Electronic Colloquium on Computational Complexity, Report No. 52 (2000)

ECCC

# Approximability and Nonapproximability by Binary Decision Diagrams
## (Extended Abstract)

Beate Bollig* and Ingo Wegener*

FB Informatik, LS2, Univ. Dortmund,
44221 Dortmund, Germany
bollig,wegener@ls2.cs.uni-dortmund.de

**Abstract.** Many BDD (binary decision diagram) models are motivated by CAD applications and have led to complexity theoretical problems and results. Motivated by applications in genetic programming Krause, Savický, and Wegener (1999) have shown that for the inner product function $IP_n$ and the direct storage access function $DSA_n$ all functions which approximate them on considerably more than half of the inputs need exponential $\pi$-OBDD size for most variable orderings $\pi$. In this paper, the results of Krause, Savický, and Wegener are generalized to more general BDD models like $k$-IBDDs and BP1s (read-once branching programs). Furthermore, the size of OBDDs for functions which approximate a function $f_n$ is compared with the size of randomized OBDDs with two sided error for $f_n$. An exponential gap is presented.

**Keywords:** Computational complexity, binary decision diagrams, approximations.

## 1 Introduction and Definitions

Branching programs (BPs) or Binary Decision Diagrams (BDDs) are a well established representation type or computation model for Boolean functions $f \in B_n$, i.e., $f : \{0,1\}^n \to \{0,1\}$.

**Definition 1.** *A branching program* (BP) *or binary decision diagram* (BDD) *on the variable set* $X_n = \{x_1, \ldots, x_n\}$ *is a directed acyclic graph with one source and two sinks labelled by the constants 0 or 1, resp. Each non-sink node* (*or inner node*) *is labelled by a Boolean variable and has two outgoing edges one labelled by 0 and the other by 1. At each node* $v$ *a Boolean function* $f_v : \{0,1\}^n \to \{0,1\}$ *is represented. For the evaluation of* $f_v(a)$ *start at* $v$*. At* $x_i$*-nodes, the outgoing* $a_i$*-edge is chosen and* $f_v(a)$ *equals the label of the sink which is finally reached on the computation path.*

*The size of a branching program* $G$ *is equal to the number of its nodes and is denoted by* $|G|$*.* BP$(f)$ *denotes the size of the smallest* BP *for a function* $f$*. The*

* Supported in part by DFG We 1066/9.

ISSN 1433-8092

*depth of a branching program is the maximum length of a path from the source to one of the sinks.*

Probabilistic methods have turned out to be useful in almost all areas of computer science. Ablayev and Karpinski (1996) have introduced randomized BPs defined analogously to probabilistic circuits. In the following we consider an alternative approach.

**Definition 2.** *A randomized* BP *(or* BDD*)* $G$ *on the variable set* $X_n = \{x_1, \ldots, x_n\}$ *is a directed acylic graph with decision nodes for Boolean variables and randomized nodes. A randomized node is an unlabelled node with two outgoing edges. The random computation path for* $a$ *is defined as follows. At decision nodes labelled by* $x_i$*, the outgoing* $a_i$*-edge is chosen. At randomized nodes, each outgoing edge is chosen independently from all other random decisions with probability* $1/2$*. The acceptance probability* $acc_G(a)$ *or* $Prob(G(a) = 1)$ *of* $G$ *on* $a$ *is the probability that the random computation path reaches a 1-sink. The rejection probability* $rej_G(a)$ *equals* $Prob(G(a) = 0) = 1 - Prob(G(a) = 1)$*.*

*$G$ represents $f \in B_n$ with two-sided $\epsilon$-bounded error, $0 \leq \epsilon < 1/2$, if $Prob(G(a) \neq f(a)) \leq \epsilon$ for all inputs $a$.*

In order to develop and strengthen lower bound techniques and for applications restricted computation models are considered.

**Definition 3.** *i) A branching program is called read $k$ times* (BP$k$) *if each variable is tested on each path at most $k$ times.*

*ii) A* BP *is called oblivious if the node set can be partitioned into levels such that edges lead from lower to higher levels and all inner nodes of one level are labelled by the same variable.*

In his seminal paper Bryant (1986) has introduced ordered binary decision diagrams (OBDDs) which are up to now the most popular representation not only for formal circuit verification but also in genetic programming.

**Definition 4.** *A permutation $\pi$ on $\{1, \ldots, n\}$ describes the variable ordering $x_{\pi(1)}, \ldots, x_{\pi(n)}$. A* $\pi$-OBDD *for a variable ordering $\pi$ is a* BP *where the sequence of tests on a path is restricted by the variable ordering $\pi$, i.e., if an edge leads from an $x_i$-node to an $x_j$-node, the condition $\pi(i) < \pi(j)$ has to be fulfilled. An* OBDD *is a* $\pi$-OBDD *for some variable ordering $\pi$.*

Unfortunately, several important and also quite simple functions have exponential OBDD size. Therefore, more general representations with good algorithmic behaviour are necessary. Jain, Abadir, Bitner, Fussell, and Abraham (1994) have introduced $k$-IBDDs.

**Definition 5.** *A* $k$-IBDD *is a branching program which can be partitioned into $k$ layers such that the $i$th layer is an* OBDD *(with possible many sources) respecting the ordering $\pi_i$ and such that the edges leaving the $i$th layer reach only nodes of the layers $j > i$ and the sinks.*

One main problem in learning theory is the search for a small size representation of a function $f \in B_n$ which coincides with the partial function $g \in B_n$ defined by the set of labelled training examples $(a_1, g(a_1)), \ldots, (a_m, g(a_m))$ where $a_i \in \{0,1\}^n$ and $m = n^{O(1)}$. The idea to look for a small size representation is based on the well-known Occam's razor principle. See Krause, Savický, and Wegener (1999) for a discussion of this background. They also have described how such a small representation can be obtained by a genetic programming approach. In particular, they have underlined the importance of the choice of a good variable ordering. Simple functions like the direct storage access function $\mathrm{DSA}_n$ and the inner product function $\mathrm{IP}_n$ have small OBDD size but almost all variable orderings lead to exponential size representations for the considered functions and even all reasonably good approximations. In learning theory we cannot expect to learn the unknown function exactly. It is good enough to obtain a representation of a function which almost looks like the unknown function. This motivates the investigation of a BDD theory for the approximate representation of Boolean functions.

**Definition 6.** *i) A function $g \in B_n$ is a $c$-approximation of $f \in B_n$ if $Pr(f(\tilde{x}) = g(\tilde{x})) \geq c$ for a random input $\tilde{x}$.*

*ii) A function $g \in B_n$ is a strong $c$-approximation of $f \in B_n$ if, for $a \in \{0,1\}$, $Pr(g(\tilde{x}) = a) \geq c$ for a random input $\tilde{x} \in f^{-1}(a)$.*

In applications, strong approximations are necessary. However, our lower bounds even hold for $c$-approximations. Since one of the two constant functions 0 and 1 is always a 1/2-approximation we consider $(1/2 + \epsilon_n)$-approximations. Using methods from information theory and communication complexity Krause, Savický, and Wegener (1999) have proved that besides $\mathrm{DSA}_n$ the inner product function $\mathrm{IP}_n$ has only $(1/2 + \epsilon_n)$-approximations of exponential $\pi$-OBDD size for almost all variable orderings $\pi$. Futhermore, they have introduced the permuted inner product function $\mathrm{PIP}_n$ for which all $(1/2+\epsilon_n)$-approximations have exponential OBDD size. The shifted inner product function $\mathrm{SIP}_n$ is related to $\mathrm{IP}_n$ and has appeared in several disguises in the literature on BDDs (see, e.g., Krause (1992)).

**Definition 7.** *i) For even $n$, $\mathrm{IP}_n(x_1, \ldots, x_n) = x_1x_2 \oplus x_3x_4 \oplus \ldots \oplus x_{n-1}x_n$ is the inner product function.*

*ii) For even $n$, let $\mathrm{IP}_n^\pi(x_1, \ldots, x_n) = \mathrm{IP}_n(x_{\pi(1)}, \ldots, x_{\pi(n)})$.*
*For $l = \lceil \log(n!) \rceil$, $\mathrm{PIP}_n(a_0, \ldots, a_{l-1}, x_1, \ldots, x_n) = \mathrm{IP}^{\pi_a}(x_1, \ldots, x_n)$ is the permuted inner product function, where $\pi_a$ is the permutation on $\{1, \ldots, n\}$ coded by the variables $a_0, \ldots, a_{l-1}$.*

*iii) Let $n = 2^l$. The function $\mathrm{SIP}_n$ is defined on $2n+l$ Boolean variables $a_{l-1}, \ldots, a_0, x_{n-1}, \ldots, x_0, y_{n-1}, \ldots, y_0$. Let*

$$\mathrm{SIP}_n^i(x,y) = \bigoplus_{0 \leq j \leq n-1} x_j y_{(j+i) \bmod n},$$

*then*

$$\mathrm{SIP}_n(a,x,y) = \bigvee_{0 \leq i \leq n-1} (|a| = i) \wedge \mathrm{SIP}_n^i(x,y).$$

Now, we give an overview on the rest of the paper. In Section 2, we present some tools from communication complexity and show how distributional communication complexity can be used to prove lower bounds on the BDD size for approximations of Boolean functions. The relation between BDDs for approximations and randomized BDDs with two-sided bounded error is investigated in Section 3. A function $f_n$ is presented which cannot well be represented by oblivious randomized BDDs of linear length but for which there exists a function $g_n$ with linear OBDD size, where $g_n$ approximates $f_n$ on almost all inputs. In Section 4, we prove that the permuted inner product function $\mathrm{PIP}_n$ has no small and good approximation by $k$-IBDDs or BP1s.

## 2 Communication Complexity and Approximations

We start with some notions from communication complexity needed later on. For a thorough introduction into communication complexity, we refer to the monographs of Hromkovič (1997) and Kushilevitz and Nisan (1997). The communication complexity of two-party protocols has been introduced by Yao (1979). It has turned out to be the strongest tool for proving lower bounds on the size of oblivious BDDs. The basic model is the following one. Alice and Bob want to cooperate for the evaluation of a Boolean function $f(x,y) \in B_{m+n}$, i.e., $f : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}$. Alice knows the first part $X = (x_1, \ldots, x_m)$ and Bob the second part $Y = (y_1, \ldots, y_n)$ of the input. Before Alice and Bob obtain their partial inputs, they reach an agreement on a communication protocol which decides who starts the communication and how the communication is organized. If Alice starts the communication, she sends the first message $m_1(x)$ depending on her partial input $x$, Bob answers with a message $m_2(y, m_1(x))$ depending on his partial input $y$ and Alice's message. Then Alice anwers and so on until one player knows $f(x,y)$. The cost of a protocol $P$ for $f$ on input $(x,y)$ is the total length of all messages sent by Alice and Bob. The cost of the protocol $P$ is the cost for a worst case input. The deterministic communication complexity $C(f)$ of $f$ is the minimum cost of $P$, over all protocols $P$ that compute $f$. Many generalizations are possible, e.g., nondeterministic protocols and randomized protocols. Furthermore, the number of communication rounds may be restricted.

Distributional communication complexity provides lower bounds for randomized protocols that are allowed two-sided error.

**Definition 8.** *Let $\mu$ be a probability distribution on $X \times Y$. The $(\mu, \epsilon)$-distributional communication complexity of $f$, $D^{\mu}_{\epsilon}(f)$ , is the cost of the best deterministic protocol that gives the correct answer for $f$ on at least a $(1-\epsilon)$-fraction of all inputs in $X \times Y$, weighted by $\mu$.*

For ease of description we rename the variables such that $\mathrm{IP}_n(x,y) = x_1 y_1 \oplus \cdots \oplus x_{n/2} y_{n/2}$. We give the $x$-variables to Alice and the $y$-variables to Bob. The distributional communication complexity of the inner product function has been studied by Chor and Goldreich (1988). They have proved the following fact which has turned out to be very useful for our investigations.

**Fact 1:** $D^{uniform}_{1/2-\epsilon}(\mathrm{IP}_n) \geq n/4 - \log(1/\epsilon)$.

The following proposition shows how lower bounds on the distributional communcation complexity of a function $f$ lead to lower bounds on the size of BDDs for approximations of $f$.

**Proposition 1.** *Let $f \in B_{m+n}$ and $g$ be a $c$-approximation of $f$ and $G$ be a $k$-*IBDD *representing $g$ with respect to the variable orderings $\pi_1, \ldots, \pi_k$ where for all $\pi_i$, $1 \leq i \leq k$, it holds that the $x$-variables are tested before the $y$-variables. Then it holds that*

$$(2k-1)\lceil \log |G| \rceil \geq D^{uniform}_{1-c}(f).$$

**Proof.** This result follows by the fact that $D^{uniform}_{1-c}(f) \leq C(g)$. The upper bound $(2k-1)\lceil \log |G| \rceil$ for $C(g)$ is well-known (Wegener (2000)). □

Communication complexity cannot be used directly for lower bound proofs on the size of BP$k$s which are not oblivious. The only remaining known lower bound technique is the method of generalized rectangles based on so-called $(k,a)$-rectangles (a name introduced by Jukna (1995)) which first has been described by Borodin, Razborov, and Smolensky (1993) and Okol'nishnikova (1993). Using results from randomized communication complexity Sauerhoff (1998) has generalized this technique to the randomized case to prove lower bounds on the size of randomized BP$k$s. Finally, Thathachar (1998) has applied Sauerhoff's technique to separate the read-$k$-times branching program hierarchy.

**Definition 9.** *A Boolean function $g \in B_n$ is called a $(k,a)$-rectangle if it can be represented as a conjunction of functions $g_i$, $1 \leq i \leq ka$, such that $g_i$ essentially depends only on variables from $X(i)$ where $|X(i)| \leq \lceil n/a \rceil$ and each variable $x_j$ belongs to at most $k$ of the $X(i)$-sets.*

We identify a $(k,a)$-rectangle $R$, which is defined as Boolean function, with the set $R^{-1}(1)$. Such a rectangle $R$ is called $f$-monochromatic if $R \subseteq f^{-1}(0)$ or $R \subseteq f^{-1}(1)$.

**Definition 10.** *A function $f \in B_n$ is called an $(s,k,a)$-step function if $\{0,1\}^n$ can be partitioned into $(2s)^{ka}$ $f$-monochromatic $(k,a)$-rectangles.*

Functions $f \in B_n$ whose BP$k$ size equals $s$ are $(s,k,a)$-step functions. Applying a method due to Yao (1983), Sauerhoff (1998) has proved that functions with small randomized BP$k$ size can be approximated with small error by $(s,k,a)$-step functions.

**Theorem 1.** *Let $\mu$ be a probability measure on $\{0,1\}^n$. If $f \in B_n$ can be represented by a randomized* BP$k$ *of size $s$ which has two-sided $\epsilon$-bounded error, there exists an $(s,k,a)$-step function $g$ such that $g$ $\mu$-approximates $f$ with $\epsilon$-bounded error, i.e., the set of all $a$ where $f(a) \neq g(a)$ has a $\mu$-measure bounded above by $\epsilon$.*

Therefore, the randomized BP$k$ size of a function $f$ is an upper bound for the BP$k$ size of an $(1-\epsilon)$-approximation of $f$ if $\mu$ is the uniform distribution. How far can the best upper bound obtained by this method be away from the BP$k$ size of an $(1-\epsilon)$-approximation of $f$? In the next section we answer this question for oblivious BDDs of linear length.

## 3 Randomized BDDs and Approximations by BDDs

**Theorem 2.** *Let $n = 2^l$. There exists a function $f_n$ on $3n+l$ Boolean variables which needs exponential size for oblivious randomized* BDDs *of linear length with two-sided $\epsilon$-bounded error, $0 \leq \epsilon < 1/2$, but has a strong $(1-1/2^n)$-approximation $g_n$ which can be represented by $\pi$-OBDDs of size $4n+2l+1$ for arbitrary $\pi$.*

**Proof.** The function $f_n : \{0,1\}^{3n+l} \to \{0,1\}$ is defined on the variable set $A_n \cup X_n \cup Y_n \cup Z_n$ where $A_n = \{a_0, \ldots, a_{l-1}\}$, $X_n = \{x_0, \ldots, x_{n-1}\}$, $Y_n = \{y_0, \ldots, y_{n-1}\}$, and $Z_n = \{z_0, \ldots, z_{n-1}\}$. Let $h_n(a,x,y) := \mathrm{SIP}_n(a,x,y)$. Then the function $f_n$ is given as

$$f_n(a,x,y,z) := \begin{cases} a_0 \oplus \cdots \oplus a_{l-1} \oplus x_0 \oplus \cdots \oplus x_{n-1} \oplus y_0 \oplus \cdots \oplus y_{n-1}, & \text{if } z \neq \mathbf{0}, \\ h_n(a,x,y), & \text{if } z = \mathbf{0}. \end{cases}$$

For the lower bound it is sufficient to show that the subfunction $h_n$ of $f_n$ cannot well be represented by oblivious randomized BDDs of linear length with two-sided $\epsilon$-bounded error. It is known that deterministic oblivious BDDs of linear length representing the function $\mathrm{SIP}_n$ have exponential size (Krause (1992)). Let $s = (s_1, \ldots, s_l)$ be a sequence of variables from $X_n$ and $Y_n$ such that no variable appears more than $k$ times. There are two sets $X'_n \subseteq X_n$ and $Y'_n \subseteq Y_n$ such that $|X'_n| \geq |X_n|/2^{2k-1}$, $|Y'_n| \geq |Y_n|/2^{2k-1}$, and the number of layers in $s$ with respect to $X'_n$ and $Y'_n$ is bounded by $2k+1$ (Alon and Maass (1988)). There are at least $(n/2^{2k-1})^2$ pairs $(x_{j_1}, y_{j_2})$ with $x_{j_1} \in X'_n$ and $y_{j_2} \in Y'_n$. By the pigeonhole principle there exists an index $i \in \{0, \ldots, n-1\}$ such that there are at least $n/2^{4k-2}$ pairs $(x_j, y_{(j+i) \bmod n})$. The $x$-variables of these pairs are contained in the set $X''_n$ and the $y$-variables in $Y''_n$ respectively. We fix the $a$-variables in such a way that $|a| = i$ and we set all variables of $X_n \cup Y_n$ which are not in $X''_n \cup Y''_n$ to 0.

Using the fact that the inner product function $\mathrm{IP}_n$ has a communication complexity of $\Omega(n)$ for randomized communication protocols with two-sided $\epsilon$-bounded error (where Alice holds the $x$- and Bob the $y$-variables and the protocols may use an arbitrary number of rounds) (see e.g. Babai, Frankl, and Simon (1986)) we obtain an exponential lower bound of $\Omega(2^{n/2^{4k-2}})$ for the size of oblivious randomized BDDs of linear length with two-sided $\epsilon$-bounded error, $0 \leq \epsilon < 1/2$, representing the function $f_n$.

Now, we prove the upper bound. The function $g_n(a,x,y,z) = a_0 \oplus \cdots a_{l-1} \oplus x_0 \oplus \cdots \oplus x_{n-1} \oplus y_0 \oplus \cdots \oplus y_{n-1}$ is obviously a strong $(1-1/2^n)$-approximation of $f_n$. The $\pi$-OBDD size of $g_n$ is $4n+2l+1$ for arbitrary $\pi$. □

The result shows that for proving lower bounds on the size of oblivious randomized BDDs for a function $f_n$ it is sufficient to find a hard subfunction which is difficult to represent. Therefore, it is also possible to apply reducibility concepts like the so-called rectangular reductions (see, e.g., Kushilevitz and Nisan (1997)). The situation is different for the representations of strong approximations for $f_n$. If the error probability is close to 1/2, we have to prove that all subfunctions of $f_n$ are hard to represent.

## 4 The Permuted Inner Product Function

The permuted inner product function $\mathrm{PIP}_n$ defined on $l := \lceil \log(n!) \rceil$ selection variables and $n$ data variables $x_1, \ldots, x_n$ has been introduced by Krause, Savický, and Wegener (1999) showing that there exists a function which cannot be approximated well by OBDDs. Now, we prove that $\mathrm{PIP}_n$ cannot be approximated well by $k$-IBDDs and BP1s either. The $l$ selection variables describe a permutation $\pi$ on $\{1, \ldots, n\}$, more precisely, for each permutation the number of codewords is one or two. The function $\mathrm{PIP}_n$ realizes $\mathrm{IP}_n$ on the data variables which are permuted according to $\pi$.

For $\mathrm{IP}_n^\pi$ we call $x_{\pi(2i-1)}$ the partner of $x_{\pi(2i)}$, $1 \leq i \leq n/2$, and vice versa. A variable is called singleton with respect to a variable ordering $\pi'$, or a $\pi'$-singleton for short if it is in the first half of the variables according to $\pi'$ but its partner is in the second half. Similar we define singletons according to a variable set $A \subseteq \{x_1, \ldots, x_n\}$, or a $A$-singletons for short if a variable is in $A$ and its partner is not.

First, we determine the number of different functions $\mathrm{IP}_n$. We consider the following experiment. In the beginning $x_1, \ldots, x_n$ are free. Then we choose the free variable with the smallest index and choose another arbitrary free variable as its partner. The chosen variables are no longer free. The process is repeated $n/2$ times. Hence, the number of different functions $\mathrm{IP}_n^\pi$ equals

$$d(n) = \prod_{1 \leq i \leq n/2} (2i-1) = 2^{\Theta(n \log n)}.$$

**Fact 2** (Krause, Savický, and Wegener (1999))**:** With probability at least $1 - e^{-4\delta^2 n}$ a random variable ordering $\pi'$ leads to at least $(1-\delta)n/8$ $\pi'$-singletons for $\mathrm{IP}_n^\pi$.

**Lemma 1.** *Let $k$ be a constant and let $\pi_1, \ldots, \pi_k$ be independent random variable orderings. With probability at least $1 - e^{-\Omega(n)}$ there is a set $I \subseteq \{x_{\pi(2i-1)} | 1 \leq i \leq n/2\}$ of at least $((1-\delta)/8)^k n = \Omega(n)$ elements with the following property. It is possible to partition each variable ordering $\pi_j$, $1 \leq j \leq k$, into a top part and a bottom part such that for each $x_{\pi(2i-1)} \in I$ the top part contains exactly one of the variables $x_{\pi(2i-1)}$ and $x_{\pi(2i)}$.*

**Proof.** Let $I_0 = \{x_{\pi(2i-1)} | 1 \leq i \leq n/2\}$. The set $I_1$ is constructed in the way described by Krause, Savický, and Wegener (1999). The set $I_j$ is constructed

in the same way based on $I_{j-1}$ and the variable ordering $\pi_j$ restricted to the variables of $I_{j-1}$ and their partners. Hence, $|I_j| \geq ((1-\delta)/8)|I_{j-1}| \geq ((1-\delta)/8)^j n$ proving the claim on the size of $I = I_k$. The error probability is bounded above by $e^{-4\delta^2|I_{j-1}|}$ implying that the total error probability is bounded by $ke^{-4\delta^2|I|} = e^{-\Omega(n)}$. □

Now, we are ready to prove the following theorem.

**Theorem 3.** *Let $k$ be a constant. Each function which is a $c$-approximation of* $\mathrm{PIP}_n$ *with $c \geq 1/2 + ke^{-4\delta^2((1-\delta)/8)^k n} + e^{-n^{1-\epsilon'}}$, $0 < \epsilon' < 1$, has a $k$-*IBDD *size which is bounded below by $2^{\Omega(n)}$.*

**Proof.** Let $\pi_1, \ldots, \pi_k$ be arbitrary variable orderings and $G$ be a $k$-IBDD according to $\pi_1, \ldots, \pi_k$ representing a $c$-approximation $g_n$ of $\mathrm{PIP}_n$. By the definition of $\mathrm{PIP}_n$ we have $N := 2^{\lceil \log(n!) \rceil}$ codewords. Therefore, some $\pi$ have a probability of $1/N$ to be chosen and others have a probability of $2/N \leq 2/(n!)$. Hence, the error probability of Lemma 1 is at most doubled and we get the result that for at least $(1 - 2ke^{-4\delta^2((1-\delta)/8)^k n})2^{\Theta(n \log n)}$ subfunctions $\mathrm{IP}_n^\pi$ of $\mathrm{PIP}_n$ obtained by replacing the selection variables with constants there are at least $((1-\delta)/8)^k n$ variables in a subset $I$ such that $\pi_1, \ldots, \pi_k$ can be partitioned into a top part and a bottom part such that for each $x_{\pi(2i-1)} \in I$ the top part contains exactly one of the variables $x_{\pi(2i-1)}$ and $x_{\pi(2i)}$. We consider such a subfunction. If we fix all other variables, we obtain the function $\mathrm{IP}^\pi_{((1-\delta)/8)^k n}$ or its negation as subfunction. Let $\gamma := 2ke^{-4\delta^2((1-\delta)/8)^k n}$.

By averaging and using Proposition 1

$$|G| \geq 2^{1/(2k-1)(D^{uniform}_{1/(1-\gamma)(1-c)}(\mathrm{IP}_{((1-\delta)/8)^k n}))}.$$

Using Fact 1 it follows that

$$|G| \geq 2^{1/(2k-1)(((1-\delta)/8)^k n/2 - \log((1/2-(1/(1-\gamma))(1-c))^{-1})}.$$

It holds that

$$\begin{aligned} -\log((1/2-(1/(1-\gamma))(1-c))^{-1}) &= \log((-1-\gamma+2c)/(2(1-\gamma))) \\ &\geq \log(-1-\gamma+2c) - 1. \end{aligned}$$

Since $c \geq 1/2 + (1/2)\gamma + e^{-n^{1-\epsilon'}}$ it follows that

$$|G| \geq 2^{1/(2k-1)(((1-\delta)/8)^k n/2 - (\log e) n^{1-\epsilon'})} = 2^{\Omega(n)}.$$

□

Now, we consider $c$-approximations of $\mathrm{PIP}_n$ represented by BP1s. First, we consider the subproblem to represent the function $\mathrm{IP}_n$ by restricted BP1s. Often it turns out to be easier to prove lower bounds on the size of BP1s if they have the following special property.

**Definition 11.** *A* BP1 *on the variables $x_1, \ldots, x_n$ is called regular if for each node $v$ the same set of variables is tested on all paths from the source to $v$. Furthermore, it is required that on each path from the source to the sinks all $n$ variables are tested.*

It is not difficult to see that an arbitrary BP1 $G$ can be converted into a regular BP1 $G'$ with size $|G'| \leq (n+1)|G|$ by inserting dummy tests.

Since the function $\mathrm{IP}_n$ can be represented by OBDDs of linear size, it is not possible to apply directly the method of generalized rectangles to prove exponential lower bounds. Our aim is to adapt the known proof technique to prove a lower bound on the size of restricted BP1s for $\mathrm{IP}_n$. Therefore, we combine an adapted version of the rectangle method with results from distributional communication complexity.

The following theorem is well-known (see e.g. Okol'nishnikova (1993)).

**Theorem 4.** *If a function $f \in B_n$ can be represented by a* BP1 *$G$ of size $s$, $\{0,1\}^n$ can be partitioned into $2s$ $f$-monochromatic rectangles.*

The proof idea of Theorem 4 is to decompose $G$ into 2-dimensional rectangles. Therefore, we define a cut through $G$ consisting of all nodes $v$ such that on each path from the source of $G$ to $v$ exactly $n/2$ variables have been tested. By our assumption on $G$ on all paths to $v$ the same set $A(v)$ of variables has been tested before reaching $v$. In this way paths using the same set of variables reaching the same node $v$ on the cut are "bundled" together. A 2-dimensional rectangle hence represents the computation paths running through a sequence $(v_0, v_1, v_2)$ of nodes, where $v_0$ is the source of $G$, $v_2$ is one of the sinks, and $v_1$ is a node on the cut of $G$. From this construction it follows immediately that properties of the paths to the nodes on the cut of $G$ correspond to properties of the 2-dimensional rectangles.

A 2-dimensional rectangle $R$ corresponds to a balanced partition of the variables into two sets $X_1$ and $X_2$. $R$ is a 2-dimensional rectangle with at least $(1-\delta)n/8$ singeltons if there are at least $(1-\delta)n/8$ variables in $X_1$ whose partners belong to $X_2$.

**Proposition 2.** *Let $G$ be a* BP1 *for $\mathrm{IP}_n$ of size $s$. If for all nodes $v$ of $G$ which can be reached after testing exactly $n/2$ variables it holds that there exist at least $(1-\delta)n/8$ $A(v)$-singletons, $\{0,1\}^n$ can be partitioned into $2s$ $\mathrm{IP}_n$-monochromatic 2-dimensional rectangles with at least $(1-\delta)n/8$ singletons.*

The discrepancy technique (see, e.g., Kushilevitz and Nisan (1997)) is known as a method for proving lower bounds for the distributional communication complexity $D^{\mu}_{\epsilon}$ by giving upper bounds on the size of rectangles that are "almost" monochromatic.

**Definition 12.** *Let $f : X \times Y \to \{0,1\}$ be a Boolean function, $R$ be any rectangle, and $\mu$ be a probability distribution on $X \times Y$. Denote*

$$Disc_{\mu}(R,\ f) = \Big|Pr_{\mu}(f(x,y) = 0 \text{ and } (x,y) \in R) - Pr_{\mu}(f(x,y) = 1 \text{ and } (x,y) \in R)\Big|.$$

*The discrepancy of $f$ according to $\mu$ is $Disc_\mu(f) = \max_R Disc_\mu(R, f)$ where the maximum is taken over all rectangles $R$.*

Sauerhoff (1998) has generalized this method by introducing the rectangle balance property. Afterwards, he has shown how this property can be used for proving lower bounds on the size of randomized BP$k$s. Here, we present the method for proving lower bounds on the size of BP1s for $(1-\epsilon)$-approximations of functions $f \in B_n$.

**Definition 13.** *Let $f \in B_n$, $\mu$ be a probability measure on $\{0,1\}^n$, and $\gamma(n)$ be a real-valued function. The function $f$ has the rectangle balance property with respect to $(\mu, \gamma(n))$ if*

$$\mu(R \cap f^{-1}(0)) \geq \mu(R \cap f^{-1}(1)) - \gamma(n)$$

*holds for each 2-dimensional rectangle $R$.*

**Theorem 5 (Sauerhoff(1998)).** *If $f \in B_n$ has the rectangle balance property with respect to $(\mu, \gamma(n))$, the size of each* BP1 *representing a $(1-\epsilon)$-approximation of $f$ is bounded below by*

$$1/2[(\mu(f^{-1}(1)) - \epsilon)/\gamma(n)].$$

Now, we adapt this method to our case. In the rest of the section we set $\mu$ to the uniform distribution.

**Corollary 1.** *Let $G$ be a regular* BP1 *representing a $(1-\epsilon)$-approximation of* $\mathrm{IP}_n$ *with the property that for all nodes $v$ of $G$ which can be reached after testing exactly $n/2$ variables it holds that there exist at least $(1-\delta)n/8$ $A(v)$-singletons according to* $\mathrm{IP}_n$*. Let $\mu$ be the uniform distribution and $\gamma(n)$ be a real-valued function. If for each 2-dimensional rectangle $R$ with $(1-\delta)n/8$ singletons it holds that $\mu(R \cap \mathrm{IP}_n^{-1}(0)) \geq \mu(R \cap \mathrm{IP}_n^{-1}(1)) - \gamma(n)$, the size of $G$ is bounded below by*

$$1/2[(\mu(\mathrm{IP}^{-1}(1)) - \epsilon)/\gamma(n)].$$

**Fact 3:** $\mu(\mathrm{IP}_n^{-1}(1)) = 1/2 - 1/2^{n/2+1}$ where $\mu$ is the uniform distribution on $\{0,1\}^n$.

Now, we prove that $\mu(R \cap \mathrm{IP}_n^{-1}(0)) \geq \mu(R \cap \mathrm{IP}_n^{-1}(1)) - 2^{-(1-\delta)n/16}$ for the uniform distribution $\mu$ and 2-dimensional rectangles $R$ with $(1-\delta)n/8$ singeltons. Since $R$ corresponds to a balanced partition of the input variables into two sets $X_1$ and $X_2$, $R$ can be described as a rectangle $A \times B$ where $A$ is a set of assignments to $X_1$ and $B$ a set of assignments to $X_2$. We consider the set $X_1' \subseteq X_1$ of all singletons in $X_1$ and the set $X_2' \subseteq X_2$ of all singletons in $X_2$. Each partial assignment to the variables outside $X_1' \cup X_2'$ restricts $R$ to some rectangle $A' \times B'$ where $A'$ contains assignments to the variables in $X_1'$ and $B'$ contains assignments to the variables in $X_2'$. Moreover, $\mathrm{IP}_n$ is restricted to the subfunction $\mathrm{IP}_{(1-\delta)n/8}$ or its negation. Finally, the uniform distribution $\mu$ is restricted to the uniform distribution $\mu'$ on $\{0,1\}^{(1-\delta)n/4}$, the input set for the variables in

$X_1' \cup X_2'$. Now, we use the fact that $Disc_{uniform}(\mathrm{IP}_{(1-\delta)n/8}) \leq 2^{-1/2(1-\delta)n/8}$ (Chor and Goldreich (1988)). We easily obtain the restricted balance property for $\mathrm{IP}_n$ and $R$ by averaging over all partial assignments. Altogether, we have proved the following result.

**Corollary 2.** *Let $\mu$ be the uniform distribution. Let $G$ be a regular* BP1 *representing a c-approximation of* $\mathrm{IP}_n$ *with* $c \geq (1/2 + 1/2^{n/2+1} + 2^{-\Omega(n^{1-\epsilon'})})$, $0 < \epsilon' \leq 1$. *If for all nodes $v$ of $G$ which can be reached after testing exactly $n/2$ variables it holds that there exist at least $(1-\delta)n/8$ $A(v)$-singletons according to $IP_n$, then the size of $G$ is bounded below by $2^{\Omega(n)}$.*

Now, we are ready to prove the following theorem.

**Theorem 6.** *There is no c-approximation of* $\mathrm{PIP}_n$ *with* $c \geq (1/2 + 1/2^{n/2+1} + 1/2^{n^{1-\epsilon'}})$, $0 < \epsilon' \leq 1$, *which can be represented by* BP1*s of polynomial size.*

**Proof.** We assume that there is a $c$-approximation $g_n$ of $\mathrm{PIP}_n$ with $c \geq (1/2 + 1/2^{n/2+1} + 1/2^{n^{1-\epsilon'}})$ which can be represented by a BP1 $G$ of polynomial size. First, we transform $G$ to a regular BP1 $G'$ representing $g_n$. By our assumption $G'$ is of polynomial size $s$. We define a cut through $G'$ consisting of all nodes $v$ such that on each path from the source of $G'$ to $v$ exactly $n/2$ data variables have been tested before reaching $v$. By our assumption on $G'$ on all paths to $v$ the same set $A(v)$ of variables has been tested. By the definition of $\mathrm{PIP}_n$ some permutations $\pi$ are described by two codewords. Therefore, we double the error probability of Fact 2. Since the number of nodes in $G'$ is bounded above by $s$ it follows that for at most $2se^{-4\delta^2 n}2^{\Theta(n \log n)}$ subfunctions $\mathrm{IP}_n^\pi$ of $\mathrm{PIP}_n$ there exists a node $v$ on the cut of $G'$ such that there are at most $(1-\delta)n/8$ $A(v)$-singletons according to $\pi$. By the pigeonhole principle there exists a subfunction $\mathrm{IP}_n^{\pi'}$ of $\mathrm{PIP}_n$ for which there are for each node $v$ on the cut of $G'$ at least $(1-\delta)n/8$ $A(v)$-singletons according to $\pi'$. Furthermore, $Pr(\mathrm{IP}_n^{\pi'}(\tilde{x}) = g_n'(\tilde{x})) \geq (1 - 2se^{-4\delta^2 n})c =: c'$ for a random input $\tilde{x}$ where $g_n'$ is the subfunction of $g_n$ choosing a setting to the selection variables which corresponds to a codeword for $\pi'$. Since $s \ll e^{2\delta^2 n}$ if $n$ is large enough, it holds that $2se^{-4\delta^2 n} < e^{-2\delta^2 n}$. Using Corollary 2 it follows

$$
\begin{aligned}
|G'| &\geq 1/2(1/2 - 1/2^{n/2+1} - (1-c'))2^{(1-\delta)n/16} \\
&= 1/2(-1/2 - 1/2^{n/2+1} + c')2^{(1-\delta)n/16} \\
&\geq 1/2(-1/2 - 1/2^{n/2+1} + (1 - e^{-2\delta^2 n})(1/2 + 1/2^{n/2+1} + 1/2^{n^{1-\epsilon'}}))2^{(1-\delta)n/16} \\
&\geq (1/2)2^{(1-\delta)n/16 - n^{1-\epsilon'} - 1} \\
&= 2^{\Omega(n)}.
\end{aligned}
$$

Since $|G| \geq |G'|/(n+1)$, there is a contradiction of our assumption that $G$ is of polynomial size.

□

## Conclusion

Problems from learning theory and genetic programming motivate the investigation of the approximative representation of Boolean functions by BDDs. Lower bounds for the approximative representation are shown to be harder to obtain as similar bounds for randomized representations. Exponential lower bounds were known for OBDDs and are proved here for the more general models $k$-IBDDs and read-once branching programs.

## References


1. Ablayev, F. and Karpinski, M. (1996). On the power of randomized ordered branching programs. Proc. of ICALP, Lecture Notes in Computer Science 1099, 348–356.
2. Alon, N. and Maass, W. (1988). Meanders and their applications in lower bound arguments. Journal of Computer and System Sciences 37, 118–129.
3. Babai, L., Frankl, P., and Simon, J. (1986). Complexity classes in communication complexity theory. Proc. of 27th FOCS, 337–347.
4. Borodin, A., Razborov, A., and Smolensky, R. (1993). On lower bounds for read-$k$-times branching programs. Comput. Complexity 3, 1–18.
5. Bryant, R. E. (1986). Graph-based algorithms for Boolean manipulation. IEEE Trans. on Computers 35, 677–691.
6. Chor, B. and Goldreich, O. (1988). Unbiased bits from sources of weak randomness and probabilistic communication complexity. SIAM J. Comp. 17(2), 230–261.
7. Hromkovič, J. (1997). *Communication Complexity and Parallel Computing.* Springer.
8. Jain, J., Abadir, M., Bitner, J., Fussell, D. S., and Abraham, J. A. (1992). Functional partitioning for verification and related problems. Brown/MIT VLSI Conference, 210–226.
9. Jukna, S. (1995). A note on read-$k$-times branching programs. RAIRO Theoretical Informatics and Applications 29, 75-83.
10. Krause, M. (1992). Separating $\oplus$-L from L, co-NL, and AL=P for oblivious Turing machines of linear access. RAIRO Theoretical Informatics and Applications 26, 507–522.
11. Krause, M., Savický, P., and Wegener, I. (1999). Approximations by OBDDs and the variable ordering problem. Proc. of ICALP, Lecture Notes in Computer Science 1644, 493–502.
12. Kushilevitz, E. and Nisan, N. (1997). *Communication Complexity.* Cambridge University Press.
13. Okol'nishnikova, E. A. (1993). On lower bounds for branching programs. Siberian Advances in Mathematics 3(1), 152–166.
14. Sauerhoff, M. (1998). Lower bounds for randomized read-k-times branching programs. Proc. 15th STACS, Lecture Notes in Computer Science 1373, 105–115.
15. Thathachar, J. (1998). On separating the read-$k$-times branching program hierarchy. Proc. of 30th Ann. ACM Symposium on Theory of Computing (STOC), 653–662.
16. Wegener, I. (2000). *Branching Programs and Binary Decision Diagrams - Theory and Applications.* SIAM Monographs on Discrete Mathematics and Applications. In print.
17. Yao, A. (1979). Some complexity questions related to distributive computing. Proc. of 11th STOC, 209–213.
18. Yao, A. (1983). Lower bounds by probabilistic arguments. Proc. of 24th FOCS, 420–428.




ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'