An Exponential Lower Bound on OBDD Refutations for Pigeonhole Formulas

Olga Tveretina[1], Carsten Sinz[1], Hans Zantema[2,3]

[1] Institute for Theoretical Computer Science,
Karlsruhe University
Am Fasanengarten 5, D-76131 Karlsruhe, Germany
`olga@ira.uka.de, sinz@ira.uka.de`
[2] Department of Computer Science, TU Eindhoven, P.O. Box 513,
5600 MB Eindhoven, The Netherlands
`h.zantema@tue.nl`
[3] Institute for Computing and Information Sciences, Radboud University
Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands

Abstract. Haken proved that every resolution refutation of the pigeon hole formula has at least exponential size. Groote and Zantema proved that a particular OBDD computation of the pigeon hole formula has an exponential size. Here we show that any arbitrary OBDD refutation of the pigeon hole formula has an exponential size, too: we prove that the size of one of the intermediate OBDDs is at least $\Omega(1.025^n)$.

1 Introduction

The pigeonhole principle, also known as Dirichlet's box principle states that n holes can hold at most n objects with one object to a hole. The propositional formulas describing this principle were introduced by Cook and Reckhow in 1979 [5]. The formula is a CNF parameterized by n. It is unsatisfiable, but after removing any single clause it becomes satisfiable, it is thus minimally unsatisfiable.

The formula has a very simple shape, a meta argument for unsatisfiability is easily given, but standard techniques for proving unsatisfiability automatically run out of time for quite small values of n. Therefore, this formula is a good benchmark to test the efficiency of an approach for deciding (un)satisfiability.

Also, on the theoretical side, it is the basis of many interesting results. A landmark result is that of Haken [7], who proved that the length of any resolution refutation of the pigeon hole formula is at least exponential in n. Surprisingly, Cook proved that it admits a polynomial refutation based on extended resolution [4].

An *Ordered Binary Decision Diagram* (OBDD) is a data structure that is used to represent Boolean functions [2, 14]. OBDDs have some interesting properties: they provide compact and canonic representations of Boolean functions, and there are efficient algorithms for performing logical operations on OBDDs. As a result, OBDDs have been successfully applied to a wide variety of tasks, particularly in VLSI design and CAD verification [10, 11]. There are some less

well-known applications as fault tree analysis [13], Bayesian reasoning and product configuration [8].

As a propositional proof system BDDs were studied, e.g., by Atserias *et al.* [1]. The authors introduce a very general proof system based on constraint propagation. BDDs are a special case of this proof system. Their proof system has four rules: *axiom*, *join*, *projection*, and *weakening*. The first two rules, *axiom* and *join*, correspond to an application of the BDD *apply* operator. *Projection* and *weakening* are introduced to reduce the size of intermediate BDDs. It was shown that the BDD proof system containing all four rules is strictly stronger than resolution [1] but it is still exponential [9].

In our paper, by the BDD proof of a formula φ we mean the computation of the corresponding BDD using the *apply*-operation, i.e. in terms of the above proof system from [1], we allow only two rules, namely *axiom* and *join*. If the formula contains n Boolean connectives, then the BDD construction requires exactly n calls of *apply*, and the exponential blow up of the size of the proof is caused by the expansion of the size of the arguments.

In [6] it was proved that a particular BDD computation of the pigeonhole formula is at least exponential. On the other hand, it was proved in [3] that the pigeonhole formula admits a polynomial size BDD refutation in a setting including existential quantification (i.e. including the *projection* rule).

In this paper we prove that, based on the notion of BDD refutation along the lines of [3] containing the classical ingredients of BDD computation, but excluding existential quantification, we have an exponential lower bound for the size of BDD refutations of the pigeonhole formula. This is much stronger than the result from [6]: there, the only computation considered first computes the conjunction of all positive clauses, then the conjunction of all negative clauses, and finally the conjunction of these two. In our setting, the clauses of the pigeonhole formula may be processed in any arbitrary order. We show that in any BDD refutation proof some of the intermediate BDDs has size at least exponential in n. As a consequence we state that the gap between polynomial and exponential in the BDD refutation framework for pigeonhole formula is caused by the rule for existential quantification.

We start with preliminaries in Section 2. In Section 3 we prove an exponential lower bound on BDD refutations for the pigeonhole formula. Finally, Section 4 contains conclusions.

2 Preliminaries

We consider propositional formulas in *Conjunctive Normal Form* (CNFs). Basic blocks for building CNFs are propositional variables that take the values false or true. The set of propositional variables is denoted by Var. A literal is either a variable x or its negation $\neg x$. A clause is a disjunction of literals, and a CNF is a conjunction of clauses. In the following, for convenience, we consider clauses as sets of variables, and a CNF as a set of clauses. By $\mathsf{Cls}(\varphi)$ we denote the set

of clauses contained in a CNF φ and by $\mathsf{Var}(\varphi)$ we denote the set of variables contained in the CNF φ.

2.1 Binary Decision Diagrams

A Binary Decision Diagram (BDD) is a a rooted, directed, acyclic graph, which consists of decision nodes and two terminal nodes 0 and 1. Each decision node is labeled by a propositional variable from Var and has two child nodes called low child and high child. The edge from a node to a low (high) child represents an assignment of the variable to 0 (1). Such a BDD is called *ordered* if different variables appear in the same order on all paths from the root. Therefore, BDDs assume that there is a total order $\prec$ on the set of variables Var.

A BDD is said to be *reduced* if the following two rules have been applied to its graph: 1) merge isomorphic subgraphs; 2) eliminate any node whose two children are isomorphic.

Reduced BDDs have the following property: For a fixed order $\prec$ on the set of variables, every propositional formula φ is uniquely represented by a reduced BDD $\mathsf{B}(\varphi, \prec)$, and two formulas φ and ψ are equivalent if and only if $\mathsf{B}(\varphi, \prec) = \mathsf{B}(\psi, \prec)$.

Given a propositional formula φ and an order on variables $\prec$, we define the size of a BDD $\mathsf{B}(\varphi, \prec)$ representing φ with respect to $\prec$ as the number of its internal nodes and denote it by $\mathsf{size}(\mathsf{B}(\varphi, \prec))$.

We give a definition of a BDD refutation adapting the definition from [3].

Definition 1 (BDD refutation). *Given a total order on variables $\prec$, a BDD refutation of an unsatisfiable CNF φ is a sequence of BDDs*

$$\mathsf{B}_1(\varphi_1, \prec), \ldots, \mathsf{B}_n(\varphi_n, \prec)$$

such that $\mathsf{B}_n(\varphi_n, \prec)$ is a BDD representing the constant false *and for each $\mathsf{B}_i(\varphi_i, \prec)$, $1 \leq i \leq n$, exactly one of the following holds.*

- *$\mathsf{B}_i(\varphi_i, \prec)$ represents one of the clauses $C \in \varphi$;*
- *there are BDDs $\mathsf{B}_{i'}(\varphi_{i'}, \prec)$ and $\mathsf{B}_{i''}(\varphi_{i''}, \prec)$ such that $1 \leq i' < i'' < i$ and $\varphi_i = \varphi_{i'} \wedge \varphi_{i''}$.*

We say that n is the length of the BDD refutation. The size of the BDD refutation is defined as $\sum_{i=1}^{n} \mathsf{size}(\mathsf{B}_i(\varphi_i, \prec))$.

When it is convenient, instead of $\mathsf{B}(\varphi, \prec)$ we write $\mathsf{B}(\varphi)$ or just B. By $\mathsf{Cls}(\mathsf{B}(\varphi))$ we mean the set of clauses and by $\mathsf{Var}(\mathsf{B}(\varphi))$ the set of variables contained in $\mathsf{B}(\varphi)$.

The size of the minimal OBDD representing a propositional formula F for a given order on variables $\prec$ is described by the following structure theorem [12]. We use $\mathbb{B}^n \to \mathbb{B}$ to denote the set of Boolean functions with domain $\{0,1\}^n$ and range $\{0,1\}$.

Theorem 1. *Suppose for a given formula φ the following holds:*

- $|\mathsf{Var}(\varphi)| = n$;
- $\prec$ is a total order on the set of variables $\mathsf{Var}(\varphi)$;
- $x_1, \ldots, x_k$ are the smallest k elements with respect to $\prec$ for some $k < n$;
- $A \subseteq \{1, \ldots, k\}$;
- $z = (z^1, \ldots, z^k) \in \mathbb{B}^k$.
- For all distinct $\overrightarrow{x}_1, \overrightarrow{x}_2 \in \mathbb{B}^k$ such that $x_1^i = x_2^i = z^i$ for all $i \notin A$ there exists a $\overrightarrow{y} \in \mathbb{B}^{n-k}$ such that $\varphi(\overrightarrow{x}_1, \overrightarrow{y}) \neq \varphi(\overrightarrow{x}_2, \overrightarrow{y})$.

Then the size of the BDD $\mathsf{B}(\varphi, \prec)$ is at least $2^{|A|}$.

The proof of the lower bound presented in Section 3 is based on Theorem 1. However, in order to obtain a lower bound we still have to solve some combinatorial problems.

2.2 The pigeonhole formula

The pigeonhole principle states that n holes can hold at most n objects with one object in a hole. It can be formulated as a set of clauses as follows.

$$\mathsf{PC}_n = \bigwedge_{i=1}^{n+1} \left(\bigvee_{j=1}^{n} P_{ij} \right)$$

$$\mathsf{NC}_n = \bigwedge_{\substack{1 \leq i < j \leq n+1 \\ 1 \leq k \leq n}} (\neg P_{ik} \vee \neg P_{jk})$$

$$\mathsf{PHP}_n = \mathsf{PC}_n \wedge \mathsf{NC}_n$$

Now we introduce notations that will be used in the rest of the paper. Let

$$\mathsf{PC}_n^* = \bigwedge_{i=1}^{n} \left(\bigvee_{j=1}^{n} P_{ij} \right) .$$

Hence, PC_n^* contains the first n clauses of PC_n. We represent PC_n^* as a matrix of variables with n rows and n columns (the clause $\bigvee_{j=1}^{n} P_{ij}$ corresponds to the i-th row). We denote this matrix by P, and the set of columns of P by P^c (the elements of P^c are indices j, $1 \leq j \leq n$). For each row in P there is a corresponding clause in PC_n^* and vice versa, therefore we will refer to a row as a clause, and to a set of rows as a set of clauses.

For a given total order on variables $\prec$, we define $S_\prec$ as the set containing the $\lfloor n^2/2 \rfloor$ smallest elements of $\mathsf{Var}(\mathsf{PC}_n^*)$ with respect to ordering $\prec$, and let $S_\succeq = \mathsf{Var}(\mathsf{PC}_n^*) \backslash S_\prec$. Moreover, we define

$$S_\prec^* = \{P_{ij} \in \mathsf{Var}(\mathsf{PHP}_n) \mid P_{ij} \preceq \max S_\prec^*\},$$

and

$$S_\succeq^* = \mathsf{Var}(\mathsf{PHP}_n) \backslash S_\prec^*.$$

Note that $S_\prec \cup S_\succeq = \mathsf{Var}(\mathsf{PC}_n^*)$ and $S_\prec^* \cup S_\succeq^* = \mathsf{Var}(\mathsf{PHP}_n)$. The sets $S_\prec$ and $S_\succeq$ are defined in such a way that the difference between the sizes of these sets is at most one, but, in contrary, this does not hold for the sets $S_\prec^*$ and $S_\succeq^*$.

For each BDD B_i in a BDD refutation of PHP_n we define

$$S_\prec^i = S_\prec^* \cap \mathsf{Var}(\mathsf{B}_i(\varphi))$$

and

$$S_\succeq^i = \mathsf{Var}(\mathsf{B}_i(\varphi)) \backslash S_\succeq^*.$$

Moreover, we define

$$\mathsf{Cls}^{neg}(\mathsf{B}(\varphi)) = \mathsf{Cls}(\mathsf{B}(\varphi)) \cap \mathsf{Cls}(\mathsf{NC}_n)$$

and

$$\mathsf{Cls}^{pos}(\mathsf{B}(\varphi)) = \mathsf{Cls}(\mathsf{B}(\varphi)) \cap \mathsf{Cls}(\mathsf{PC}_n).$$

3 The main result

In this section we prove a lower bound for BDD refutations on pigeonhole formulas. The proof of our lower bound is inspired by the proof of a lower bound of a particular BDD refutation given in [6].

First we introduce technical lemmas that we use to prove the main result.

Lemma 1. *Consider a matrix $M = \{m_{ij}\}$, $1 \le i \le n$, $1 \le j \le n$. Let the matrix entries be colored equally white and black, i.e. the difference between the number of white entries and the number of black entries is at most one. Let $m = \lfloor cn \rfloor$ for $c = \frac{1}{2} - \frac{1}{4}\sqrt{2} \approx 0.146$. Then at least one of the following holds.*

- *One can choose m rows, and in every of these rows a white and a black entry, such that all these $2m$ entries are in different columns.*
- *One can choose m columns, and in every of these columns a white and a black entry, such that all these $2m$ entries are in different rows.*

Proof. Starting by the given matrix repeat the following process as long as possible.

> Choose a row in the matrix containing both a white and a black entry. Remove both the column containing the white entry and the column containing the black entry. Also remove the chosen row.

Assume this repetition stops after k steps. If $k \ge m$ the first property of the lemma holds and we are done. In the remaining case the remaining matrix consists of $n - k$ rows of $n - 2k$ entries, where every row either only consists of white entries or only of black entries. Assume that at least $n - 2m$ of these rows are totally black. Using $k < m$ we conclude that the number of black entries in this remaining matrix is at least

$$(n - 2m)(n - 2k) > (n - 2m)^2 \ge \frac{1}{2}n^2,$$

contradicting the assumption that at most half of the entries are black (possibly up to one). So at least $n - k - (n - 2m) = 2m - k > m$ of these rows are totally white. By symmetry also at least m of these rows are totally white. As the length of these rows are $n - k > n - m > m$, the second property of the lemma is easily fulfilled. $\qquad\square$

By fine-tuning the argument the constant c in Lemma 1 can be improved. We conjecture that it also holds for $c = 1 - \frac{1}{2}\sqrt{2} \approx 0.293$. Choosing the $n \times n$ matrix in which the left upper $k \times k$-square is black for $k \approx \frac{n}{\sqrt{2}}$ and the rest is white, one observes that this value will be sharp. As our main result involves an exponential lower bound, we do not focus on the precise optimal value of c.

The final OBDD representing the pigeonhole formula is a terminal node 0. Hence, we have to show that for an arbitrary order on variables and for an arbitrary oder of proceeding clauses of PHP_n there is an intermediate OBDD that has a size exponential in n.

The following lemma generalizes a well-known fact about binary trees claiming the existence of subtrees with a weight lying between a and 2a (for any definition of "weight" as a sum of the weights of its leaves).

Lemma 2. *Let C be a finite set, $R \subseteq C$ with $|R| \geq 2$, and $B_1, \ldots, B_l \subseteq C$ a sequence with:*

1. $B_l = C$
2. *For each B_i ($1 \leq i \leq l$), either $B_i = \emptyset$, $B_i = \{c\}$ for $c \in C$, or $B_i = B_j \cup B_k$ for some j, k with $j < k < i$.*

Then, for each a with $\frac{1}{|R|} < a \leq \frac{1}{2}$, there is a $j < l$ such that

$$a|R| \leq |B_j \cap R| < 2a|R| \ .$$

Proof. We give a proof by contradiction. Suppose, for each B_j, either

$$|B_j \cap R| < a|R| \qquad \text{or} \qquad |B_j \cap R| \geq 2a|R| \ .$$

As $B_l \cap R = C \cap R = R$, the inequality $|B_l \cap R| \geq 2a|R|$ holds for the final element B_l of the sequence. On the other hand, for singletons $B_j = \{c\}$, we have $|B_j \cap R| = 0 < a|R|$ for $c \notin R$, and $|B_j \cap R| = 1 < a|R|$ for $c \in R$, as $a > 1/|R|$. Moreover, for $B_i = \emptyset$, $|B_i \cap R| < a|R|$ obviously holds. Following now the predecessors of B_l (via the construction by set union) in the sequence B_i backwards, we finally arrive at an index k for which the following holds:

- $|B_k \cap R| \geq 2a|R|$, and
- $B_k = B_{k'} \cup B_{k''}$, where $|B_{k'} \cap R| < a|R|$ and $|B_{k''} \cap R| < a|R|$.

As $B_k \cap R = (B_{k'} \cup B_{k''}) \cap R = (B_{k'} \cap R) \cup (B_{k''} \cap R)$, and thus $|B_k \cap R| \leq |B_{k'} \cap R| + |B_{k''} \cap R| < 2a|R|$, we arrive at a contradiction to $|B_k \cap R| \geq 2a|R|$. $\qquad\square$

Lemma 3. *Suppose* $B_1, \ldots, B_l$ *is a BDD refutation of* PHP_n *and* $R \subseteq \mathsf{Cls}(\mathsf{PC}_n)$ *with* $|R| > 4$. *Then there is an* $i < l$ *such that*

$$|R|/4 \leq |\mathsf{Cls}(B_i) \cap R| < 2|R|/4 \ .$$

Proof. Follows from Lemma 2. □

Definition 2. *Let* $B_1, \ldots, B_l$ *is a BDD refutation of* PHP_n. *For each* $i \leq l$ *define* J_i *as the set of columns from* P^c *as follows:*

$$J_i = \{j \in P^c \mid \exists a, b : \neg P_{aj} \vee \neg P_{bj} \in \mathsf{Cls}(B_i), \ P_{aj} \in S_\prec, \ and \ P_{bj} \in S_\succeq\}.$$

Lemma 4. *Suppose* $B_1, \ldots, B_l$ *is a BDD refutation of* PHP_n *for a total order on variables* $\prec$, *and* $P' \subseteq P^c$ *with* $|P'| > 4$. *Then there is an* $i < l$ *such that*

$$|P'|/4 \leq |J_i \cap P'| < |P'|/2.$$

Proof. Follows from Lemma 2, using $C = P^c$, $R = P'$, $a = 1/4$, and $J_1, \ldots, J_l$ for the sequence $(B_i)_{1 \leq i \leq l}$, for which the precondition of Lemma 2 holds, as is easily checked. □

The following theorem defines a lower bound exponential in n on a BDD refutation of PHP_n.

Theorem 2. *For every order* $\prec$ *on the set of variables, both time and space complexity of each BDD proof of* PHP_n *is* $\Omega(1.025^n)$.

Proof. Let $n > 34$, and $B_1, \ldots, B_l$ be a BDD refutation of PHP_n. We prove that for an arbitrary total order on variables $\prec$ there is $i \leq l$ such that

$$\mathsf{size}(B_i) \geq 2^{n(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4}.$$

Since $2^{(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4} > 1.025$ we have

$$\mathsf{size}(B_i) > 1.025^n$$

and the theorem holds.

We apply Lemma 1 to the matrix representing PC_n^*. Then one of the following holds.

- There is a set of $\lfloor n(\frac{1}{2} - \frac{1}{4}\sqrt{2}) \rfloor$ rows (we denote this set by R) and there is a set of $2\lfloor n(\frac{1}{2} - \frac{1}{4}\sqrt{2}) \rfloor$ entries (we denote this set by S^R) such that the following holds:
  - For each $r \in R$ there are $P_{ra}, P_{rb} \in S^R$ such that $P_{ra} \in S_\prec$ and $P_{rb} \in S_\prec$.
  - For distinct $P_{ab}, P_{cd} \in S^R$, $b \neq d$.

We define
$$R^m = \mathsf{Cls}(B_i) \cap R \ .$$

As $n > 34$, $|R| = \lfloor n(\frac{1}{2} - \frac{1}{4}\sqrt{2})\rfloor \geq 5$, and we can apply Lemma 3. Thus we know that there is an $i < l$ such that

$$|R|/4 \leq |R^m| < 2|R|/4.$$

We get
$$2|R^m| + 1 \leq |R|.$$

For each row $r \in R^m$ we fix an entry that is in the set $S_\prec$. We collect these elements in the set A. For each row $r \in R^m$ we also fix an entry that is in $S_\succeq$ and collect these elements in the set Y. Let

$$R^c = \{j \mid \exists i : P_{ij} \in A \cup Y\}.$$

Taking into account that $2|R^m| + 1 \leq |R|$ we compute

$$|\mathsf{Cls}^{pos}(B_i)| \leq (n+1) - (|R| - |R^m|) \leq (n+1) - ((2|R^m|+1) - |R^m|) = n - |R^m|.$$

We denote $\overline{R^m} = \mathsf{Cls}^{pos}(B_i) \backslash R^m$. By definition $R^m \subseteq \mathsf{Cls}^{pos}(B_i)$. Hence, we obtain
$$|\overline{R^m}| = |\mathsf{Cls}^{pos}(B_i)| - |R^m| \leq n - 2|R^m|.$$

Let $J = n - |R^c|$. Since we have chosen the set of rows R^m as satisfying the conditions of Lemma 1, we get $|R^c| = 2|R^m|$ and

$$J = n - 2|R^m|$$

and
$$|\overline{R^m}| \leq |J|.$$

For each $C \in \overline{R^m}$ we fix one variable and collect these variables in the set X that the following holds. For distinct $P_{ab}, P_{cd} \in X$, $b \neq d$. This is possible because $|\overline{R^m}| \leq |J|$.
We define $X_\prec = S_\prec \cap X$ and $X_\succeq = S_\succeq \cap X$.
We apply Lemma 1 on
$$k = |S^i_\prec|.$$

For $j = 1, \ldots, k$ we define $z_j = 1$ if $z_j \in A$ or $z_j \in X_\prec$, otherwise we define $z_j = 0$.
Choose $\overrightarrow{x}, \overrightarrow{x}'$ satisfying $\overrightarrow{x} \neq \overrightarrow{x}'$ and $x_j = x'_j = z_j$ for all $z_j \notin A$. Then there is j' such that $x_{j'} \neq x'_{j'}$.
Let $\overrightarrow{y} = (y_{k+1}, \ldots, y_q)$, where $q = |\mathsf{Var}(B_i)|$, be the vector defined by $y_j = 1$ if $y_j \in X_\succeq$ and $y_j = 0$ for all $y_j \in S^i_\succeq \backslash (Y \cup X_\succeq)$. If $y_j \in Y$ then we choose $y_j = 0$ if it is in the same row as x_i and $y_j = 1$ otherwise.
Hence, the subset of clauses represented by B_i evaluates to x_i for the assignment $(\overrightarrow{x}, \overrightarrow{y})$ and to x'_i for the assignment $(\overrightarrow{x}', \overrightarrow{y})$.

The size of the set A is at least $n(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4$ by construction. Hence, by Lemma 1, we conclude that

$$\mathsf{size}(B_i) \geq 2^{|A|} \geq 2^{|R|/4} \geq 2^{n(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4}$$

for sufficiently large n.

- There is a set of $\lfloor n(\frac{1}{2} - \frac{1}{4}\sqrt{2}) \rfloor$ columns (we denote this set by Q) and there is a set containing $2\lfloor n(\frac{1}{2} - \frac{1}{4}\sqrt{2}) \rfloor$ entries (we denote this set by S^Q) such that the following holds:
  - Each column $q \in Q$ contains one element from the set $S_\prec$ and one element from the set $S_\succeq$.
  - For distinct $m_{ab}, m_{cd} \in S^Q$, $a \neq c$.

  Suppose $m = \lfloor n(\frac{1}{2} - \frac{1}{4}\sqrt{2}) \rfloor$.
  Let

  $$Q^c = \{j \mid \exists a, b : \neg P_{aj} \vee \neg P_{bj} \in \mathsf{Cls}(B_i) \ \& \ P_{aj} \in S_\prec \ \& \ P_{bj} \in S_\succeq\}.$$

  Then, by Lemma 4, there is B_i for $i < l$ such that

  $$m/4 \leq |Q^c| < m/2.$$

  For each column in Q^c we fix one entry that is in the set $S_\prec$. We collect these elements in A. For each column in Q^c we also fix one entry that is in the set $S_\prec$. We collect these elements in Y.
  Let
  $$Q^r = \{i \mid \exists j : P_{ij} \in A \cup Y\}.$$

  Let
  $$\overline{Q^c} = Q \backslash Q^c.$$

  Then
  $$\overline{Q^c} > m/2.$$

  For each $j \in \overline{Q^c}$ we fix $P_{a_jj}, P_{b_jj} \in S^Q$, where $P_{a_jj} \in S_\prec$ and $P_{b_jj} \in S_\succeq$. We collect P_{a_jj} in $X_\prec$ and we collect P_{b_jj} in $X_\succeq$ for all $j \in \overline{Q^c}$.
  We define
  $$\overline{Q^r} = \{a \mid \exists b : P_{ab} \in X_\prec \cup X_\succeq\}.$$

  By Lemma 1 all entries collected in $\overline{Q^r}$ are from different rows. Hence, we obtain
  $$|\overline{Q^r}| = 2|\overline{Q^c}|.$$

  Taking into account that $\overline{Q^c} > m/2$ we get
  $$\overline{Q^r} > 2m/2 = m$$

  and since $\overline{Q^r}$ is a natural number we get
  $$\overline{Q^r} \geq m + 1.$$

We denote
$$Q^* = \mathsf{Cls}^{pos}(B_i) \backslash \overline{Q^r}.$$
The set of clauses $\mathsf{Cls}^{pos}(B_i)$ can contain an arbitrary subset of clauses from PC^n, i.e.
$$1 \leq |\mathsf{Cls}^{pos}(B_i)| \leq n + 1.$$
We take into account that $|\overline{Q^r}| \geq m + 1$ and compute
$$|\mathsf{Cls}^{pos}| \leq (n + 1) - |\overline{Q^r}| \leq (n + 1) - (m + 1) = n - m.$$
We define $J = \{j \mid \exists a : P_{aj} \in \mathsf{Var}(\mathsf{PHP}_n) \ \& \ j \notin Q\}$. Then
$$|J| = n - |Q| = n - m.$$
Therefore,
$$|Q^*| \leq |J|.$$
For each row $r \in Q^*$ we fix one entry and collect these entries in the set X. We require that the entries collected in X satisfy the following properties.

- r contains at least one entry such that this entry is in one of the columns of J;
- each column is J contains at most one fixed entry;

Since $|Q^*| \leq |J|$, there is such a set X.

We denote $X_{\prec}^i = S_{\prec}^i \cap X_{\prec}$; $X_{\succeq}^i = S_{\succeq}^i \cap X_{\succeq}$; $X^* = S_{\prec}^i \cap X$ and $X^{**} = S_{\succeq}^i \cap X$.

We apply Lemma 1 on $k = |S_{\prec}^i|$.

For $j = 1, \ldots, k$ we define $z_j = 1$ if $z_j \in A$ or $z_j \in X_{\prec}^i$ or $z_j \in X^*$, and we define $z_j = 0$ in all other cases.

Choose $\overrightarrow{x}, \overrightarrow{x}'$ satisfying $\overrightarrow{x} \neq \overrightarrow{x}'$ and $x_j = x_j' = z_j$ for all $z_j \notin A$. Then there is $j' \notin \{1, \ldots, k\}$ such that $x_{j'} \neq x_{j'}'$.

Let $\overrightarrow{y} = (y_{k+1}, \ldots, y_q)$, where $q = |\mathsf{Var}(B_i)|$, be the vector defined by $y_j = 1$ for all $y_j \in X_{\succeq}^i$, $y_j \in X_{\succeq}^{**}$.

For $y_j \in Y$ we define $y_j = 1$ if it is in the same column as $x_{j'}$ and $y_j = 0$ otherwise.

We choose $y_j = 0$ in all other cases.

Therefore, for each row there is an entry that is assigned to 1 and for each column except j' and columns from the set $\overline{Q^c}$ there is at most one entry assigned to 1. If a column t is contained in the set $\overline{Q^c}$ then two entries in this column can be assigned to 1. By construction, for each column t in the set $\overline{Q^c}$ there is a clause $\neg P_{s't} \vee \neg P_{s''t} \notin \mathsf{Cls}(B_i)$. Therefore, assigning $P_{s't}$ and $\neg P_{s''t}$ simultaniously to 1 does not violate the satisfiability of the subformula represented by B_i.

Hence, the subset of clauses represented by B_i evaluates to x_i for the assignment $(\overrightarrow{x}, \overrightarrow{y})$ and to x_i' for the assignment $(\overrightarrow{x}', \overrightarrow{y})$.

The size of the set A is at least $n(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4$ by construction. Hence, by Lemma 1, we conclude that

$$\mathsf{size}(B_i) \geq 2^{|A|} \geq 2^{|R|/4} \geq 2^{n(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4}$$

for sufficiently large n. □

4 Conclusions

In this paper we have shown that the BDD proof system containing two rules, *axiom* and *join*, has lower bounds exponential in n on refutations for the pigeonhole formulas. On the other hand, it has been shown in [3] that BDD refutations of the same formulas can be given of polynomial size if the *projection* rule is added to the above two rules.

Therefore, the result presented in this paper implies that the projection rule is responsible for the gap between polynomial and exponential, just like the rule in extended resolution is responsible for a similar gap.

References

1. A. Atserias, P. Kolaitis, and M. Vardi. Constraint propagation as a proof system. In *Principles and Practice of Constraint Programming CP 2004*, volume 3258 of *LNCS*, pages 77–91, 2004.
2. R. Bryant. Graph-based algorithms for boolean function manipulation. *IEEE Transactions on Computers*, 8(C-35):677–691, 1986.
3. W. Chén and W. Zhang. A direct construction of polynomial-size OBDD proof of pigeon hole problem. *Information Processing Letters*, 109(10):472–477, 2009.
4. S. Cook. A short proof of the pigeon hole principle using extended resolution. *ACM SIGACT News*, 8(4):28–32, 1976.
5. S. Cook and R. Reckhow. The relative efficiency of propositional proof systems. *Journal of Symbolic Logic*, 44(1):36–50, 1979.
6. J. F. Groote and H. Zantema. Resolution and binary decision diagrams cannot simulate each other polynomially. *Discrete Applied Mathematics*, 130(2):157–171, 2003.
7. A. Haken. The intractability of resolution. *Theoretical Computer Science*, 39:297–308, 1985.
8. O. Heling-Tveretina and G. Provan. On approximate knowledge compilation with weighted decision diagrams. In *IC-AI*, pages 470–475, 2007.
9. J. Krajíček. An exponential lower bound for a constraint propagation proof system based on ordered binary decision diagrams. *The Journal of Symbolic Logic*, 73(1):227–237, 2008.
10. Ch. Meinel and T. Theobald. *Algorithms and Data Structures in VLSI-Design: OBDD - Foundations and Applications*. Springer-Verlag, Berlin, Heidelberg, New York, 1998.
11. Shin-ichi Minato. *Binary Decision Diagrams and Applications for VLSI CAD (VLSI, Computer Architecture and Digital Signal Processing)*. Springer-Verlag Gmbh, 1996.
12. D. Sieling and I. Wegener. Nc-algorithms for operations on binary decision diagrams. *Parallel Processing Letters*, 3:3–12, 1993.
13. R. Sinnamon and J. Andrews. Fault tree analysis and binary decision diagrams. In *International Symposium on Product Quality and Integrity*, pages 215–222, 1996.
14. I. Wegener. *Branching programs and binary decision diagrams: theory and applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.