

# An Exponential Lower Bound on OBDD Refutations for Pigeonhole Formulas

Olga Tveretina\*

Institute for Theoretical Computer Science  
Karlsruhe University  
Am Fasanengarten 5, D-76131 Karlsruhe  
Germany  
olga@ira.uka.de

Carsten Sinz\*

Institute for Theoretical Computer Science  
Karlsruhe University  
Am Fasanengarten 5, D-76131 Karlsruhe  
Germany  
sinz@ira.uka.de

Hans Zantema

Department of Computer Science  
TU Eindhoven, The Netherlands  
Institute for Computing and Information Sciences  
Radboud University, The Netherlands  
h.zantema@tue.nl

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

## 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), also referred as a reduced OBDD (ROBDD) or just a BDD, is a data structure that is used to represent Boolean functions [2, 12].

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 [9]. There are some less well-known applications as fault tree analysis [11], Bayesian reasoning and product configuration.

---

\*This work was supported in part by the "Concept for the Future" of Karlsruhe Institute of Technology within the framework of the German Excellence Initiative.

As a propositional proof system OBDDs were studied, e.g., by Atserias *et al.* [1]. The authors introduce a very general proof system based on constraint propagation. OBDDs 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 OBDD *apply* operator. *Projection* and *weakening* are introduced to reduce the size of intermediate OBDDs. It was shown that the OBDD proof system containing all four rules is strictly stronger than resolution [1] but it is still exponential [8].

In our paper, by the OBDD proof of a formula  $\varphi$  we mean the computation of the corresponding OBDD 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 OBDD 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 OBDD 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 OBDD refutation in a setting including existential quantification (i.e. including the *projection* rule).

In this paper we prove that, based on the notion of OBDD refutation along the lines of [3] containing the classical ingredients of OBDD computation, but excluding existential quantification, we have an exponential lower bound for the size of OBDD 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 OBDD refutation proof some of the intermediate OBDDs has size at least exponential in  $n$ . As a consequence we state that the gap between polynomial and exponential in the OBDD 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 OBDD 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  $\text{Var}$ . 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  $\text{Cls}(\varphi)$  we denote the set of clauses contained in a CNF  $\varphi$  and by  $\text{Var}(\varphi)$  we denote the set of variables contained in the CNF  $\varphi$ .

### 2.1 Ordered Binary Decision Diagrams

An Ordered Binary Decision Diagram (OBDD) is 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  $\text{Var}$  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 structure is called *ordered* because different variables appear in the same order on all paths from the root. Therefore, OBDDs assume that there is a total order  $\prec$  on the set of variables  $\text{Var}$ .

A OBDD 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. In our paper we

consider only reduced OBDDs.

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

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

**Definition 2.1 (OBDD refutation)** *Given a total order on variables  $\prec$ , a OBDD refutation of an unsatisfiable CNF  $\varphi$  is a sequence of OBDDs  $B_1(\varphi_1, \prec), \dots, B_n(\varphi_n, \prec)$  such that  $B_n(\varphi_n, \prec)$  is a OBDD representing the constant false and for each  $B_i(\varphi_i, \prec)$ ,  $1 \leq i \leq n$ , exactly one of the following holds.*

- (Axiom)  $B_i(\varphi_i, \prec)$  represents one of the clauses  $C \in \varphi$ ;
- (Join) there are OBDDs  $B_{i'}(\varphi_{i'}, \prec)$  and  $B_{i''}(\varphi_{i''}, \prec)$  such that  $1 \leq i' < i'' < i$  and  $\varphi_i = \varphi_{i'} \wedge \varphi_{i''}$ .

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

When it is convenient, instead of  $B(\varphi, \prec)$  we write  $B(\varphi)$  or just  $B$ . If a OBDD  $B$  represents a CNF  $\varphi$  then by  $\text{Cls}(B)$  we mean  $\text{Cls}(\varphi)$  and by  $\text{Var}(B)$  we mean  $\text{Var}(\varphi)$ .

The size of the minimal OBDD representing a propositional formula  $\varphi$  for a given order on variables  $\prec$  is described by the following structure theorem [10, 6]. We use  $\mathbb{B} = \{0, 1\}$  to denote the set of Boolean constants.

**Theorem 2.2** *Suppose for a given formula  $\varphi$  the following holds:*

- $|\text{Var}(\varphi)| = n$ ;
- $\prec$  is a total order on the set of variables  $\text{Var}(\varphi)$ ;
- $x_1, \dots, x_k$  are the smallest  $k$  elements with respect to  $\prec$  for some  $k < n$ ;
- $A \subseteq \{1, \dots, k\}$ ;
- $z = (z^1, \dots, z^k) \in \mathbb{B}^k$ .
- For all distinct  $\vec{x}_1, \vec{x}_2 \in \mathbb{B}^k$  such that  $x_i^1 = x_i^2 = z^i$  for all  $i \notin A$  there exists a  $\vec{y} \in \mathbb{B}^{n-k}$  such that  $\varphi(\vec{x}_1, \vec{y}) \neq \varphi(\vec{x}_2, \vec{y})$ .

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

The proof of the lower bound presented in Section 3.5 is based on Theorem 2.2. 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.

$$\text{PC}_n = \bigwedge_{i=1}^{n+1} \left( \bigvee_{j=1}^n P_{ij} \right), \quad \text{NC}_n = \bigwedge_{\substack{1 \leq i < j \leq n+1 \\ 1 \leq k \leq n}} (\neg P_{ik} \vee \neg P_{jk})$$

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

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

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

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  $P$ . 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  $\text{Var}(PC_n^*)$  with respect to ordering  $\prec$ , and let  $S_{\succeq} = \text{Var}(PC_n^*) \setminus S_{\prec}$ . Moreover, we define

$$S_{\prec}^* = \{P_{ij} \in \text{Var}(PHP_n) \mid P_{ij} \preceq \max S_{\prec}\},$$

and

$$S_{\succeq}^* = \text{Var}(PHP_n) \setminus S_{\prec}^*.$$

Note that  $S_{\prec} \cup S_{\succeq} = \text{Var}(PC_n^*)$  and  $S_{\prec}^* \cup S_{\succeq}^* = \text{Var}(PHP_n)$ . 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 OBDD  $B_i$  in a OBDD refutation of  $PHP_n$  we define

$$S_{\prec}^i = S_{\prec}^* \cap \text{Var}(B_i) \text{ and } S_{\succeq}^i = \text{Var}(B_i) \setminus S_{\prec}^*.$$

Moreover, we define

$$\text{Cls}^{neg}(B_i) = \text{Cls}(B_i) \cap \text{Cls}(NC_n) \text{ and } \text{Cls}^{pos}(B_i) = \text{Cls}(B_i) \cap \text{Cls}(PC_n).$$

### 3 The main result

The proof of our lower bound is inspired by the proof of a lower bound of a particular OBDD refutation given in [6].

**Lemma 3.1** *Consider a matrix  $M = \{m_{ij}\}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq n$ . 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  $k \geq m$  the first property of the lemma holds and we are done. In the remaining case the remaining matrix consists of  $n - k$  rows with  $n - 2k$  entries in each row, 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

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

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.

By fine-tuning the argument the constant  $c$  in Lemma 3.1 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 pigeonhole formula is an unsatisfiable CNF and, hence, the OBDD representing  $\text{PHP}_n$  is just a terminal node 0. Therefore, we have to show that for an arbitrary order on variables and an arbitrary way to combine clauses there is an intermediate OBDD of a size exponential in  $n$ . We start our proof by the simple observations describing some properties of intermediate OBDDs. And 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 3.2** *Let  $C$  be a finite set,  $R \subseteq C$  with  $|R| \geq 2$ , and  $B_1, \dots, B_l \subseteq C$  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

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

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  $|B_k \cap R| \geq 2a|R|$ .

**Lemma 3.3** *Suppose  $B_1, \dots, B_l$  is a BDD refutation of  $\text{PHP}_n$  and  $R \subseteq \text{Cls}(\text{PC}_n)$  with  $|R| > 4$ . Then there is an  $i < l$  such that*

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

**Proof** Follows from Lemma 3.2.

Let  $B_1, \dots, B_l$  is a BDD refutation of  $\text{PHP}_n$ . For each  $i \leq l$  define  $J_i$  as the set of columns from  $P^c$  as follows:

$$J_i = \{j \in \{1, \dots, n\} \mid \exists a, b : \neg P_{aj} \vee \neg P_{bj} \in \text{Cls}(B_i), P_{aj} \in S_{\prec}, \text{ and } P_{bj} \in S_{\succeq}\}.$$

**Lemma 3.4** *Suppose  $B_1, \dots, B_l$  is a BDD refutation of  $\text{PHP}_n$  for a total order on variables  $\prec$ , and  $P' \subseteq \{1, \dots, n\}$  with  $|P'| > 4$ . Then there is an  $i < l$  such that*

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

**Proof** Follows from Lemma 3.2, using  $C = \{1, \dots, n\}$ ,  $R = P'$ ,  $a = 1/4$ , and  $J_1, \dots, J_l$  for the sequence  $(B_i)_{1 \leq i \leq l}$ , for which the precondition of Lemma 3.2 holds, as is easily checked.

**Theorem 3.5** *For every order  $\prec$  on the set of variables, the size of each OBDD refutation of  $\text{PHP}_n$  is  $\Omega(1.025^n)$ .*

**Proof** Let  $n > 34$ , and  $B_1, \dots, B_l$  be a OBDD refutation of  $\text{PHP}_n$ . We prove that for an arbitrary total order on variables  $\prec$  there is an  $i \leq l$  such that  $\text{size}(B_i) \geq 2^{n(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4}$ . Since  $2^{(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4} > 1.025$  we have  $\text{size}(B_i) > 1.025^n$  and the theorem holds.

We apply Lemma 3.1 to the matrix representing  $\text{PC}_n^*$ . 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  $P_{rb} \in S_{\succeq}$ .
  - For distinct  $P_{ab}, P_{cd} \in S^R$ ,  $b \neq d$ .

We define

$$R^i = \text{Cls}(B_i) \cap R .$$

As  $n > 34$ ,  $|R| = \lfloor n(\frac{1}{2} - \frac{1}{4}\sqrt{2}) \rfloor \geq 5$ , and we can apply Lemma 3.3. Thus we know that there is an  $i < l$  such that

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

We get

$$2|R^i| + 1 \leq |R|.$$

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

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

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

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

We denote  $\overline{R^i} = \text{Cls}^{\text{pos}}(B_i) \setminus R^i$ . By definition  $R^i \subseteq \text{Cls}^{\text{pos}}(B_i)$ . Hence, we obtain

$$|\overline{R^i}| = |\text{Cls}^{\text{pos}}(B_i)| - |R^i| \leq n - 2|R^i|.$$

Let  $J = n - |R^j|$ . Since we have chosen the set of rows  $R^i$  as satisfying the conditions of Lemma 3.1, we get  $|R^j| = 2|R^i|$  and

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

and

$$|\overline{R^i}| \leq |J|.$$

For each  $C \in \overline{R^i}$  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  $|\overline{R^i}| \leq |J|$ .

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

We apply Lemma 2.2 on

$$k = |S_{\prec}^i|.$$

For  $j = 1, \dots, k$  we define  $z_j = 1$  if  $z_j \in A$  or  $z_j \in X_{\prec}$ , otherwise we define  $z_j = 0$ .

Choose  $\vec{x}, \vec{x}'$  satisfying  $\vec{x} \neq \vec{x}'$  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  $\vec{y} = (y_{k+1}, \dots, y_q)$ , where  $q = |\text{Var}(B_i)|$ , be the vector defined by  $y_j = 1$  if  $y_j \in X_{\succeq}$  and  $y_j = 0$  for all  $y_j \in S_{\succeq}^i \setminus (Y \cup X_{\succeq})$ . 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  $x_{j'}$  for the assignment  $(\vec{x}, \vec{y})$  and to  $x'_{j'}$  for the assignment  $(\vec{x}', \vec{y})$ .

The size of the set  $A$  is at least  $n(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4$  by construction. Hence, by Lemma 2.2, we conclude that  $\text{size}(B_i) \geq 2^{|A|} \geq 2^{|R|/4} \geq 2^{n(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4}$  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:
  - For each  $q \in Q$  there are  $P_{aq}, P_{bq} \in S^Q$  such that  $P_{aq} \in S_{\prec}$  and  $P_{bq} \in S_{\succeq}$ .
  - For distinct  $P_{ab}, P_{cd} \in S^Q$ ,  $a \neq c$ .

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

Let

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

Then, by Lemma 3.4, there is  $B_i$  for  $i < l$  such that

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

For each  $j \in Q^c$  we choose  $\neg P_{aj} \vee \neg P_{bj}$  such that  $\neg P_{aj} \vee \neg P_{bj} \in \text{Cls}(B_i)$ , where  $P_{aj} \in S_{\prec}$  and  $P_{bj} \in S_{\succeq}$ . We collect  $P_{aj}$  in  $A$  and  $P_{bj}$  in  $Y$ .

Let

$$Q^r = \{a \mid \exists j : P_{aj} \in A \cup Y\}.$$

Let

$$\overline{Q^c} = Q \setminus Q^c.$$

Then

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

For each  $j \in \overline{Q^c}$  we fix  $P_{a_j j}, P_{b_j j} \in S^Q$ , where  $P_{a_j j} \in S_{\prec}^*$  and  $P_{b_j j} \in S_{\succeq}^*$ . We collect  $P_{a_j j}$  in  $X_{\prec}$  and we collect  $P_{b_j j}$  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 3.1 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

$$Q^* = \text{Cls}^{pos}(B_i) \setminus \overline{Q^r}.$$

The set of clauses  $\text{Cls}^{pos}(B_i)$  can contain an arbitrary subset of clauses from  $\text{PC}^n$ , i.e.

$$1 \leq |\text{Cls}^{pos}(B_i)| \leq n + 1.$$

We take into account that  $|\overline{Q^r}| \geq m + 1$  and compute

$$|\text{Cls}^{pos}| \leq (n + 1) - |\overline{Q^r}| \leq (n + 1) - (m + 1) = n - m.$$

We define  $J = \{j \mid \exists a : P_{aj} \in \text{Var}(\text{PHP}_n) \ \& \ j \notin Q\}$ . 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  $W$ . 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 in  $J$  contains at most one fixed entry.

Since  $|Q^*| \leq |J|$ , there is such a set  $W$ . We denote  $X_{\prec}^i = S_{\prec}^i \cap X_{\prec}$ ;  $X_{\succeq}^i = S_{\succeq}^i \cap X_{\succeq}$ ;  $W_{\prec} = S_{\prec}^i \cap W$  and  $W_{\succeq} = S_{\succeq}^i \cap W$ . We apply Lemma 2.2 on

$$k = |S_{\prec}^i|.$$

For  $j = 1, \dots, k$  we define  $z_j = 1$  if  $z_j \in A \cup X_{\prec}^i \cup W_{\prec}$ , and we define  $z_j = 0$  in all other cases. We choose  $\vec{x}, \vec{x}'$  satisfying  $\vec{x} \neq \vec{x}'$  and  $x_j = x'_j = z_j$  for all  $z_j \notin A$ . Then there is  $j' \notin \{1, \dots, k\}$  such that  $x_{j'} \neq x'_{j'}$ . Let

$$\vec{y} = (y_{k+1}, \dots, y_q),$$

where  $q = |\text{Var}(B_i)|$ , be the vector defined by  $y_j = 1$  for all  $y_j \in X_{\succeq}^i, y_j \in W_{\succeq}$ . 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  $\neg P_{s't} \vee \neg P_{s''t} \notin \text{Cls}(B_i)$ . Therefore, assigning  $P_{s't}$  and  $\neg P_{s''t}$  simultaneously 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  $x_{j'}$  for the assignment  $(\vec{x}, \vec{y})$  and to  $x'_{j'}$  for the assignment  $(\vec{x}', \vec{y})$ .

The size of the set  $A$  is at least  $n(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4$  by construction. Hence, by Lemma 2.2, we conclude that  $\text{size}(B_i) \geq 2^{|A|} \geq 2^{|R|/4} \geq 2^{n(\frac{1}{2} - \frac{1}{4}\sqrt{2})/4}$  for sufficiently large  $n$ .



## 4 Conclusions

This paper improved an earlier result in which the use of the OBDD proof system is restricted, in a way that the proof must follow the structure of a given formula. We have shown that the OBDD 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 OBDD 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 & M. Vardi (2004): *Constraint Propagation as a Proof System*. In: *Principles and Practice of Constraint Programming CP 2004*, LNCS 3258. pp. 77–91.
- [2] R. Bryant (1986): *Graph-based algorithms for Boolean function manipulation*. *IEEE Transactions on Computers* 8(C-35), pp. 677–691.
- [3] W. Chén & W. Zhang (2009): *A direct construction of polynomial-size OBDD proof of pigeon hole problem*. *Information Processing Letters* 109(10), pp. 472–477.
- [4] S. Cook (1976): *A short proof of the pigeon hole principle using extended resolution*. *ACM SIGACT News* 8(4), pp. 28–32.
- [5] S. Cook & R. Reckhow (1979): *The Relative Efficiency of Propositional Proof Systems*. *Journal of Symbolic Logic* 44(1), pp. 36–50.
- [6] J. F. Groote & H. Zantema (2003): *Resolution and binary decision diagrams cannot simulate each other polynomially*. *Discrete Applied Mathematics* 130, pp. 157–171.
- [7] A. Haken (1985): *The Intractability of Resolution*. *Theoretical Computer Science* 39, pp. 297–308.
- [8] J. Krajíček (2008): *An exponential lower bound for a constraint propagation proof system based on ordered binary decision diagrams*. *The Journal of Symbolic Logic* 73(1), pp. 227–237.
- [9] Ch. Meinel & T. Theobald (1998): *Algorithms and Data Structures in VLSI-Design: OBDD - Foundations and Applications*. Springer-Verlag, Berlin, Heidelberg, New York.
- [10] D. Sieling & I. Wegener (1993): *NC-Algorithms for Operations on Binary Decision Diagrams*. *Parallel Processing Letters* 3, pp. 3–12.
- [11] R. Sinnamón & J. Andrews (1996): *Fault tree analysis and binary decision diagrams*. In: *International Symposium on Product Quality and Integrity*. pp. 215–222.
- [12] I. Wegener (2000): *Branching programs and binary decision diagrams: theory and applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.