bound and the union bound it follows that item i is false with probability at most $s \cdot e^{-pt/3}$. Similarly for the remaining items. ◀

▶ **Definition 12.** Denote by G_ρ the graph with vertices $A_D \cup A_V \cup A_I \cup A_{RL} \cup \bigcup_{i \in [s-1]} B_i$, and with edges only between vertices on neighboring levels, such that (i, j) is connected by an edge to $(i-1, j')$ if and only if $h_i(L(i, j, \cdot)) = j'$ (then $(i-1, j')$ is called the *left child* of (i, j)) or $h_i(R(i, j, \cdot)) = j'$ (then $(i-1, j')$ is the *right child* of (i, j)).

The following lemma will be used later to show that a random restriction likely does not falsify any clause of $\mathrm{REF}^F_{s,t}$.

▶ **Lemma 13.** *With probability at least* $1 - 3p - 67p^3st$*, the following are satisfied.*

(i) $(s, t) \notin (A_D \cup A_{RL} \cup A_V)$*.*

(ii) *There is no triple* $((i_1, j_1), (i_2, j_2), (i_3, j_3))$ *of elements of* $[s] \times [t]$*, such that all the following hold:*

(a) *For each* $u \in [3]$ *there is* $X \in \{D, V, I, RL\}$ *with* $(i_u, j_u) \in A_X$*,*

(b) $|\{(i_u, j_u, X) : u \in [3], X \in \{D, V, I, RL\}, (i_u, j_u) \in A_X\}| \geq 3$*,*

(c) *the subgraph of* G_ρ *consisting of the vertices that are in the triple and their children and all edges that go from a vertex of the triple to its children, is connected.*

Proof. The probability that item i is true is $(1-p)^3 \geq 1 - 3p$.

Regarding item ii, we distinguish several cases based on the relative positions of the elements in a triple $((i_1, j_1), (i_2, j_2), (i_3, j_3))$. Note that the order in which the elements of the triple are listed does not matter in what we are proving, but some of the elements may coincide. When considering the cases, recall that due to our choice of the function h_i in the definition of ρ, two vertices in G_ρ cannot share a child.

In case all the elements of the triple are the same, iib is satisfied only if the element is chosen to A_X for three distinct values of X. This cannot happen on level 1, and on the other levels it happens with probability p^3. There are st many triples considered in the present case, so by the union bound the probability that there is any such triple satisfying all conditions in ii is at most p^3st.

In case $(i_1, j_1) \neq (i_2, j_2) = (i_3, j_3)$, condition iic is satisfied only if $i_1 = i_2 + 1$ or $i_2 = i_1 + 1$. In each of these two subcases, there are at most st^2 such triples. In the former subcase, we must have $(i_1, j_1) \in A_{RL}$ and at the same time $h_{i_1}(R(i_1, j_1, \cdot)) = j_2$ or $h_{i_1}(L(i_1, j_1, \cdot)) = j_2$. This happens with probability at most $2p/t$. Also, (i_2, j_2) has to be in A_X and $A_{X'}$ for distinct X, X', which happens with probability at most $3p^2$. So, the probability that any triple considered in this subcase satisfies iia - iic is at most $st^2 \cdot 6p^3/t = 6p^3st$. In the latter subcase, (i_2, j_2) has to be in A_{RL}, connected to (i_1, j_1), and additionally it has to be in A_D or A_V, while (i_1, j_1) has to be in arbitrary possible A_X. This happens with probability at most $2p/t \cdot 2p \cdot 3p = 12p^3/t$, so the probability that any such triple satisfies iia - iic is at most $12p^3st$.

In case all the elements of the triple are distinct, we again consider two subcases: first, $i_1 = i_2 + 1 = i_3 + 2$, and second, $i_1 - 1 = i_2 = i_3$. Each subcase concerns at most st^3 triples. In the first subcase, (i_3, j_3) has to be a child of (i_2, j_2), which in turn has to be a child of (i_1, j_1), and (i_3, j_3) also has to be in arbitrary possible A_X. This happens with

probability at most $12p^3/t^2$. Hence the probability that any such triple satisfies iia - iic is at most $12p^3st$. In the second subcase, (i_1, j_1) has to have children (i_2, j_2) and (i_3, j_3), and each child has to be in some A_X for any suitable X. This happens with probability at most $2p/(t(t-1)) \cdot (3p)^2 = 18p^3/(t(t-1)) \leq 36p^3/t^2$. Hence the probability that any such triple satisfies iia - iic is at most $36p^3st$. ◀

We now define some specific ways to measure a clause and we use them in the next lemma to describe how a clause simplifies under a restriction.

▶ **Definition 14.** Let E be a clause in $\Pi \restriction \rho$, and let $(i, j) \in [s] \times [t]$. If E contains a literal of a variable from $D(i, j, \cdot, \cdot)$ (resp. $R(i, j, \cdot)$; $L(i, j, \cdot)$; $V(i, j, \cdot)$; $I(j, \cdot)$ and $i = 1$), we say that the pair (i, j) is *D-mentioned* (resp. *R-mentioned*; *L-mentioned*; *V-mentioned*; *I-mentioned*) *in* E.

We say that (i, j) is *V-important* (resp. *L-important*; *R-important*; *I-important*) *in* E if E contains the negative literal of a variable in $V(i, j, \cdot)$ (resp. $L(i, j, \cdot); R(i, j, \cdot); I(j, \cdot)$ and $i = 1$) or if E contains at least $n/2$ (resp. $t/2$; $t/2$; $r/2$) positive literals of variables in $V(i, j, \cdot)$ (resp. $L(i, j, \cdot)$; $R(i, j, \cdot)$; $I(j, \cdot)$ and $i = 1$). A pair is *D-important in* E if it is D-mentioned in E.

▶ **Lemma 15.** *With probability at least* $1 - \max\left\{e^{-\frac{pw}{3}} + 2s \cdot e^{-\frac{pt}{3}}, e^{-\frac{pt}{8r}}\right\} \cdot 2^{t^\delta}$*, for every clause* E *in* $\Pi \restriction \rho$ *all of the following are satisfied.*

- **(i)** *At most* w *many pairs* (i, j) *are D-mentioned in* E.
- **(ii)** *At most* w *many pairs* $(1, j)$ *are I-important in* E.
- **(iii)** *At most* w *many pairs* (i, j) *are V-important in* E.
- **(iv)** *At most* w *many pairs* (i, j) *are L-important in* E.
- **(v)** *At most* w *many pairs* (i, j) *are R-important in* E.
- **(vi)** *For each* $m \in [r]$, $|\{j : I(j, m) \in E\}| \leq \frac{t}{4}$.
- **(vii)** *For each* $i \in \{s - n + 1, \ldots, s - 1\}$ *and* $k \in [n]$, $|\{j : V(i, j, k) \in E\}| \leq \frac{t}{4}$.

Proof. It is sufficient to prove that if E' is a clause in Π that violates any of i - vii, then with probability at least $1 - \max\left\{e^{-\frac{pw}{3}} + 2s \cdot e^{-\frac{pt}{3}}, e^{-\frac{pt}{8r}}\right\}$, E' is satisfied by ρ. Since Π has length at most 2^{t^δ}, the lemma then follows by the union bound.

Regarding item i, assume that E' in Π D-mentions more than w pairs (i, j). This means that a literal of a variable in $D(i, j, \cdot, \cdot)$ is in E' for more than w many pairs (i, j). For each such (i, j), such a literal is satisfied by ρ with probability at least $p/2$. So the probability that none of these literals in E' is satisfied is at most $(1 - p/2)^w < e^{-pw/2}$.

Regarding item ii, suppose that more than w pairs $(1, j)$ are I-important in E'. For each such $(1, j)$, the probability that $(1, j) \in A_I \setminus A_D$ is $p(1 - p)$, and provided this happens, the probability that ρ satisfies a literal in E' of a variable in $I(j, \cdot)$ is at least $\min\{(r - 1)/r, 1/2\} = 1/2$. Hence the probability that E' is not satisfied by ρ is at most $(1 - p(1 - p)/2)^w < (1 - p/3)^w < e^{-pw/3}$ (the first inequality follows from (18)).

Regarding item iii, a calculation similar to that for ii gives that a clause E' in Π with more than w many V-important pairs (i, j) is not satisfied by ρ with probability at most $(1 - p/2)^w < e^{-pw/2}$.

Regarding item iv, suppose that more than w many pairs (i, j) from $\{2, \ldots, s\} \times [t]$ are L-important in E'. For each $i \in \{2, \ldots, s\}$, assume without loss of generality that the set of pairs (i, j) that are L-important in E' is the set $\{(i, 1), \ldots, (i, w_i)\}$; denote it by W_i. Note that the distribution of ρ does not change if we choose A_i and h_i in t many steps as follows. Start with $A_{i,0} = h_{i,0} = \emptyset$. At step $j = 1, 2, \ldots, t$, first add (i, j) to $A_{i,j-1}$ with probability p to get $A_{i,j}$. Then, if $|A_{i,j}| \leq 2pt$ and $(i, j) \in A_{i,j}$, choose at random two distinct

elements j', j'' from $[t] \setminus \mathrm{im}(h_{i,j-1})$, and define $h_{i,j} := h_{i,j-1} \cup \{(L(i,j,\cdot), j'), (R(i,j,\cdot), j'')\}$. If $|A_{i,j}| \leq 2pt$ and $(i,j) \notin A_{i,j}$, define $h_{i,j} := h_{i,j-1}$. If $|A_{i,j}| > 2pt$ define $h_{i,j} := \emptyset$. This finishes step j. Finally, define $A_i := A_{i,t}$ and $h_i := h_{i,t}$.

For $i \in \{2, \ldots, s\}$, let H_i be the set of literals in E' of a variable in $L(i,j,\cdot)$ for some $(i,j) \in W_i$. Also, for $(i,j) \in W_i$, let $T_{i,j}$ be the set of those $j' \in [t]$ such that the partial assignment given by setting $L(i,j,\cdot)$ to j' satisfies some literal in H_i. We know that $|T_{i,j}| \geq t/2$ for each $(i,j) \in W_i$.

The event that no literal in H_i is satisfied by ρ is a subset of the union of events (a) $|A_{i,t}| > 2pt$, and (b) $|A_{i,w_i}| \leq 2pt$ and for each $(i,j) \in A_{i,w_i}$, $h_{i,j}(L(i,j,\cdot)) \notin T_{i,j}$. Event (a) happens with probability at most $e^{-pt/3}$ by the Chernoff bound. We bound the probability of event (b). For each $j \in [w_i]$, if $(i,j) \in A_{i,j}$ and $|A_{i,j}| \leq 2pt$, then the probability that $h_{i,j}(L(i,j,\cdot)) \in T_{i,j}$ is at least $(|T_{i,j} \setminus \mathrm{im}(h_{i,j-1})|)/t \geq (t/2 - 4pt)/t = (1-8p)/2 \geq 1/3$ (the last inequality follows from (18)). Therefore, denoting $\ell := \min\{2pt, w_i\}$, the probability of event (b) is at most

$$\sum_{k=0}^{\ell} \binom{w_i}{k} p^k (1-p)^{w_i-k} \left(\frac{2}{3}\right)^k \leq \sum_{k=0}^{w_i} \binom{w_i}{k} \left(\frac{2p}{3}\right)^k (1-p)^{w_i-k} = (1-p/3)^{w_i}.$$

Thus, the probability that no literal in H_i is satisfied by ρ is at most $e^{-pt/3} + e^{-pw_i/3}$, and, denoting $S := \{i \in \{2, \ldots, s\} : w_i \neq 0\}$, the probability that no literal in $\bigcup_{i \in S} H_i$ is satisfied by ρ is at most

$$\begin{aligned}
\prod_{i \in S} \left(e^{-\frac{pw_i}{3}} + e^{-\frac{pt}{3}}\right) &\leq e^{-\frac{pw}{3}} + \sum_{k=1}^{|S|} \binom{|S|}{k} e^{-\frac{ptk}{3}} \\
&\leq e^{-\frac{pw}{3}} + |S| \cdot e^{-\frac{pt}{3}} \sum_{k=1}^{|S|} \binom{|S|-1}{k-1} e^{-\frac{pt(k-1)}{3}} \\
&= e^{-\frac{pw}{3}} + |S| \cdot e^{-\frac{pt}{3}} \cdot \left(1 + e^{-\frac{pt}{3}}\right)^{|S|-1} \\
&\leq e^{-\frac{pw}{3}} + s \cdot e^{-\frac{pt}{3}} \cdot e^{e^{\ln(t) - \frac{pt}{3}}} \leq e^{-\frac{pw}{3}} + 2s \cdot e^{-\frac{pt}{3}},
\end{aligned}$$

where the penultimate inequality follows from $|S| - 1 \leq s \leq t$, and the last inequality follows from (19).

Item v is handled in the same way as iv.

Regarding item vi, suppose that for some $m \in [r]$ there are more than $t/4$ of $I(j,m)$ in E'. Similarly to the case ii, each such $I(j,m)$ is satisfied by ρ with independent probability at least $p(1-p)/r > p/(2r)$, so E' is not satisfied with probability at most $(1-p/(2r))^{t/4} < e^{-pt/(8r)}$.

Item vii is treated similarly to vi, with the resulting probability of not satisfying E' being $(1-p/n)^{t/4} < e^{-pt/(4n)} < e^{-pt/(8r)}$, where the last inequality follows from (16). ◀

By (17) and by Lemmas 11, 13, and 15, there is a restriction ρ satisfying all the assertions of the lemmas. Fix any such ρ.

▶ **Definition 16.** A partial assignment σ to the variables of $\mathrm{REF}^F_{s,t}$ is called an *admissible assignment* if it extends ρ and satisfies all the following conditions.

(C1) For each $(i,j) \in [s] \times [t]$, $D(i,j,\cdot,\cdot)$ (resp. $V(i,j,\cdot)$, $I(j,\cdot)$, $L(i,j,\cdot)$, $R(i,j,\cdot)$) either is set to some clause (resp. some $k \in [n]$, some $m \in [r]$, some $j' \in [t]$, some $j' \in [t]$) by σ or contains no variable that is in $\mathrm{dom}(\sigma)$.

(C2) For each $(i,j) \in [s] \times [t]$, if $L(i,j,\cdot)$ or $R(i,j,\cdot)$ is set to some $j' \in [t]$, then both $D(i,j,\cdot,\cdot)$ and $D(i-1,j',\cdot,\cdot)$ are set.

(C3) For each $(i,j) \in [s] \times [t]$, if $D(i,j,\cdot,\cdot)$ is set, then $V(i,j,\cdot)$ is set (if $i \in \{2,\dots,s\}$) or $I(j,\cdot)$ is set (if $i = 1$).

(C4) For each $(i,j) \in [s] \times [t]$, if $D(i,j,\cdot,\cdot)$ is set to a clause $C_{i,j}$, then $C_{i,j}$ is non-tautological and has at least $\min\{s-i,n\}$ many literals. If $D(i,j,\cdot,\cdot)$ is set to a clause $C_{i,j}$ with less than n literals and $V(i,j,\cdot)$ is set to some $k \in [n]$, then none of the literals of x_k is in $C_{i,j}$.

(C5) If $D(s,t,\cdot,\cdot)$ is set, it is set to the empty clause.

(C6) For each $j \in [t]$, if $D(1,j,\cdot,\cdot)$ and $I(j,\cdot)$ are set, then σ satisfies all clauses in (1) with this j.

(C7) For each $i \in \{2,\dots,s\}, j, j' \in [t]$, if $L(i,j,\cdot)$ (resp. $R(i,j,\cdot)$) is set to j' and both $V(i,j,\cdot)$, $D(i-1,j',\cdot,\cdot)$ are set, then σ satisfies all clauses in (3) (resp. (4)) with these i,j,j'.

(C8) For each $i \in \{2,\dots,s\}, j, j' \in [t]$, if $L(i,j,\cdot)$ (resp. $R(i,j,\cdot)$) is set to j' and $V(i,j,\cdot)$, $D(i,j,\cdot,\cdot)$, $D(i-1,j',\cdot,\cdot)$ are set, then σ satisfies all clauses in (5) (resp. (6)) with these i,j,j'.

(C9) For each $i \in \{2,\dots,s\}$, the binary relation $h_{\sigma,i} := \{(Z(i,j,\cdot),j') : j,j' \in [t], Z \in \{L,R\},$ and $Z(i,j,\cdot)$ is set to j' by $\sigma\}$ is a partial injection from $\{Z(i,j,\cdot) : j \in [t], Z \in \{L,R\}\}$ to $[t]$.

For the proofs of the following three lemmas, see the full version [5].

▶ **Lemma 17.** *No clause in* $\mathrm{REF}^F_{s,t} \restriction \rho$ *is falsified by any admissible assignment.*

▶ **Lemma 18.** *There is an admissible assignment.*

▶ **Lemma 19.** *Suppose that a clause E in $\Pi \restriction \rho$ is obtained by the resolution rule from clauses E_0 and E_1. Suppose further that there is an admissible assignment σ which satisfies both conditions*

(i) *every literal in E of a variable in* $\mathrm{dom}(\sigma)$ *is falsified by σ,*

(ii) *for each $Z \in \{D,V,I,R,L\}$, each Z-variable with a home pair Z-important in E is in* $\mathrm{dom}(\sigma)$.

Then there is an admissible assignment τ and $b \in \{0,1\}$ such that i and ii hold with τ in place of σ and E_b in place of E.

These three lemmas easily imply a contradiction, which concludes the proof of Theorem 8.

▶ Remark 20. If we assume $s = n+1$ in Theorem 8 (instead of assuming only $s \geq n+1$) then we can allow t to be smaller: it is enough to assume that $t \geq r^{2+\epsilon}$. This can be useful if one wants to reduce the number of variables of $\mathrm{REF}^F_{s,t}$ while keeping the lower bound of the theorem valid. The latter can be shown by making only the following modification in the proof of Theorem 8: change the definition of p to $p = s^{-1/3}t^{-a'}$ with $a' = \min\left\{\frac{1+\epsilon}{3+\epsilon}, \frac{1}{2}\right\}$, and change the definition of w to $w = s^{1/3}t^{3/5}$.

We note that if in the definition of $\mathrm{REF}^F_{s,t}$ we encode the functions determined by V- and I-variables in binary instead of in unary, the assumption $t \geq r^{3+\epsilon}$ in Theorem 8 is not necessary (and the proof of the theorem simplifies somewhat), and, in addition, the L- and R-variables can be encoded in binary too (with some further simplifications of the proof). This reduces the number of variables of $\mathrm{REF}^F_{s,t}$ in two ways, by allowing a smaller t and by using a more efficient encoding.

▶ Remark 21. Most of the obstacles our proof has to overcome are caused by the nature of the object described by $\mathrm{REF}^F_{s,t}$ and by the fact that the functions determined by V,I,L,R-variables are encoded in unary, rather than in binary. This forces us to work with several

notions of width of two kinds, and we cannot keep as an invariant of the maintained partial assignment that it falsifies all literals of a clause as we traverse the refutation (as is the case e.g. in [11]). Moreover, keeping falsified just the literals with important indices and adding some simple conditions about not directly falsifying an axiom (a method which works e.g. in [8] for the pigeonhole principle) is not enough either, because we need to be prepared to consistently answer the prover's questions about clauses situated at remote parts of the same not too small component (learnt through the L- and R-variables). This is further complicated by the need to respond by adding a fresh literal to a clause that has too few literals to make sure its width grows fast enough (such clauses originate in the component of the empty clause), and by the necessity to arrive to a weakening of a clause in F when asked how a clause on level 2 is derived; both are more difficult to meet under the unary encoding and pose specific requirements on random restrictions. Our strategy stores some useful information in the form of negating some other literals than just those with important indices in a clause, as can be seen in the hierarchy of setting of variables of different kinds in Definition 16.

5 Reflection Principle for Resolution

We express the negation of the reflection principle for Resolution by a CNF in the form of a conjunction $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{s,t}$. The only shared variables by the formulas $\mathrm{SAT}^{n,r}$ and $\mathrm{REF}^{n,r}_{s,t}$ encode a CNF with r clauses in n variables. The meaning of $\mathrm{SAT}^{n,r}$ is that the encoded CNF is satisfiable, while the meaning of $\mathrm{REF}^{n,r}_{s,t}$ is that it has a resolution refutation of s levels of t clauses. A formal definition is given next.

Formula $\mathrm{SAT}^{n,r}$ has the following variables. Variables $C(m,k,b)$, $m \in [r], k \in [n], b \in \{0,1\}$, encode clauses C_m as follows: $C(m,k,1)$ (resp. $C(m,k,0)$) means that the literal x_k (resp. $\neg x_k$) is in C_m. Variables $T(k)$, $k \in [n]$, and variables $T(m,k,b)$, $m \in [r], k \in [n], b \in \{0,1\}$, encode that an assignment to variables $x_1, \ldots, x_n$ satisfies the CNF $\{C_1, \ldots, C_r\}$. The meaning of $T(k)$ is that the literal x_k is satisfied by the assignment. The meaning of $T(m,k,1)$ (resp. $T(m,k,0)$) is that clause C_m is satisfied through the literal x_k (resp. $\neg x_k$).

We list the clauses of $\mathrm{SAT}^{n,r}$:

$$T(m,1,1) \vee T(m,1,0) \vee \ldots \vee T(m,n,1) \vee T(m,n,0) \qquad m \in [r], \quad (20)$$

$$\neg T(m,k,1) \vee T(k) \qquad m \in [r], k \in [n], \quad (21)$$

$$\neg T(m,k,0) \vee \neg T(k) \qquad m \in [r], k \in [n], \quad (22)$$

$$\neg T(m,k,b) \vee C(m,k,b) \qquad m \in [r], k \in [n], b \in \{0,1\}, \quad (23)$$

The meaning of (20) is that clause C_m is satisfied through at least one literal. The meaning of (21) and (22) is that if C_m is satisfied through a literal, then the literal is satisfied. The meaning of (23) is that if C_m is satisfied through a literal, then it contains the literal.

Variables of $\mathrm{REF}^{n,r}_{s,t}$ are the variables $C(m,k,b)$ of $\mathrm{SAT}^{n,r}$ together with all the variables of $\mathrm{REF}^{F}_{s,t}$ for some (and every) F of r clauses in n variables. That is, $\mathrm{REF}^{n,r}_{s,t}$ has the following variables:

$C(m,k,b)$	$m \in [r], k \in [n], b \in \{0,1\},$
$D(i,j,k,b)$	$i \in [s], j \in [t], k \in [n], b \in \{0,1\},$
$R(i,j,j')$ and $L(i,j,j')$	$i \in \{2,\ldots,s\}, j,j' \in [t],$
$V(i,j,k)$	$i \in \{2,\ldots,s\}, j \in [t], k \in [n],$
$I(j,m)$	$j \in [t], m \in [r].$

The clauses of $\mathrm{REF}^{n,r}_{s,t}$ are (2) - (15) of $\mathrm{REF}^{F}_{s,t}$ together with the following clauses (to replace clauses (1)):

$$\neg I(j,m) \vee \neg C(m,k,b) \vee D(1,j,k,b) \qquad j \in [t], m \in [r], k \in [n], b \in \{0,1\}, \tag{24}$$

saying that if clause $C_{1,j}$ is a weakening of clause C_m, then the former contains each literal of the latter.

We now prove the lower bound for $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{s,t}$ stated in the Introduction as Theorem 3 and restated below as Theorem 22.

▶ **Theorem 22.** *For every $c > 4$ there is $\delta > 0$ and an integer n_0 such that if n, r, s, t are integers satisfying*

$$t \geq s \geq n+1, \qquad r \geq n \geq n_0, \qquad n^c \geq t \geq r^4, \tag{25}$$

then any resolution refutation of $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{s,t}$ has length greater than 2^{n^δ} (which is exponential in the size of the formula).

Proof. Fix $c > 4$. We first observe that if Π is a resolution refutation of $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{s,t}$ and σ is a partial assignment such that its domain are all C-variables, then $\Pi \restriction \sigma$ is either a refutation of $\mathrm{REF}^{n,r}_{s,t} \restriction \sigma$, or a refutation of $\mathrm{SAT}^{n,r} \restriction \sigma$. This is because $\Pi \restriction \sigma$ is a resolution refutation and the two restricted formulas do not share any variables.

Let F be a CNF with r clauses in n variables, and let σ_F be a partial assignment such that its domain are all C-variables and σ_F evaluates them so that they describe the clauses of F. Notice that $\mathrm{REF}^{n,r}_{s,t} \restriction \sigma_F$ is $\mathrm{REF}^{F}_{s,t}$, since σ_F turns the clauses (24) into the clauses (1) (and removes the satisfied clauses). Therefore, in the case that $\Pi \restriction \sigma_F$ is a refutation of $\mathrm{REF}^{n,r}_{s,t} \restriction \sigma_F$ and F is unsatisfiable, the lower bound of Theorem 8 applies (setting $\epsilon = 1$ in that theorem, there is n_0 such that conditions (16) on n, r, s, t follow from (25)): the theorem yields some $\delta_1 > 0$ such that the length of $\Pi \restriction \sigma_F$ is at least $2^{n^{\delta_1}}$.

Let us now consider the case that $\Pi \restriction \sigma_F$ is a refutation of $\mathrm{SAT}^{n,r} \restriction \sigma_F$. Let SAT^{F} stand for $\mathrm{SAT}^{n,r} \restriction \sigma_F$. There is a substitution τ to the variables of SAT^{F} that turns the clauses of SAT^{F} into all the clauses of F together with some tautological clauses. It is defined as follows. If $\sigma_F(C(m,k,b)) = 0$, then $\tau(T(m,k,b)) = 0$. This satisfies (21) - (23) and deletes $T(m,k,b)$ from (20). If $\sigma_F(C(m,k,b)) = 1$, then (23) has been satisfied and we define $\tau(T(m,k,b)) = x^b_k$ and $\tau(T(k)) = x_k$. This choice turns (21) - (22) into a tautological clause and correctly substitutes the remaining literals of (20) to yield the m-th clause of F. Thus, if $\Pi \restriction \sigma_F$ is a refutation of $\mathrm{SAT}^{n,r} \restriction \sigma_F$, the substitution τ takes it into a not larger resolution refutation of F (since tautological clauses can be removed from any resolution refutation).

It remains to take any unsatisfiable formula F whose number of clauses is polynomially related to the number of variables and that requires resolution refutations of exponential length, e.g. the pigeonhole principle [6]. A trivial modification of F to serve also in the extreme case $r = n$ allowed by (25) will yield $\delta_2 > 0$ such that any resolution refutation of F has length greater than $2^{n^{\delta_2}}$, where n is the number of variables of F.

Setting δ to the minimum of δ_1 and δ_2 concludes the proof of the theorem. ◀

A similar proof gives Theorem 4. We restate the theorem below for convenience.

▶ **Theorem 23.** *Let $\gamma > 0$ and let $\{A_n\}_{n \geq 1}$ be a family of unsatisfiable CNFs such that A_n is in n variables, has the number of clauses polynomial in n, and has no resolution refutations of length at most 2^{n^γ}. Then there is $\delta > 0$ and a polynomial p such that $A_n \wedge \mathrm{REF}^{A_n}_{n+1,p(n)}$ has no resolution refutations of length at most 2^{n^δ} and has polynomial size* Res(2) *refutations.*

Proof. Let $p(n) \geq \max\{r^4, t_0\}$, where r is the maximum of the number of clauses of A_n and n, and t_0 is given by Theorem 8 for $\epsilon = 1$. This theorem and the assumptions on A_n give the required lower bound. To get the upper bound, start with the Res(2) refutation of $\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{n+1,p(n)}$ given by Theorem 5. Define substitutions σ_{A_n} and τ like in the proof of Theorem 22 with A_n in place of F, and observe again that $((\mathrm{SAT}^{n,r} \wedge \mathrm{REF}^{n,r}_{n+1,p(n)}) \restriction \sigma_{A_n}) \restriction \tau$ is $A_n \wedge \mathrm{REF}^{A_n}_{n+1,p(n)}$ together with some tautological clauses. ◀

References

1. Albert Atserias and María Luisa Bonet. On the automatizability of resolution and related propositional proof systems. *Information and Computation*, 189(2):182–201, 2004.
2. Albert Atserias and Moritz Müller. Automating Resolution is NP-Hard. *arXiv e-prints*, April 2019. `arXiv:1904.02991v1`.
3. María Luisa Bonet, Toniann Pitassi, and Ran Raz. On Interpolation and Automatization for Frege Systems. *SIAM J. Comput.*, 29(6):1939–1967, 2000. `doi:10.1137/S0097539798353230`.
4. Stephen A. Cook and Robert A. Reckhow. The Relative Efficiency of Propositional Proof Systems. *The Journal of Symbolic Logic*, 44(1):36–50, 1979.
5. Michal Garlík. Resolution Lower Bounds for Refutation Statements. *arXiv e-prints*, May 2019. `arXiv:arXiv:1905.12372v1`.
6. Armin Haken. The intractability of resolution. *Theoretical Computer Science*, 39(2–3):297–308, 1985.
7. Jan Krajíček, Alan Skelley, and Neil Thapen. NP search problems in low fragments of bounded arithmetic. *The Journal of Symbolic Logic*, 72(2):649–672, 2007.
8. Pavel Pudlák. Proofs as Games. *American Mathematical Monthly*, 107(6):541–550, 2000. `doi:10.2307/2589349`.
9. Pavel Pudlák. On reducibility and symmetry of disjoint NP-pairs. *Theoretical Computer Science*, 295:323–339, 2003.
10. Nathan Segerlind, Samuel R. Buss, and Russel Impagliazzo. A switching lemma for small restrictions and lower bounds for k-DNF resolution. *SIAM Journal on Computing*, 33(5):1171–1200, 2004.
11. Neil Thapen. A Tradeoff Between Length and Width in Resolution. *Theory of Computing*, 12(5):1–14, 2016.