Random Resolution Refutations[*][†]

Pavel Pudlák[1] and Neil Thapen[2]

1 **Institute of Mathematics, Czech Academy of Sciences, Prague, Czech Republic**
pudlak@math.cas.cz

2 **Institute of Mathematics, Czech Academy of Sciences, Prague, Czech Republic**
thapen@math.cas.cz

Abstract

We study the *random resolution* refutation system defined in [Buss et al. 2014]. This attempts to capture the notion of a resolution refutation that may make mistakes but is correct most of the time. By proving the equivalence of several different definitions, we show that this concept is robust. On the other hand, if $\mathbf{P} \neq \mathbf{NP}$, then random resolution cannot be polynomially simulated by any proof system in which correctness of proofs is checkable in polynomial time.

We prove several upper and lower bounds on the width and size of random resolution refutations of explicit and random unsatisfiable CNF formulas. Our main result is a separation between polylogarithmic width random resolution and quasipolynomial size resolution, which solves the problem stated in [Buss et al. 2014]. We also prove exponential size lower bounds on random resolution refutations of the pigeonhole principle CNFs, and of a family of CNFs which have polynomial size refutations in constant depth Frege.

1998 ACM Subject Classification F.4.1 Mathematical Logic, F.1.3 Complexity Measures and Classes

Keywords and phrases Proof complexity, random, resolution

Digital Object Identifier 10.4230/LIPIcs.CCC.2017.1

1 Introduction

The following system for refuting propositional CNFs was introduced in [3]. Let F be a CNF in variables $x_1, \dots, x_n$ and let $0 < \varepsilon < 1$.

▶ **Definition 1.** An *ε-random resolution distribution*, or *ε-RR distribution*, of F is a probability distribution $\mathcal{D}$ on pairs $(B_i, \Pi_i)_{i \sim \mathcal{D}}$ such that

1. for each $i \in \mathcal{D}$, B_i is a CNF in variables $x_1, \dots, x_n$ and Π_i is a resolution refutation of $F \wedge B_i$
2. for every $\alpha \in \{0,1\}^n$, $\Pr_{i \sim \mathcal{D}}[B_i \text{ is satisfied by } \alpha] \geq 1 - \varepsilon$.

The *size* and the *width* of $\mathcal{D}$ are defined respectively as the maximum size and maximum width of the refutations Π_i (if these maxima exist).

This is sound as a refutational system, in the sense that if F has an ε-RR distribution then F is unsatisfiable. To see this, consider any assignment $\alpha \in \{0,1\}^n$. Since $\varepsilon < 1$, there is at least one pair (B_i, Π_i) such that α satisfies B_i and Π_i is a resolution refutation of $F \wedge B_i$. So α cannot also satisfy F, by the soundness of resolution. The system is also complete,

[*] The full version of the paper is available as [13], https://eccc.weizmann.ac.il/report/2016/175/.

[†] Partially supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement 339691. The Institute of Mathematics of the Czech Academy of Sciences is supported by RVO:67985840.

32nd Computational Complexity Conference (CCC 2017).
Editor: Ryan O'Donnell; Article No. 1; pp. 1:1–1:10
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

since resolution is complete and we can take $\mathcal{D}$ to consist of a single pair (B, Π) where B is any tautology and Π is a (possibly exponential sized) resolution refutation of F.

On the other hand, as defined, it is not a propositional proof system in the sense of Cook and Reckhow [6], because it is defined by a semantic condition that presumably cannot be tested in polynomial time (see Proposition 10). Nevertheless it makes perfect sense to compare the complexity of proofs in it with proofs in the standard proof systems, in particular with resolution and bounded depth Frege. We prove some results in this direction in this work. Note also that the definition is particular to resolution, and we must take care if we try to generalize it. For example, if we instead define a random Frege distribution system, in which B and Π can contain arbitrary formulas, then we can trivially refute any unsatisfiable F by setting $B = \neg F$.

As with some concepts of probabilistic computation studied in computation complexity theory, one can use the linear programming duality to give an equivalent definition of the system based on probability distributions over inputs rather than over proofs (see Definition 3). This is very useful if one needs to prove lower bounds. Another essentially equivalent formulation is in terms of semantic resolution derivations. This means, roughly speaking, that instead of having an auxiliary formula that is satisfied with high probability, we consider semantic derivations with respect to a large subset of inputs, where lines in the proof are clauses. In a sense, this captures better the intuitive idea of a proof with errors.

Let us also mention that while we tend to think of the error ε as something small, there is a simple amplification lemma that allows us to shrink the error at some cost in proof size. Thus, for the questions we are interested in, without loss of generality we can take $\varepsilon = \frac{1}{2}$.

The definition was first proposed by Stefan Dantchev. Its appearance in [3] is ultimately motivated by an open problem in bounded arithmetic. We will not go into detail about the connection to bounded arithmetic in this abstract, and instead will discuss the problem in terms of constant depth Frege proof systems. One of the longstanding open problems in proof complexity is to prove (or disprove) that the set of polylogarithmic width CNFs with quasipolynomial size refutations in depth i Frege strictly increases as i increases.[1] The simplest instance of this problem is to separate $R(\log)$, which is effectively a low depth Frege system, from higher levels of the constant depth Frege hierarchy. The system $R(\log)$, introduced in [10], is an extension of resolution in which in place of literals one can use small conjunctions, of logarithmic size in the length of the proof. We can separate $R(\log)$ from weaker fragments of constant depth Frege, but not from stronger ones (see for example [14]).

The system $R(\log)$ corresponds to a particular fragment T^2_2 of bounded arithmetic. Since this problem has been notoriously open for many years, it was proposed in [3] to consider, in place of T^2_2, theories of similar strength but of a rather different nature, based on Jeřábek's approximate counting [8]. Interestingly, it turned out that although there are no proof systems associated in the usual sense with these theories, one could use random resolution to prove separation of one of them from higher fragments. It suffices to find a narrow CNF which does not have a narrow 1/2-random resolution distribution, but has a quasipolynomial size refutation in some constant depth Frege system. The problem of separating the theory from other fragments of bounded arithmetic was eventually solved by different means [2], but the problem of proving lower bounds on the width of random resolution distributions remained open.

[1] We emphasize that we are interested in this question for quasipolynomial size proofs. This matches the natural question in bounded arithmetic, and a separation for polynomial size is known [7], using a padded pigeonhole principle $\mathrm{PHP}_{(\log n)^k}$ which has short proofs in some depth i, but is such that the exponential size lower bound for PHP in depth $i-1$ gives a quasipolynomial lower bound for the padded version.

In this paper we solve this problem by proving that the propositional translation of the *coloured polynomial local search* principle CPLS, introduced in [12], which has polynomial size resolution refutations, does not have narrow 1/2-random resolution distributions. Previously, lower bounds on random resolution have only been known for treelike refutations [3] or for relatively small errors ε [11].

The proof is based on a lemma that looks like a rudimentary version of the switching lemmas used in propositional proof complexity (see the discussion at the start of Section 3). Although this does not solve the big open problem of giving a new separation in constant depth Frege, we believe that our result may pave the way to a solution. It has been conjectured that in order to separate constant depth Frege system, we need only to prove switching lemmas for certain more complicated tautologies, generalizations of CPLS (see for example [15]). Nevertheless, all attempts in this direction have failed so far because of the complexity of the associated combinatorial problems. Our proof gives us some hope that eventually it will be possible to prove such lemmas.

The full version of the paper is available as an ECCC technical report [13]. In this extended abstract we present definitions and statements of the main results. We state our theorems without detailed proofs with one exception, which is the simplest lower bound from the full paper and is intended to demonstrate our lower bound technique.

2 Basic properties and alternative definitions

We first introduce some notation. We identify CNF formulas with sets of clauses. We will use 0 (false) and 1 (true) to represent truth values. For a formula F and an assignment α of truth values to its variables, we denote by $F[\alpha]$ the truth value to which the formula is evaluated by α. If ρ is a partial assignment, we denote by F^ρ the formula obtained by substituting ρ into F and simplifying the formula (that is, replacing a conjunction by 0 if one conjunct is 0, etc.).

The *width* of a clause is the number of literals it contains. The *width* and *size* of a refutation are respectively the width of its widest clause and the total number of clauses. A k-CNF is a CNF in which every clause has width at most k.

We will often use the notation $p_1 \wedge \dots \wedge p_r \rightarrow q_1 \vee \dots \vee q_s$ to stand for the clause $\neg p_1 \vee \dots \vee \neg p_r \vee q_1 \vee \dots \vee q_s$, where $p_1, \dots, p_r, q_1, \dots, q_r$ can be any literals. In this notation the resolution rule can, for example, have the form: from $A \wedge p \rightarrow C$ and $A \wedge \neg p \rightarrow D$ conclude $A \rightarrow C \vee D$, where p is a literal, A is a conjunction of literals and C and D are clauses.

If p is a literal, we will sometimes write $p = 1$ instead of the literal p and $p = 0$ instead of the literal $\neg p$. Similarly we will write $p \neq 1$ or $p \neq 0$ to mean respectively $\neg p$ or p. If $p_1, \dots, p_r$ are literals and $\beta \in \{0,1\}^r$ we write $\bar{p} = \beta$ to stand for the conjunction $\bigwedge_{1 \leq i \leq k} p_i = \beta_i$ where each conjunct is formally either p_i or its negation, as above; and $\bar{p} \neq \beta$ to stand for the disjunction $\bigvee_{1 \leq i \leq k} p_i \neq \beta_i$.

We write $[n]$ for $\{0, \dots, n-1\}$. When we formalize combinatorial principles as CNFs, if the principle involves a function $f : [n] \rightarrow [m]$ we will often formalize f by introducing variables for its "bit-graph". That is, for each $x < n$ we introduce $\log m$ variables $(f(x))_0, \dots, (f(x))_{\log m - 1}$ representing the value of $f(x)$ in binary. For the sake of simplicity, in this situation we will assume that m is a power of 2. For $y < m$ we will write $f(x) = y$ to stand for the conjunction $\bigwedge_i (f(x))_i = \beta_i$, where $\beta \in \{0,1\}^{\log m}$ is y written in binary, and we will write $f(x) \neq y$ for the disjunction $\bigvee_i (f(x))_i \neq \beta_i$.

Because we deal with propositional refutation systems, rather than proof systems, for us the natural translation into propositional logic of a true first order principle, such as the pigeonhole principle PHP, is a family of unsatisfiable CNFs that we want to refute, rather than a family of tautologous DNFs that we want to prove. Therefore we will use the same name, PHP, for both this family of CNFs and the original principle. It should be clear from the context which is meant, and the propositional version will often be written with a size parameter, for example as PHP_n.

In the rest of this section, let F be a CNF in variables $x_1, \ldots, x_n$ and let $0 < \varepsilon < 1$. Our definition of the size of an ε-RR distribution above does not take into account the size of the sample space (that is, of the set of pairs (B, Π) appearing in $\mathcal{D}$) but one can show that the size of the sample space can be bounded, at the cost of slightly increasing the error ε. Also we can decrease the error ε at the cost of increasing the width and size.

▶ **Lemma 2.** *Suppose F has an ε-RR distribution of width w and size s. Then*

1. *it also has 2ε-RR distribution of the same size and width, in which the sample space has size $O(n/\varepsilon)$, and*
2. *for every $k \geq 1$ it also has an ε^k-RR distribution of width at most kw and size $O(s^k)$.*

We will now give two more definitions equivalent to random resolution distributions.

▶ **Definition 3.** Let Δ be a probability distribution on $\{0,1\}^n$. An *(ε, Δ)-random resolution refutation*, or *(ε, Δ)-RR refutation*, of F is a pair (B, Π) such that

1. B is a CNF in variables $x_1, \ldots, x_n$ and Π is a resolution refutation of $F \wedge B$
2. $\Pr_{\alpha \sim \Delta}[B[\alpha] = 1] \geq 1 - \varepsilon$.

This definition is in general not sound, for a fixed Δ. However, if an (ε, Δ)-RR refutation exists for *all* distributions Δ, then this is equivalent to the existence of an ε-RR distribution, as follows.

▶ **Proposition 4.** *The following are equivalent.*

1. *F has an ε-RR distribution of width w and size s.*
2. *F has an (ε, Δ)-RR refutation of width w and size s for every distribution Δ on $\{0,1\}^n$.*

Proof. One direction is just an averaging argument. The other is a consequence of the minimax theorem. ◀

The following generalization of this will be useful for proving lower bounds.

▶ **Proposition 5.** *Proposition 4 still holds if we allow Δ to range over distributions on partial assignments, rather than total assignments. In this case we change item 2 in Definition 3 to* $\Pr_{\rho \sim \Delta}[B[\rho] = 0] \leq \varepsilon$.

Semantic derivations were introduced in [9]. We will use the special case defined by clauses.

▶ **Definition 6.** Let $\mathcal{A} \subseteq \{0,1\}^n$ be a nonempty set of truth assignments. We say that a formula C is a *semantic consequence over $\mathcal{A}$* of formulas $C_1, \ldots, C_r$, written $C_1, \ldots, C_r \vDash^{\mathcal{A}} C$, if every assignment in $\mathcal{A}$ that satisfies $C_1, \ldots, C_r$ also satisfies C.

A *semantic resolution refutation of F over $\mathcal{A}$* is a sequence Π of *clauses*, ending with the empty clause, in which every clause either belongs to F or is a semantic consequence over $\mathcal{A}$ of at most two earlier clauses.

▶ **Definition 7.** Let Δ be a probability distribution on $\{0,1\}^n$. An *(ε, Δ)-semantic refutation* of F is a pair $(\mathcal{A}, \Pi)$ such that

1. Π is a semantic refutation of F over $\mathcal{A}$, and
2. $\Pr_{\alpha \sim \Delta}[\alpha \in \mathcal{A}] \geq 1 - \varepsilon$.

▶ **Proposition 8.** *If F has an (ε, Δ)-RR refutation of width w and size s, then it also has an (ε, Δ)-semantic resolution refutation of width $\leq w$ and size $\leq s$.*

In the opposite direction, if F has an (ε, Δ)-semantic refutation of width w and size s, then it also has an (ε, Δ)-RR refutation of width $O(w)$ and size at most $O(sw^2)$.

We show that random 3-CNFs with sufficiently high density have small RR distributions, while as is well-known, they only have exponentially large resolution refutations [5].

▶ **Proposition 9.** *A random 3-CNF F with n variables and $64n$ clauses has a $1/2$-RR distribution of constant width and constant size with probability exponentially close to 1.*

Proof. With exponentially high probability, F has the property that no single assignment satisfies more than a fraction $15/16$ of its clauses. We define a distribution as follows: choose a clause C_i of the form $y_1 \vee y_2 \vee y_3$ from F uniformly at random, let the auxiliary formula B_i be the CNF $\neg y_1 \wedge \neg y_2 \wedge \neg y_3$, and let Π_i be the three-step derivation of the empty clause from C_i and B_i. Then with high probability this is a 15/16-RR distribution for F, which can be amplified to a 1/2-RR distribution using Lemma 2. ◀

This gives a separation of narrow random resolution from resolution in one direction (Theorem 16 will give the opposite direction). We can also prove such a separation for an explicit sequence of CNFs, the *retraction weak pigeonhole principle* (see [4, 8]) that asserts that there is no pair of functions $f : [2n] \to [n]$ and $g : [n] \to [2n]$ such that $g(f(x)) = x$ for all $x < n$.

We can now address the natural question of whether random resolution can be presented as a standard propositional proof system in the sense of Cook and Reckhow [6], or at least whether it can be polynomially simulated by such a system. Because we want to compare other systems with random resolution, we adapt the definition to refutation systems – this makes no difference to the result, since any proof system can be considered as a refutation system and vice versa. The essential property of Cook and Reckhow's definition is that one can test the correctness of refutations in polynomial time, that is, that the binary relation "Π is a refutation of F" is decidable in deterministic polynomial time. The other two properties, soundness and completeness, are satisfied by random resolution.

In order to state our question formally, we must say which object we choose to represent a refutation in random resolution, and what polynomial simulation means. We will consider 1/2-RR distributions in which all samples have the same weight. Such a distribution can be written down simply as a list of pairs (B_i, Π_i), and by Lemma 2 we do not lose anything important if we only consider 1/2-RR distributions in this form. Polynomial simulation of refutation systems can be defined in our situation, where correctness may not be decidable in polynomial time, in essentially the same way as for standard refutation systems.

▶ **Proposition 10.** *If $\mathbf{P} \neq \mathbf{NP}$, then $1/2$-RR cannot be polynomially simulated by any Cook-Reckhow refutation system, and in particular is not itself a Cook-Reckhow refutation system.*

Proof. This is a corollary of the PCP Theorem, which can be stated as follows (see Theorem 11.9 in [1]): there exists a polynomial time computable function g and a constant $\delta < 1$ such that for every CNF formula F, $g(F)$ is a 3-CNF formula such that

1. if F is satisfiable, then $g(F)$ is also satisfiable
2. otherwise, every assignment satisfies at most a fraction δ of the clauses of $g(F)$.

Using the construction from the proof of Lemma 9, this implies that if F is unsatisfiable then $g(F)$ has a δ-RR distribution, which furthermore is constructable in polynomial time.

The error δ can be reduced to $1/2$ by Lemma 2 and, again, this can be done in polynomial time. Let h denote the polynomial time computable function h that from a given unsatisfiable CNF formula F produces a $1/2$-RR distribution that refutes $g(F)$.

Suppose that $1/2$-RR can be polynomially simulated by a refutation system given by a polynomial time binary relation R and let f be the simulation. Then we can test whether F is satisfiable by computing $R(f(h(F), g(F)), g(F))$. ◀

3 Lower bounds for the bit pigeonhole principle

We present the first of our three main lower bounds on RR distributions. Before going into details, we outline the basic structure that the proofs will follow.

To prove width lower bounds on a $1/2$-RR distribution for a CNF F, we use Proposition 5 to convert the distribution into a $(1/2, \mathcal{R})$-RR refutation (B, Π) with respect to a distribution $\mathcal{R}$ on partial assignments (we will use the terms "restriction" and "partial assignment" interchangeably). The crucial thing is to choose the distribution $\mathcal{R}$ carefully.

The ideal would be that there are many restrictions ρ from $\mathcal{R}$ which make the auxiliary formula B true, thus making it vanish and leaving us with a resolution refutation for which we already have a lower bound. To this end we use a sort of rudimentary version of the switching lemma, which we call a *fixing lemma* (a different lemma in each case, because it depends on the formula F). Intuitively this shows that, with reasonably high probability, ρ fixes the value of B to either true or false. From the definition of a $(1/2, \mathcal{R})$-RR refutation we know that $B^\rho = 0$ with probability at most a half, so we can conclude that many restrictions ρ make B true.

However, in practice it is not possible to achieve the ideal that ρ makes B true. Instead we only ask that the restricted formula B^ρ cannot be falsified by any "legal" extension $\sigma \supseteq \rho$. What counts as a legal extension depends on F – for example, for the pigeonhole principle it will be a partial assignment that encodes a matching. The definition is chosen so so that we can both prove the fixing lemma and then prove a width lower bound on Π by an adversary argument, in which the adversary only works with legal extensions of ρ.

The proof of a fixing lemma should, in principle, be a special case of a proof of a switching lemma, since we are essentially switching a CNF to a decision tree of height 0, or to a trivial DNF. However in the one case we consider in which a switching lemma is known, for the (non-bit) pigeonhole principle, we do not use it directly, but rather prove our own fixing lemma. One reason is that the usual lemma works with *syntactic* transformations of formulas and does not seem to guarantee that our *semantic* condition on B, that B is satisfied with high probability, is preserved. For the CPLS formula in the next section, there is unlikely to be any traditional switching lemma. This is because, understood very broadly, such a lemma would imply strong size lower bounds on CPLS in constant depth Frege, while we know that CPLS already has polynomial size refutations in resolution.

We continue with our lower bound proof for the bit pigeonhole principle. Let $n = 2^k$. BPHP_n is contradictory CNF asserting that a function f is an injection from $[n+1]$ to $[n]$. It has variables $(f(x))_j$ for each $x < n+1$ and $j < k$, for the jth bit of the value of $f(x)$, and consists of clauses

$$f(x) \neq y \vee f(x') \neq y$$

for all $x < x' < n+1$ and all $y < n$, using the bit-graph notation described in Section 2.

In our proof, we will only consider partial assignments in which, for every x, either all or none of the variables $(f(x))_j$ are set. We identify such assignments with the corresponding partial functions from $n+1$ pigeons to n holes.

Given a probability p, define the distribution $\mathcal{R}_p$ of partial injections ρ from $[n+1]$ into $[n]$ as follows: choose the domain of ρ by putting each pigeon into the domain independently at random with probability $1-p$, then choose uniformly at random from all possible partial injections with this domain (if all $n+1$ pigeons get put into the domain, we just take ρ to be empty). For the rest of the proof, set $p = n^{-2/3}$ and $w = n^{1/4}$.

▶ **Lemma 11** (Fixing Lemma). *Let n be sufficiently large. Suppose B is a w-CNF such that* $\Pr[B^\rho = 0] \leq 1/2$. *Then*

$$\Pr[\textit{there exists a partial injection } \sigma \supseteq \rho \textit{ with } B^\sigma = 0] \leq 3/4.$$

Proof. Let S be the set of $\rho \in \mathcal{R}_p$ for which there exists a partial injection $\sigma \supseteq \rho$ which falsifies B. Partition S into the set S^0 of restrictions which falsify B, and the set S^1 of restrictions which do not falsify B themselves, but which have an extension to a partial injection which falsifies B. We know $\Pr[S^0] \leq 1/2$, so it remains to bound the size of S^1.

Consider any $\rho \in S^1$. No clause in B is falsified by ρ, but there must be at least one clause which is falsified in some partial injection $\sigma \supseteq \rho$. Let C be the first such clause and let σ be such an extension of ρ falsifying it. The literals in C appear in some fixed order. Let x be the first pigeon mentioned in C which is not in the domain of ρ, and let $i < w$ be the position in C at which the first variable from pigeon x appears. Let σ' be σ restricted to the pigeons in the domain of ρ together with pigeon x, that is, $\sigma' = \rho \cup \{\langle x, \sigma(x)\rangle\}$.

Define a function θ on S^1 by $\theta : \rho \mapsto (\sigma', i)$, where σ' and i are chosen as above. Then θ is an injection, because we can first recover C from $\theta(\rho)$ as the first clause of B which is falsified in some extension of σ' to a partial injection; then we can recover x as the pigeon associated with the variable at position i in C; and finally we can recover ρ from σ' by unsetting pigeon x.

If a restriction ρ sets $m > 0$ pigeons, then the probability of ρ is

$$\Pr[\rho] = (1-p)^m p^{n+1-m} \frac{(n-m)!}{n!}.$$

Hence $\Pr[\sigma']/\Pr[\rho] = (1-p)/p(n-m)$. By the Chernoff bound, the number $n-m$ of unset pigeons is smaller than $2pn$ with exponentially high probability in n. Let S_{bad} be the set of restrictions for which this bound fails, so that $\Pr[\sigma']/\Pr[\rho] > (1-p)/2p^2n > 1/4p^2n$ for $\rho \in S^1 \setminus S_{\text{bad}}$. Partition $S^1 \setminus S_{\text{bad}}$ into subsets $S_0, \ldots, S_{w-1}$ according to the second component i of θ. On each S_i, the first component θ_1 of θ is an injection from $\mathcal{R}_p$ to $\mathcal{R}_p$ which increases probability by at least $1/4p^2n$. Therefore

$$\Pr[\theta_1[S_i]] = \sum_{\rho \in S_i} \Pr[\theta_1(\rho)] > \frac{1}{4p^2n} \sum_{\rho \in S_i} \Pr[\rho] = \frac{1}{4p^2n} \Pr[S_i].$$

Since $\Pr[\theta_1[S_i]] \leq 1$ we can conclude that $\Pr[S_i] < 4p^2n$, and hence that $\Pr[S^1 \setminus S_{\text{bad}}] < 4p^2nw = 4n^{-1/12}$. Since $\Pr[S_{\text{bad}}]$ is also exponentially small, the result follows. ◀

▶ **Theorem 12.** BPHP_n *has no $1/2$-RR distribution of width $w = n^{1/4}$.*

Proof. We will show that BPHP_n has no $(1/2, \mathcal{R}_p)$-RR refutation with this width. Suppose for a contradiction that there is such a refutation (B, Π), where B is the auxiliary w-CNF which is false in $\mathcal{R}_p$ with probability at most $1/2$.

By Lemma 11, for a random $\rho \in \mathcal{R}_p$ with probability at least $1/4$ there is no extension of ρ to a partial injection which falsifies any clause from B. Thus by the Chernoff bound we can fix one such restriction ρ which also leaves at least $pn/2 = n^{1/3}/2$ holes free.

Now consider any clause C in the refutation Π. Suppose we have a partial injection $\sigma \supseteq \rho$ that falsifies C, and suppose that C is derived by resolution from clauses $D \vee v$ and $E \vee \neg v$, where v is a variable $(f(x))_j$ for some $x < n+1$ and $j < k$. Since $|C| \leq n^{1/4}$ we can find $\sigma' \subseteq \sigma$ that falsifies C and sets at most $n^{1/4}$ pigeons not set in ρ. Hence we can find a free hole to assign to pigeon x, thus extending σ' to a partial injection which falsifies either $D \vee v$ or $E \vee \neg v$.

In this way, working inductively up through the refutation, we can find a partial injection $\sigma \supseteq \rho$ which falsifies some initial clause. But this is a contradiction, since a partial injection cannot falsify any clause from BPHP_n, and by our choice of ρ a partial injection extending ρ cannot falsify any clause from B. ◀

We show size lower bounds by combining the argument of Theorem 12 with a standard application of random restrictions to remove clauses mentioning many pigeons from Π.

▶ **Theorem 13.** BPHP_n *has no* $1/2$*-RR distribution of subexponential size, that is, of size less than* 2^{n^ε} *for some* $\varepsilon > 0$.

4 Main lower bounds and separations

4.1 A separation of resolution from narrow RR

The *coloured polynomial local search* principle (CPLS) was introduced in [12]. The propositional version of it was studied in [16]. We refer to those two papers for more on the principle, and only remark here that it is a good candidate for proving separations of this kind because it is in some sense "complete" among narrow CNFs with short resolution refutations [12], while at the same time its combinatorial structure is simple enough that we are able to come up with useful random restrictions. We take our definitions from [16].

Consider a leveled directed graph whose nodes consist of all pairs (i, x) from $[a] \times [b]$. We refer to (i, x) as *node x on level i*. If $i < a-1$, this node has a single neighbour in the graph, node $f_i(x)$ on level $i+1$. Every node in the graph is coloured with some set of colours from $[c]$. CPLS expresses that the following three sentences cannot all be true at once.

1. Node 0 on level 0 has no colours.
2. For every node x on every level $i < a-1$, if the neighbour $f_i(x)$ of x on level $i+1$ has any colour y, then x also has colour y.
3. Every node x on the bottom level $a-1$ has at least one colour, $u(x)$.

We will express this principle as a family of propositional contradictions. Let a be any natural number and let b and c be powers of two. We will define a CNF formula $\mathrm{CPLS}_{a,b,c}$, in the following propositional variables.

- For each $i < a$, $x < b$ and $y < c$, there is a variable $G_i(x, y)$, expressing whether colour y is present at node (i, x).
- For each $i < a$, $x < b$ and $j < \log b$, there is a variable $(f_i(x))_j$, standing for the jth bit of the value of $f_i(x)$.
- For each $x < b$ and $j < \log c$, there is a variable $(u(x))_j$, standing for the jth bit of the value of $u(x)$.

▶ **Definition 14.** The formula $\mathrm{CPLS}_{a,b,c}$ consists of the following three sets of clauses, which we will call Axioms 1, 2 and 3:

Axiom 1. For each $y < c$, the clause $\neg G_0(0, y)$.

Axiom 2. For each $i < a - 1$, each pair $x, x' < b$ and each $y < c$, the clause

$$f_i(x) = x' \wedge G_{i+1}(x', y) \to G_i(x, y) .$$

Axiom 3. For each $x < b$ and each $y < c$, the clause

$$u(x) = y \to G_{a-1}(x, y) .$$

▶ **Proposition 15.** $\mathrm{CPLS}_{a,b,c}$ *has polynomial size resolution refutations.*

Proof. For $i < a$, let M_i be the set of clauses $\{\bigvee_{y<c} G_i(x, y) : x < b\}$ expressing that every node at level i has a colour. We can derive M_{a-1} from Axiom 3. Then repeatedly using Axiom 2 we can derive M_{a-2}, M_{a-3}, etc. Once we have M_0 we can derive a contradiction from Axiom 1. For more detail see [16]. ◀

▶ **Theorem 16.** *For all sufficiently large n, the formula $\mathrm{CPLS}_{n,n,\lfloor n^{1/7} \rfloor}$ does not have a $1/2$-RR distribution of width $n^{1/8}$.*

4.2 A separation of constant-depth Frege from RR

We exhibit a narrow CNF which requires exponential size 1/2-RR distributions but which, unlike the pigeonhole principle, has polynomial size refutations in constant depth Frege, in fact in Res(2). Here Res(2) is an extension of resolution in which clauses may contain conjunctions of pairs of literals (see [10]).

The formula is CPLS^2, a variant of CPLS. For each i, x, y, instead of the single variable $G_i(x, y)$ it has two variables $G_i^0(x, y)$ and $G_i^1(x, y)$. To express that colour y is present at node (i, x) we now use the conjunction $G_i^0(x, y) \wedge G_i^1(x, y)$.

As before, the formula expresses that node $(0, 0)$ has no colours; that every colour present at node $(i + 1, f_i(x))$ is also present at node (i, x); and that colour $u(x)$ is present at node $(a - 1, x)$.

▶ **Proposition 17.** $\mathrm{CPLS}^2_{a,b,c}$ *has polynomial size Res(2) refutations.*

▶ **Theorem 18.** *For all sufficiently large n, the formula $\mathrm{CPLS}^2_{n,n,\lfloor n^{1/7} \rfloor}$ does not have a $1/2$-RR distribution of size $\leq 2^{n^{1/17}}$.*

4.3 Lower bounds for the pigeonhole principle

We now consider the usual formalization of the pigeonhole principle, rather than the bit-graph version. The CNF formula PHP_n has variables p_{ij} for $i \in [n + 1]$ and $j \in [n]$ and consists of clauses $\bigvee_{j=1}^n p_{ij}$ for all $i \in [n + 1]$ and $\neg p_{ij} \vee \neg p_{i'j}$ for all $i, i' \in [n + 1]$ with $i \neq i'$ and all $j \in [n]$.

▶ **Theorem 19.** PHP_n *has no $1/2$-RR distribution of size less than $2^{\Omega(n^{1/12})}$.*

Acknowledgments. We would like to thank Pavel Hrubeš and Jan Krajíček for discussions and valuable suggestions.

References

1 Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009. doi:10.1017/CBO9780511804090.

2 Albert Atserias and Neil Thapen. The ordering principle in a fragment of approximate counting. *ACM Trans. Comput. Logic*, 15(4):29:1–11, 2014. doi:10.1145/2629555.

3 Samuel R. Buss, Leszek Aleksander Kołodziejczyk, and Neil Thapen. Fragments of approximate counting. *The Journal of Symbolic Logic*, 79(2):496–525, 2014. doi:10.1017/jsl.2013.37.

4 Mario Chiari and Jan Krajíček. Witnessing functions in bounded arithmetic and search problems. *The Journal of Symbolic Logic*, 63(03):1095–1115, 1998. doi:10.2307/2586729.

5 Vašek Chvátal and Endre Szemerédi. Many hard examples for resolution. *Journal of the ACM*, 35(4):759–768, 1988. doi:10.1145/48014.48016.

6 Stephen Cook and Robert Reckhow. The relative efficiency of propositional proof systems. *The Journal of Symbolic Logic*, 44(1):36–50, 1979. doi:10.2307/2273702.

7 Russell Impagliazzo and Jan Krajíček. A note on conservativity relations among bounded arithmetic theories. *Mathematical Logic Quarterly*, 48(3):375–377, 2002. doi:10.1002/1521-3870(200204)48:3<375::AID-MALQ375>3.0.CO;2-L.

8 Emil Jeřábek. On independence of variants of the weak pigeonhole principle. *Journal of Logic and Computation*, 17(3):587–604, 2007. doi:10.1093/logcom/exm017.

9 Jan Krajíček. Interpolation theorems, lower bounds for proof systems, and independence results for bounded arithmetic. *The Journal of Symbolic Logic*, 62(02):457–486, 1997. doi:10.2307/2275541.

10 Jan Krajíček. On the weak pigeonhole principle. *Fundamenta Mathematicae*, 170:123–140, 2001. doi:10.4064/fm170-1-8.

11 Jan Krajíček. A feasible interpolation for random resolution, 2016. Preprint arXiv:1604.06560. arXiv:1604.06560.

12 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. doi:10.2178/jsl/1185803628.

13 Pavel Pudlák and Neil Thapen. Random resolution refutations, 2016. Electronic Colloquium on Computational Complexity, TR16-175. URL: https://eccc.weizmann.ac.il/report/2016/175/.

14 Alexander Razborov. Pseudorandom generators hard for k-DNF resolution and polynomial calculus resolution. *Annals of Mathematics*, 181(2):415–472, 2015. doi:10.4007/annals.2015.181.2.1.

15 Alan Skelley and Neil Thapen. The provably total search problems of bounded arithmetic. *Proceedings of the London Mathematical Society*, 103(1):106–138, 2011. doi:10.1112/plms/pdq044.

16 Neil Thapen. A tradeoff between length and width in resolution. *Theory of Computing*, 12(5):1–14, 2016. doi:10.4086/toc.2016.v012a005.