Electronic Colloquium on Computational Complexity, Report No. 18 (2013)

Exponential Lower Bounds for Refuting Random Formulas Using Ordered Binary Decision Diagrams

Luke Friedman[⋆] and Yixin Xu[⋆⋆]

Rutgers University, Piscataway, NJ, USA
lbfried/yixinxu@cs.rutgers.edu

Abstract. A propositional proof system based on ordered binary decision diagrams (OBDDs) was introduced by Atserias et al. in [3]. Krajíček proved exponential lower bounds for a strong variant of this system using feasible interpolation [14], and Tveretina et al. proved exponential lower bounds for restricted versions of this system for refuting formulas derived from the Pigeonhole Principle [20]. In this paper we prove the first lower bounds for refuting randomly generated unsatisfiable formulas in restricted versions of this OBDD-based proof system. In particular we consider two systems OBDD* and OBDD+; OBDD* is restricted by having a fixed, predetermined variable order for all OBDDs in its refutations, and OBDD+ is restricted by having a fixed order in which the clauses of the input formula must be processed. We show that for some constant $\epsilon > 0$, with high probability an OBDD* refutation of an unsatisfiable random 3-CNF formula must be of size at least $2^{\epsilon n}$, and an OBDD+ refutation of an unsatisfiable random 3-XOR formula must be of size at least $2^{\epsilon n}$.

1 Introduction

Propositional proof complexity is both an approach for attacking the famous P vs. NP problem, and also for obtaining a better theoretical understanding of algorithms for the satisfiability problem. A whole landscape of proof systems of varying strengths has been mapped out and studied – see for instance [17] for general background in this field. From a complexity theory standpoint the situation is similar to that of circuit complexity – for certain restricted systems such as the resolution system exponential lower bounds on the size of refuting many different families of unsatisfiable propositional formulas have been proved. However, for strong systems such as extended Frege, researchers have failed to prove even super-linear lower bounds for any family of unsatisfiable formulas, despite the fact that if such a system had polynomial-size refutations of all unsatisfiable formulas this would imply that NP = CO-NP.

[⋆] Partially supported by NSF grants CCF-0832787 and CCF-1064785

[⋆⋆] Partially supported by NSF grant CCF-0832787

ISSN 1433-8092

In this paper we prove the first lower bounds on the size of refuting randomly-generated unsatisfiable 3-CNF and 3-XOR formulas in proof systems based on ordered binary decision diagrams. Random CNF formulas have been studied extensively, both as a benchmark for measuring in some sense the average case performance of SAT solving algorithms, and also as a tool for proving proof complexity lower bounds. It is well-known that if a random 3-CNF formula on n variables is generated with Δn clauses for large enough constant Δ, then with high probability the formula will be unsatisfiable. The lack of structure in these formulas makes them hard to refute; indeed, it is conceivable that they require exponential size refutations in *any* proof system, and since even generating candidate hard formulas for strong proof systems can be difficult [16], they are a natural choice for lower bound proofs and developing new techniques for them is a worthwhile task.

Along with random 3-CNF formulas, we also consider random 3-XOR formulas, which are formulas whose clauses are satisfied if and only if one or three of its literals are satisfied. Unlike in the 3-CNF case, determining satisfiability of a 3-XOR formula is known to be computable in polynomial time, since such formulas can be equivalently represented as a system of linear equations over $\mathbf{F}_2$, and then an algorithm such as Gaussian elimination can be used to test the solvability of the system. However, random 3-XOR formulas retain a lot of the important properties of random 3-CNF formulas, and because they are easier to reason about they have been useful in proving lower bounds for weak proof systems (e.g. [2]).

Ordered Binary Decision Diagrams (OBDDs) are data structures for representing Boolean functions that were originally introduced in [6] and have found a wide variety of applications in areas of computer science such as VLSI design and model checking. They have also emerged as a basis for SAT solving algorithms that have been demonstrated to be competitive on certain classes of formulas with the state-of-the-art DPLL based solvers that are generally used in practice [15],[13]. Informally they are read-once branching programs where variables must be queried according to a fixed order. Part of what makes OBDDs so useful is that their relatively rigid structure makes it possible to manipulate them efficently: For any given Boolean function f on n variables and variable order π there is a unique (up to isomorphism) minimal OBDD computing f, and operations such as taking the conjunction of two OBDDs and determining whether an OBDD representing a function f_1 majorizes an OBDD representing a function f_2 (i.e for all x, $f_1(x) \geq f_2(x)$) are computable in polynomial time [6].

A refutation system based on OBDDs was introduced in [3]. The basic idea of such a system is simple: Given an unsatisfiable 3-CNF (or 3-XOR) formula $\mathcal{F}$, an OBDD refutation of $\mathcal{F}$ with respect to a variable order π is a sequence $\mathrm{OBDD}_1, \mathrm{OBDD}_2, \ldots \mathrm{OBDD}_t \equiv 0$, where each OBDD_i uses the variable order π and is either the OBDD representation of a clause from $\mathcal{F}$ (an axiom), or is the conjunction of two OBDDs derived earlier (i.e. $\mathrm{OBDD}_i = \mathrm{OBDD}_j \wedge \mathrm{OBDD}_k$ for some $j, k < i$). One can also include a weakening rule, so that OBDD_i may also be an OBDD such that OBDD_i majorizes OBDD_j for some $j < i$. Such a refutation system is sound and complete, and because computing the conjunction of two OBDDs can be done in polynomial time (as well as determining whether one OBDD majorizes another in the case of a weakening), verifying whether a refutation is correct is also polynomial time computable. Thus these OBDD-based systems qualify as propositional proof systems in the formal sense introduced by Cook and Reckhow [8]. The OBDDs representing axioms in this type of refutation are small, as well as the final OBDD OBDD_t. Therefore, if a refutation in one of these OBDD-based systems has a polynomial number of steps, whether it is polynomial size or not depends only on whether one of the intermediate OBDDs computed along the way has super-polynomial size. The only non-deterministic choices the prover must make are which variable order π to use, and in what order to combine OBDDs. (If a weakening rule exists, the prover must also choose when and how to use it). These choices can be crucial however in determining the size of the refutation; for instance, it is a simple exercise to show that for certain functions the OBDD representation has size $O(n)$ according to one variable order yet size $\Omega(2^n)$ according to another order.

By restricting the options the prover has in making these choices, one can define different variants of this OBDD-based system that have varying strengths. One reason for doing so is that no current OBDD-based SAT solver takes full advantage of the power offered by the underlying OBDD proof system in its unrestricted form. This is a common phenomenon in SAT solving – basing solvers on more powerful proof systems does not necessarily make the solvers better. The reason is that as the proof systems become more powerful, trying to deterministically make the non-deterministic choices of the proof system becomes an increasingly difficult task. This is highlighted by the fact that the best general purpose SAT solvers in use today are variants of the DPLL algorithm, which is based on the resolution system, one of the weakest proof systems that has been studied. In the case of OBDD based systems, it is not clear how to best make use of the full weakening rule, and even determining the best variable order to use in an OBDD representation of a single function is an NP-complete problem [5]. Particularly when considering random formulas, because of the symmetry and lack of structure it seems unlikely that one variable order would be exponentially better than another, or even if such a good order did exist that it could be found efficiently. However, the sheer number of different possible variable orders make proving such a fact difficult from a technical standpoint.

From a theoretical point of view, restricting these OBDD systems creates interesting intermediate systems. It was proved in [3] that allowing unrestricted use of the weakening rule makes the OBDD proof system as strong as CP*, a variant of the cutting planes system where coefficients are represented in unary, that is strictly stronger than resolution and for which the only known lower bounds are based on feasible interpolation. However, if we do not allow weakening the story changes significantly – in this case there exist certain families of unsatisfiable formulas for which the smallest OBDD refutations are exponentially larger than the smallest resolution refutations [20]. Despite this apparent weakness, it has not been proved that the Frege system, a powerful system that could even conceivably be optimal, can polynomially simulate this restricted OBDD system. The reason is that the lines of Frege systems are formulas, which cannot directly simulate the dag-like structure of OBDDs. Thus studying different variations of restricted OBDD-based systems is one possible route towards bridging the gap between systems we know to be weak and those for which we do not have lower bounds on natural families of formulas.

Krajíček gave exponential lower bounds for the OBDD-based system of [3] in its full generality using a form of the feasible interpolation method [14], and these are currently the only lower bounds known for this strongest variant. Tveretina et al. showed that if the weakening rule is disallowed, then an OBDD-based refutation of the pigeonhole principle must have exponential size [20], building upon a similar result from Groote and Zantema [11], who had also restricted the system to only consider specific variable orders. In this paper we take a first step towards understanding the limitations of OBDD-based systems to refute *random* formulas by proving exponential lower bounds for certain restricted variants.

In particular we consider two restricted OBDD-based systems, which we can denote by OBDD* and OBDD+. In both systems the weakening rule is excluded. In the OBDD* system the variable order that will be used for the refutation is fixed before the random formula is chosen. Because random formulas are generated symmetrically with respect to the variables, without loss of generality we can fix the identity order $\mathcal{I}$ that orders a set of variables $x_1, x_2, \dots x_n$ as $x_1 < x_2 < \cdots < x_n$. In the OBDD* system the prover has the freedom to combine OBDDs during the refutation in an arbitrary way. In the OBDD+ system, the prover has the freedom to choose any variable order π *after* seeing the random formula ϕ that is to be refuted. However, during the refutation, the clauses of ϕ (represented as OBDDs) must be combined in a predetermined fashion corresponding to some canonical ordering of the clauses in ϕ.

The following two theorems are our main results:

Theorem 1. *Let Δ be a sufficiently large constant. There exists an $\epsilon > 0$, such that with high probability when ϕ is a random 3-CNF formula on n variables with clause density Δn, ϕ is unsatisfiable and any OBDD* refutation of ϕ must have size at least $2^{\epsilon n}$.*[1]

Theorem 2. *Let Δ be a sufficiently large constant. There exists an $\epsilon > 0$ such that with high probability when ϕ is a random 3-XOR formula on n variables with clause density Δn, ϕ is unsatisfiable and any OBDD+ refutation of ϕ must have size at least $2^{\epsilon n}$.*

The progress we have made in this paper is summarized in Figure 1.

[1] This theorem can be proved almost identically in the case where we consider a random 3-XOR formula as well. Also, a close inspection of the proof shows that for either the 3-CNF or 3-XOR case, if instead of fixing the variable order $\mathcal{I}$ we allow the prover to fix any set S of $2^{\delta n}$ variable orders for sufficiently small δ before seeing the random formula ϕ, then to choose one of the variable orders from S after seeing ϕ, the theorem still holds in this scenario as well.

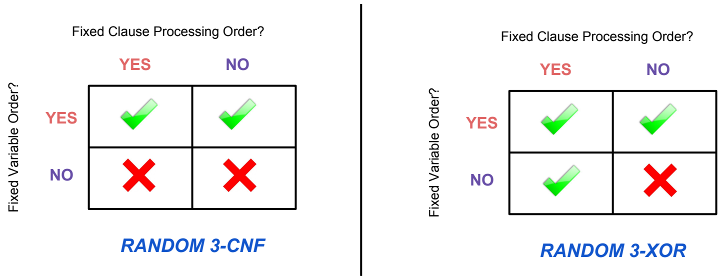

Fig. 1. *A summary of the results from this paper. We consider different OBDD-based proof systems, none of which include a weakening rule. The systems differ according to two possible restrictions: (1) Is the variable order that will be used in the refutation fixed before the random formula is chosen? (2) Are the clauses processed in the refutation according to a canonical order in which they appear in the input formula? We consider both random 3-CNF and random 3-XOR formulas. A check mark appears in the box corresponding to a given proof system and type of random formula if we prove exponential lower bounds for this combination in this paper, and an X appears in the box if proving lower bounds in this case is still open.*

2 Preliminaries and Notations

We will denote a set of n Boolean variables as $\{x_1, \ldots, x_n\}$. A literal x_i^j, $j \in \{0,1\}$, is either a variable or its negation. An assignment α to a set of n variables is a function $[n] \rightarrow \{0,1\}$, where $[n]$ denotes the set $\{1, 2, \ldots, n\}$. α satisfies a literal x_i^j if and only if $\alpha(i) = j$.

A clause C is a set of literals. An assignment α satisfies C as a CNF clause if and only if α satisfies some literal in C. α satisfies C as an XOR clause if and only if α satisfies an odd number of literals in C. A 3-CNF (3-XOR) formula $\mathcal{F}$ over n variables is a list of clauses $(C_1, \ldots, C_m)$, where each of the clauses contains three literals from variables in the set $\{x_1, \ldots x_n\}$. It is satisfied by an assignment α if and only if every clause in $\mathcal{F}$ is satisfied by α as a CNF (XOR) clause. If it is irrelevant whether we are referring to a 3-CNF formula or a 3-XOR formula, we will often refer to the formula simply as a 3-formula.

Definition 1 (Random 3-formula). *A random 3-formula ϕ on n variables with clause density Δ is a 3-formula $(C_1, \ldots, C_{\Delta n})$, where each clause C_i is chosen uniformly at random from all of the $2^3\binom{n}{3}$ possible clauses.*

Let π be a total order on a set of variables $\{x_1, \ldots, x_n\}$. We will refer to π simply as an *order*. Alternatively, we can view π as a permutation such that $\pi(i) = j$ if and only if the i-th variable in the order of π is x_j. We will also write $\pi^{-1}(j) = i$ to indicate that $\pi(i) = j$. We also define the identity order $\mathcal{I}$ such that for all i, $\mathcal{I}(i) = i$.

Let $f : \{0,1\}^n \rightarrow \{0,1\}$ be a Boolean function on n variables and let $\mathbf{z} \in \{0,1\}^t$ for $t \leq n$. We define $f|_{\pi,\mathbf{z}}$ to be the function $f' : \{0,1\}^{n-t} \rightarrow \{0,1\}$ that is the function f restricted so that for each $1 \leq i \leq t$, if $\pi(i) = j$, then x_j is fixed to the constant value $\mathbf{z}_i$.

Definition 2 (OBDD). *Given an order π on $\{x_1, \dots, x_n\}$, an ordered binary decision diagram with respect to π, denoted by OBDD_π, is a branching program with the following structure. An OBDD_π is a layered directed acyclic graph with layers 1 through $n+1$. Layer 1 contains a single root node, and layer $(n+1)$ contains two final nodes, one labeled with the value 0 and the other labeled with the value 1. Every node in layers 1 through n has outdegree two: such a node v on level i has one outgoing edge to a node on level $i+1$ labeled with the value 0, and another outgoing edge to a node on level $i+1$ labeled with the value 1.*

An OBDD_π defines a Boolean function $\{0,1\}^n \to \{0,1\}$ in the following way. For an assignment α on n variables, we start at the root node, and for $i = 1$ to n, advance along the edge labeled with $\alpha(\pi(i))$. When this process is complete, we will have arrived at one of the final nodes. If this final node is labeled with 0, then we define $\mathrm{OBDD}_\pi(\alpha) = 0$, and otherwise we define $\mathrm{OBDD}_\pi(\alpha) = 1$, where now we are associating α with an n bit string in the natural way.

$|\mathrm{OBDD}_\pi|$ denotes the size (the number of nodes) of the OBDD.

An important property of OBDDs is that for a given Boolean function $f : \{0,1\}^n \to \{0,1\}$ and an ordering π, there is a unique minimal OBDD_π up to isomorphism computing f [6]. Thus for a given f we can safely refer to $\mathrm{OBDD}_\pi(f)$ as *the* OBDD computing f according to π.

The following simple theorem (and corollary) provide general techniques for proving lower bounds on $|\mathrm{OBDD}_\pi(f)|$.

Theorem 3 ([18]). *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function on n variables and π an order. Let $k = |\{f|_{\pi,\mathbf{z}} : \mathbf{z} \in \{0,1\}^t\}|$ (i.e., k counts the number of distinct subfunctions of f that can be produced by fixing the first t variables according to π). Then the t-th level of $OBDD_\pi(f)$ contains k nodes.*

Corollary 1 ([19]). *Let f be a Boolean function on n variables and π an order. Suppose the following conditions hold*

1. *$x_1, \cdots, x_t$ are the least t variables according to π for some $t < n$.*
2. *$B \subseteq \{1, \dots, t\}$.*
3. *$\mathbf{z} \in \{0,1\}^t$.*
4. *For all $\mathbf{x}, \mathbf{x}' \in \{0,1\}^t$, if $\mathbf{x} \neq \mathbf{x}'$ and $\mathbf{x}_i = \mathbf{x}'_i = \mathbf{z}_i$ for all $i \notin B$, then there exists $\mathbf{y} \in \{0,1\}^{n-t}$ such that $f(\mathbf{x}, \mathbf{y}) \neq f(\mathbf{x}', \mathbf{y})$.*

Then $|\mathrm{OBDD}_\pi(f)| \geq 2^{|B|}$.

Definition 3 ($\mathbf{OBDD}^*_\pi$ refutation). *Given an unsatisfiable 3-formula $\mathcal{F}$ and an order π, an $OBDD^*_\pi$ refutation of $\mathcal{F}$ is a sequence $OBDD_\pi(f_1), OBDD_\pi(f_2), \cdots, OBDD_\pi(f_t \equiv 0)$ such that for each f_i one of the following conditions is satisfied:*

1. *f_i is a clause of $\mathcal{F}$. (In this case we say that f_i is an* axiom*.)*

2. $f_i = f_j \wedge f_k$ *for some* $j, k < i$.

The size of the $OBDD^*_\pi$ *refutation is defined as* $\sum_{i=1}^{t} |OBDD_\pi(f_i)|$.

We define $S^*_\pi(\mathcal{F})$ *to be the minimum size of any* OBDD^*_π *refutation of* $\mathcal{F}$. *In this paper we focus on* $\pi = \mathcal{I}$ *and thus will refer to* $S^*_{\mathcal{I}}(\mathcal{F})$.

Definition 4 ($\mathbf{OBDD}^+_\pi$ refutation). *An* OBDD^+_π *refutation of an unsatisfiable 3-formula* $\mathcal{F} = (C_1, \ldots, C_m)$ *is an* OBDD^*_π *refutation where the clauses of* $\mathcal{F}$ *are processed one at a time in order. Precisely, an* OBDD^+_π *refutation of* $\mathcal{F}$ *is a sequence* $OBDD_\pi(f_1), OBDD_\pi(f_2), \cdots, OBDD_\pi(f_{2m} = 0)$ *where for* $1 \leq i \leq m$, $f_i = C_i$, $f_{m+1} = C_1$, *and for* $m + 2 \leq j \leq 2m$, $f_j = f_{j-1} \wedge f_{j-m}$. *We define* $S^+_\pi(\mathcal{F})$ *to be the size of the unique* OBDD^+_π *refutation of* $\mathcal{F}$, *and we define* $S^+(\mathcal{F})$ *to be the minimum over* π *of* $S^+_\pi(\mathcal{F})$.

We will make use of the following bounds related to satisfiability thresholds.

Theorem 4. *[9] There exists* $\Delta^* \leq 4.51$ *such that for large* n, *w.h.p a random 3-CNF formula with* n *variables and clause density* $\Delta > \Delta^*$ *will be unsatisfiable.*

Theorem 5. *[10] There exists* $\Delta^* \leq 0.91$ *such that for large* n, *w.h.p a random 3-XOR formula with* n *variables and clause density* $\Delta > \Delta^*$ *will be unsatisfiable.*

We will also need the following lemma, which is a restatement of a result that appeared in [7]. For S a subset of the clauses of a 3-formula $\mathcal{F}$, let $var(S)$ be the set of all variables that appear in at least one of the clauses of S (ignoring the sign of the literal). We call a 3-formula $\mathcal{F}$ on n variables an (x, y)-expander if for all subsets S of the clauses of $\mathcal{F}$ such that $|S| \leq xn$, $|var(S)| \geq y|S|$.

Lemma 1. *[7] For all* $y < 2$ *and* $\Delta > 0$, *there exists positive* x *such that w.h.p a random 3-formula on* n *variables with clause density* Δ *will be an* (x, y)*-expander.*

Finally, we need two results on systems of distinct representatives that follow from Hall's marriage theorem. For a clause C, let $var(C)$ be the set of variables appearing in C, and for a set of clauses S, let $var(S) = \cup_{C \in S} var(C)$. We say a subset S of clauses has a system of distinct representatives (SDR) if there is a one-to-one function $\sigma : S \rightarrow var(S)$ such that for all $C \in S$, $\sigma(C) \in var(C)$.

Lemma 2. *[12] Let* S *be a subset of clauses.* S *has an SDR if and only if for all* $S' \subseteq S$, $|var(S')| \geq |S'|$.

Lemma 3. *[7] Let* S *be a set of clauses and* V *a set of variables.* S *has an SDR* σ *with at most* t *elements of* V *in the range of* σ *if and only if it has an SDR and for all* $S' \subseteq S$, $|S'| - |var(S') \setminus V| \leq t$.

3 Proof of the OBDD* Case

The purpose of this section is to prove Theorem 1, which we now restate.

Theorem 6 (restatement of Theorem 1). *Let $\Delta > 4.51$. There exists a constant $\epsilon > 0$ such that, with high probability when ϕ is a random 3-CNF formula on n variables with clause density Δ, ϕ is unsatisfiable and $S^*_{\mathcal{I}}(\phi) \geq 2^{\epsilon n}$.*

The main work in our proof of Theorem 6 is proving the following lemma.

Lemma 4. *Let $\Delta > 4.51$. There exist constants $\delta, \epsilon > 0$ such that, with high probability when ϕ is a random 3-CNF formula on n variables with clause density Δ, ϕ is a $(\delta, 1.9)$ expander and the following holds: Let S be any subset of the clauses of ϕ such that $\delta n/2 \leq |S| \leq \delta n$, and let f_S be the conjunction of these clauses. Then $|OBDD_{\mathcal{I}}(f_S)| \geq 2^{\epsilon n}$.*

Proof (of Theorem 6). Because $\Delta > 4.51$, by Theorem 4 with high probability ϕ will be unsatisfiable. Let $P = OBDD_{\mathcal{I}}(f_1), OBDD_{\mathcal{I}}(f_2), \cdots, OBDD_{\mathcal{I}}(f_t = 0)$ be an OBDD$^*_{\mathcal{I}}$ refutation of ϕ. Each f_i is a conjunction of some subset of clauses S of ϕ. Let $|f_i|$ denote $|S|$.

By Lemma 1, there exists a constant δ such that with high probability ϕ is a $(\delta, 1.9)$ expander. By Lemma 2 this means that any subset S of clauses of ϕ with $|S| \leq \delta n$ has an SDR. Any set of clauses S that has an SDR σ is satisfiable, since an assignment that for each clause $C \in S$ sets $\sigma(C)$ to the value that satisfies C will satisfy S. Therefore, since f_t is the constant 0 function, which is trivially unsatisfiable, $|f_t| \geq \delta n$. For every f_i that is an axiom, we have $|f_i| = 1$. If $f_i = f_j \wedge f_k$ for some $j, k < i$, then $|f_i| \leq |f_j| + |f_k|$. Therefore, for each $i \in t$, $|f_i| \leq 2 \max_{j<i} |f_j|$. This implies that there exists $i \in [t]$ such that $\delta n/2 \leq |f_i| \leq \delta n$. By Lemma 4, $|OBDD_{\mathcal{I}}(f_i)| \geq 2^{\epsilon n}$, so P has size at least $2^{\epsilon n}$.

The remainder of this section is devoted to proving Lemma 4. First we prove a few other lemmas that will be useful towards this goal. Some of these proofs are deferred to the Appendix for space reasons.

Lemma 5. *Let $\Delta > 0$ and $0 < \delta < \Delta$ be some constant. There exists $\epsilon > 0$, such that with high probability when ϕ is a random 3-formula on n variables with clause density Δ, for any set T of ϵn variables, the number of clauses from ϕ that contain a variable from T is less than δn.*

Lemma 6. *Let $\Delta > 0$ and $0 < \delta < \Delta$ be some constant. There exists $\epsilon > 0$, such that with high probability when ϕ is a random 3-formula on n variables with clause density Δ, the following property holds: For all sets S of clauses from ϕ with $|S| \geq \delta n$, there exists a set of clauses $T \subseteq S$ with $|T| = \epsilon n$ such that the clauses in T are disjoint (i.e. no two clauses of T share a common variable).*

Definition 5 (splits). *Let t be a positive integer less than n and $\mathcal{F}$ a 3-formula. For a clause $C \in \mathcal{F}$, we say that t left-splits C according to an order π if there is exactly one variable $x_i \in var(C)$ such that $\pi^{-1}(i) \leq t$. In this case we define $\text{left}_{C,t,\pi} = x_i$. Similarly, we say that t right-splits C according to an order π if there is exactly one variable $x_i \in var(C)$ such that $\pi^{-1}(i) > t$, and in this case define $\text{right}_{C,t,\pi} = x_i$. If t either right-splits or left-splits C, then we will sometimes simply say that t splits C.*

Lemma 7. *Let $\Delta, \delta > 0$ be any constants, and for some $0 < \epsilon < 1$, let $\Gamma_\epsilon = \{\lceil \epsilon n \rceil, \lceil 2\epsilon n \rceil, \lceil 3\epsilon n \rceil, \cdots, \lceil (1-\epsilon)n \rceil\}$. Then with high probability when ϕ is a random 3-formula on n variables with clause density Δ, for any set of clauses S from ϕ, with $|S| \geq \delta n$, there exists $t \in \Gamma_\epsilon$ such that at least $(\delta - 7\epsilon\Delta)\epsilon n$ of the clauses are left-split by t according to $\mathcal{I}$.*

Lemma 8. *Let $\Delta > 4.51$, and let δ' be the constant that comes out of Lemma 1 such that with high probability a random 3-formula with clause density Δ is a $(\delta', 1.9)$ expander. Let $\delta < \delta'$ be some constant. There exists constants $\gamma, \epsilon > 0$, such that with high probability when ϕ is a random 3-formula on n variables with clause density Δ, the following property holds: For all sets T of clauses from ϕ, with $\delta n \leq |T| \leq \delta' n$, there exists $S \subseteq T$ such that*

1. *$|S| = \gamma n$.*
2. *There exists $t \in \Gamma_\epsilon$ such that every clause $C \in S$ is left-split by t according to $\mathcal{I}$.*
3. *The clauses of S are disjoint.*
4. *T has an SDR σ, such that for every clause $C \in S$, exactly one variable in C is in the range of σ.*

Proof. Suppose the conclusions of Lemma 1, Lemma 6, and Lemma 7 hold with respect to ϕ (which occurs with high probability). Let T be a set of clauses from ϕ such that $\delta n \leq |T| \leq \delta' n$. By Lemma 7, there exists a constant ϵ such that for $\lambda = (\delta - 7\epsilon\Delta)\epsilon > 0$, at least λn clauses from T are left-split by $t \in \Gamma_\epsilon$ according to $\mathcal{I}$. Call this set of λn clauses U.

By Lemma 6, for some constant λ' we can find a set of $\lambda' n$ disjoint clauses $U' \subseteq U$. Now we invoke Lemma 3 to show that there exists an SDR σ for T such that at most $1.6\lambda' n$ of the $3\lambda' n$ variables in $var(U')$ are in the range of σ. To do this it suffices to show that for any set of clauses $S' \subseteq T$, $|S'| - |var(S') \setminus var(U')| \leq 1.6\lambda' n$. If $|S'| \leq 1.6\lambda' n$ then trivially the inequality is satisfied. Otherwise, if $|S'| > 1.6\lambda' n$, then because ψ is a $(\delta', 1.9)$ expander, $|var(S')| \geq 1.9|S'|$, so

$$|S'| - |var(S') \backslash var(U')| \leq -0.9|S'| + 3\lambda' n \leq -1.44\lambda' n + 3\lambda' n \leq 1.6\lambda' n$$

Because there are at most $1.6\lambda' n$ variables from $var(U')$ in the range of σ, there must exist a set of clauses $S \subseteq U'$, with $|S| = 0.4\lambda' n$, such that for every clause $C \in S$, exactly one variable in C is in the range of σ. This set S satisfies the requirements of the lemma.

We are now ready to prove Lemma 4.

Proof (of Lemma 4).
By Lemma 1, there exists $\delta > 0$ such that with high probability ϕ is a $(\delta, 1.9)$ expander, and also with high probability the conclusion of Lemma 8 holds.

Let S be a subset of the clauses of ϕ such that $\delta n/2 \leq |S| \leq \delta n$. Let $S' \subseteq S$ be the set guaranteed to exist by Lemma 8 with the four properties from that lemma, and σ the corresponding SDR for S. Let $\mathrm{left}_{S'} = \{\mathrm{left}_{C,t,\mathcal{I}} : C \in S'\}$.

In order to prove the lemma we will make use of Corollary 1 to show that $|OBDD_{\mathcal{I}}(f_S)| \geq 2^{|\mathrm{left}_{S'}|} \geq 2^{\epsilon n}$ for some constant $\epsilon > 0$. Our set B from that theorem will be $\mathrm{left}_{S'}$. Define $\mathbf{z} \in \{0,1\}^t$ as follows. For each $1 \leq i \leq t$ such that $x_i = \sigma(C)$ for some clause C, let $\mathbf{z}_i$ be the value that satisfies the clause C. Assign all other values of $\mathbf{z}$ arbitrarily.

To finish the proof of the lemma, we need that for all $\mathbf{x}, \mathbf{x}' \in \{0,1\}^t$, if $\mathbf{x} \neq \mathbf{x}'$ and $\mathbf{x}_i = \mathbf{x}'_i = \mathbf{z}_i$ for all $i \notin B$, then there exists $\mathbf{y} \in \{0,1\}^{n-t}$ such that $\phi(\mathbf{x}, \mathbf{y}) \neq \phi(\mathbf{x}', \mathbf{y})$.

Let $\mathbf{x}, \mathbf{x}' \in \{0,1\}^t$ such that $\mathbf{x} \neq \mathbf{x}'$ and $\mathbf{x}_i = \mathbf{x}'_i = \mathbf{z}_i$ for all $i \notin B$. Let j be an index such that $\mathbf{x}_j \neq \mathbf{x}'_j$. Let C be the clause from S' such that $\mathbf{x}_j = \mathrm{left}_{C,t,\mathcal{I}}$.

Define $\mathbf{y}$ as follows. Let p and q be the two indices other than j such that $\mathbf{x}_p$ and $\mathbf{x}_q$ are in the clause C. Define $\mathbf{y}_{p-t}$ and $\mathbf{y}_{q-t}$ each to be the value that does not satisfy the clause C. For each clause $D \neq C$ such that $D \in S'$, let r and s be the two indices greater than t such that $\mathbf{x}_r$ and $\mathbf{x}_s$ are in the clause D. Define $\mathbf{y}_{r-t}$ and $\mathbf{y}_{s-t}$ each to be the value that satisfies D. For any index i such that $t < i \leq n$ and $\mathbf{x}_i = \sigma(E)$ for some clause E other than C, define $\mathbf{y}_{i-t}$ to be the value that satisfies the clause E. Assign all other values of $\mathbf{y}$ arbitrarily. Note that because the clauses in S' are disjoint and for each clause C in S' exactly one of the variables of C is in the range of σ, it is always possible to form the partial assignment $\mathbf{y}$ according to these rules. (Note in particular that if $\sigma(E) = x$ for some clause $E \notin S'$, then $x \notin var(S')$).

Either $\mathbf{x}_j$ satisfies the clause C, or $\mathbf{x}'_j$ does. Assume without loss of generality that $\mathbf{x}_j$ does. Then $\phi(\mathbf{x}', \mathbf{y}) = 0$, since the assignment $(\mathbf{x}', \mathbf{y})$ does not satisfy the clause C. However, $\phi(\mathbf{x}, \mathbf{y}) = 1$, since the assignment $(\mathbf{x}, \mathbf{y})$ satisfies every clause in ϕ. This completes the proof.

4 Proof of the OBDD+ Case

The purpose of this section is to prove Theorem 2, which we now restate.

Theorem 7 (Restatement of Theorem 2). *Let $\Delta > 0.91$. There exists a constant $\epsilon > 0$ such that, with high probability when ϕ is a random 3-XOR formula on n variables with clause density Δ, ϕ is unsatisfiable and $S^+(\phi) \geq 2^{\epsilon n}$.*

The following three lemmas are needed in the proof; for space reasons we are forced to stick their proofs and the proof of Theorem 7 in the appendix.

Lemma 9. *For $0 < \epsilon < 1$, let $\Gamma_\epsilon = \{\lceil \epsilon n \rceil, \lceil 2\epsilon n \rceil, \lceil 3\epsilon n \rceil, \ldots, \lceil (1-\epsilon)n \rceil\}$. Let $\Delta > 0.5$. There exists $\epsilon, \delta > 0$ such that, with high probability when ϕ is a random 3-formula on n variables with clause density Δ, the following property holds: For* any *order π, there exists some $t_\pi \in \Gamma_\epsilon$ such that more than δn of the clauses from ϕ are split by t_π according to π.* [2]

[2] In fact, using a slightly more complicated first moment argument, one can prove the stronger statement that this lemma holds even if we fix $t_\pi = n/2$.

Lemma 10. *There exists $\lambda > 0$ such that, with high probability when ϕ is a random 3-XOR formula with clause density $\Delta = 0.6$, ϕ is a $(0.6, 1+\lambda)$ expander.*

Lemma 11. *Let ψ be a 3-formula over n variables with clause density Δ such that ψ is a $(\Delta, 1 + \delta)$-expander for some $\delta > 0$. Let $U \subseteq \psi$ be a set of disjoint clauses with $|U| = \lambda n$ for some $\lambda > 0$. Let Ψ be a set of variables and f be a bijection from Ψ to U such that for all $x \in \Psi$, x appears in the clause $f(x)$. Then there exists $\epsilon > 0$ such that there is an SDR σ on ψ for which at least ϵn of the variables from Ψ are not in the range of σ.*

5 Future Work

The obvious open problem is to prove lower bounds for refuting random 3-CNF or 3-XOR formulas in an OBDD-based refutation system where neither the variable order nor the order in which clauses are processed in the refutation is constrained. Although it might seem that one could tweak our techniques to get this result, it may be that this is more difficult than appears at first glance.

For instance, suppose we tried to use the same approach of focusing in on a particular OBDD in the refutation of a random 3-CNF formula such that the OBDD represents the conjunction of about δn clauses, for some appropriately chosen fixed δ, in the hopes of showing that the OBDD must be of exponential size. In the restricted systems from this paper, we were able to choose δ to be an arbitarily small constant. However, in the unrestricted system (still without weakening), we would be forced to choose δ to be greater than about $1/6$, because a random formula with clause density just above the threshold *does* contain sub-formulas with about $n/6$ clauses that have small OBDD representations for some variable order. (For instance, one can look for a large set of disjoint clauses and then choose a variable order where the variables from each clause are adjacent in the order). It is much more difficult to reason about sub-formulas in this regime; for instance, to the best of our knowledge it has not even been proved that a random 3-CNF formula with clause density just above the threshold with high probability does not contain an unsatisfiable sub-formula consisting of $n/6$ clauses (let alone that all sub-formulas slightly larger than this must have large OBDD representations). Certainly one cannot rely on the existence of SDRs when considering sub-formulas with clause density this close to the threshold.

Making progress will probably require using a more sophisticated analysis of the structure of random formulas than we do in this paper. A large amount of research has been done on investigating the structure of random CNF and XOR formulas with densities below the respective satisfiability thresholds, including understanding the solution space structure of such formulas and the occurrence of various phase transitions. (See for example [1] for a survey of this work, along with more general information about SAT solving and random formulas). Finding a way to leverage this type of knowledge in this context is probably a key step towards achieving these more difficult lower bounds.

References

1. D. Achlioptas. Random satisfiability. *Handbook of Satisfiability*, pages 245–270, 2009.
2. M. Alekhnovich. Lower bounds for k-DNF resolution on random 3-CNFs. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 251–256, 2005.
3. A. Atserias, P. Kolaitis, and M. Vardi. Constraint propagation as a proof system. In *Tenth International Conference on Principles and Practice of Constraint Programming*, pages 77–91, 2004.
4. E. Ben-Sasson and N. Galesi. Space complexity of random formulae in resolution. In *Proceedings of the Sixteenth Annual IEEE Conference on Computational Complexity*, pages 42–51, 2001.
5. B. Bollig and I. Wegener. Improving the variable ordering of OBDDs is NP-complete. *IEEE Transactions on Computers*, 45(9):993–1002, 1996.
6. R. E. Bryant. Graph-based algorithms for boolean function manipulation. *IEEE Transactions on Computing*, 35:677–691, 1986.
7. V. Chvátal and E. Szémeredi. Many hard examples for resolution. *Journal of the ACM*, 35(4):759–768, 1988.
8. S.A. Cook and R. Reckhow. The relative efficiency of propositional proof systems. *Journal of Symbolic Logic*, 44:36–50, 1979.
9. O. Dubois, Y. Boufkhad, and J. Mandler. Typical random 3-sat formulae and the satisfiability threshold. Technical Report (10)003, ECCC, 2003.
10. O. Dubois and J. Mandler. The 3-XORSAT threshold. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 769–778, 2002.
11. J. F. Groote and H. Zantema. Resolution and binary decision diagrams cannot simulate each other polynomially. *Discrete Applied Mathematics*, 130:157–171, 2003.
12. P. Hall. On representatives of subsets. *J. London Math. Soc.*, 10:26–30, 1935.
13. J. Huang and A. Darwiche. Toward good elimination ordering for symbolic SAT solving. In *Proceedings of the Sixteenth IEEE Conference on Tools with Artificial Intelligence*, pages 566–573, 2004.
14. J. Krajíček. An exponential lower bound for a constraint propagation proof system based on ordered binary decision diagrams. Technical Report (07)007, Electronic Colloquium on Computational Complexity, 2007.
15. G. Pan and M. Vardi. Search vs. symbolic techniques in satisfiability solving. In *The International Conference on Theory and Applications of Satisfiability Testing*, 2004.
16. A. A. Razborov. Pseudorandom generators hard for k-DNF resolution and polynomial calculus resolution. *manuscript*, 2003.
17. N. Segerlind. The complexity of propositional proofs. *Bulletin of Symbolic Logic*, 54:40–44, 2007.
18. D. Seiling and I. Wegener. NC-algorithms for operations on binary decision diagrams. *Parallel Processing Letters*, 3(1):3–12, 1993.
19. O. Tveretina, C. Sinz, and H. Zantema. An exponential lower bound on OBDD refutations for pigeonhole formulas. In *Athens Colloquium on Algorithms and Complexity.* Electronic Proceedings in Theoretical Computer Science, 2009.
20. O. Tveretina, C. Sinz, and H. Zantema. Ordered binary decision diagrams, pigeonhole formulas and beyond. *Journal on Satisfiability, Boolean Modeling and Computation*, 7:35–38, 2010.

6 Appendix

Acknowledgements: The first author would like to thank Eric Allender, Russell Impagliazzo, Mike Molloy, Mike Saks, and Iddo Tzameret for useful discussions.

Proof (of Lemma 5). Let T be a set of variables such that $|T| = \epsilon n$. For any constant $c > 1$, the probability that a particular clause of ϕ contains a variable from T is less than $3c\epsilon$. Therefore, the expected number of clauses from ϕ that do not contain a variable from T is at least $(1 - 3c\epsilon)\Delta n$. Let $b = (1 - 3c\epsilon)\Delta$. Let X be the random variable that counts the number of clauses from ϕ that do not contain a variable from T. In order for the number of clauses from ϕ that contain a variable from T to be less than δn, we need that $bn - X < \delta n$, so $X > (b - \delta)n = (1 - \frac{\delta}{b})bn$.

By a Chernoff bound, the probability that $X \leq (1 - \frac{\delta}{b})bn$ is less than $e^{-bn\left(\frac{\delta}{b}\right)^2/2} = e^{-\left(\frac{\delta^2}{2b}\right)n}$. The total number of sets of variables that contain exactly ϵn variables is $\binom{n}{\epsilon n} \leq \left(\frac{en}{\epsilon n}\right)^{\epsilon n} = \left(\frac{e}{\epsilon}\right)^{\epsilon n}$. By a union bound, the probability that there exists some set of variables T with $|T| = \epsilon n$ such that more than δn clauses from ϕ contain a variable from T is less than

$$e^{-\left(\frac{\delta^2}{2b}\right)n} \cdot \left(\frac{e}{\epsilon}\right)^{\epsilon n}.$$

By choosing a small enough value of ϵ, this probability is exponentially small in n.

Proof (of Lemma 6). Let ϵ' be the constant that comes out of Lemma 5 corresponding to δ. Take a maximal set $T \subseteq S$ such that the clauses in T are disjoint. Let $\epsilon = \frac{\epsilon'}{3}$. We have that $|T| \geq \epsilon n$, otherwise $var(T)$ is a set of $\epsilon' n$ variables that appear in at least δn clauses, which would violate Lemma 5.

Proof (of Lemma 7). Say two variables x_i and x_j have distance d if $|i - j| = d$. Let C be a clause from ϕ. If all variables from C have pairwise distance at least ϵn, then some element of Γ_ϵ must left-split C according to $\mathcal{I}$. The probability that any two given variables from C have distance at least ϵn is at least $1 - 2\epsilon$, so the probability that all variables from C have pairwise distance at least ϵn is at least $1 - 6\epsilon$.

Therefore the expected number of clauses that are left-split according to $\mathcal{I}$ by at least one element of Γ_ϵ is at least $\Delta n(1 - 6\epsilon)$. Therefore, by a Chernoff bound, with high probability, for any constant $c > 1 - (1 - 6\epsilon) = 6\epsilon$, at most $c\Delta n$ clauses from ϕ are not left-split according to $\mathcal{I}$ by any element of Γ_ϵ.

Assume the above property of ϕ holds and consider an arbitrary set of δn clauses from ϕ. By an averaging argument, there exists $t \in \Gamma_\epsilon$ such that $(\delta - c\Delta)\epsilon n$ of these clauses are left-split according to $\mathcal{I}$ by t. Choosing $c = 7\epsilon$ completes the proof of the lemma.

Proof (of Lemma 9). By Lemma 1, for all $y < 2$, there exists an $x > 0$ such that with high probability ϕ will be an (x, y) expander. We will pick some $y < 2, \epsilon > 0$, and $\delta > 0$ to be determined later in the proof.

Suppose for contradiction that there exists a π such that no $t \in \Gamma_\epsilon$ splits more than δn of the clauses from ϕ. Then there exists $S \subseteq \phi$, with $|S| \geq \Delta n - \frac{\delta n}{\epsilon}$, such that for each clause $C \in S$, all the variables from C are contained within an interval of Γ_ϵ: i.e. there exists $t \in \{0\} \cup \Gamma_\epsilon$ such that for all $x_i \in C$, $\pi^{-1}[i] > t$ and $\pi^{-1}[i] < t + \epsilon n$. In this case we will say that the clause C is contained in the interval t according to π.

By an averaging argument there exists some $t \in \{0\} \cup \Gamma_\epsilon$ and $T \subseteq S$ with $|T| \geq \epsilon(\Delta n - \frac{\delta n}{\epsilon}) = \epsilon \Delta n - \delta n$ such that every clause from T is contained in the interval t.

Suppose that we choose y, ϵ, and δ in such a way that

1. $y < 2$
2. $y(\epsilon \Delta n - \delta n) > \epsilon n$.
3. $xn \geq \epsilon \Delta n - \delta n$

Then this would be a contradiction, because it would imply by the expansion properties of ϕ that not all of the clauses from T can in fact be contained in the interval t.

All that remains is to show that we can satisfy these inequalities simultaneously.

In order to simultaneously satisfy the first two inequalities, we must be able to satisfy the following inequality:

$$2(\epsilon \Delta n - \delta n) \geq \epsilon n$$

Solving for δ, we get that

$$\begin{aligned} 2\epsilon \Delta n - 2\delta n &\geq \epsilon n \\ 2\delta n &\leq 2\epsilon \Delta n - \epsilon n \\ \delta &\leq \epsilon \Delta - \frac{\epsilon}{2} \\ \delta &\leq \epsilon(\Delta - \frac{1}{2}) \end{aligned}$$

Note that because $\Delta > \frac{1}{2}$, δ can satisfy this inequality and still be positive.

A stronger condition than the third inequality is that $\epsilon \leq \frac{x}{\Delta}$. Thus by choosing y to be sufficiently close to 2, then choosing ϵ and δ as shown above we get our contradiction and the lemma is proved.

Proof (of Lemma 10).

This type of calculation is by now standard (see for instance [4], Lemma 5.1), although in our case it is slightly messier than usual because of the more specific bounds that we need.

Let $\delta = 0.01$ and $\varDelta = 0.6$. Let A_i denote the event that a set of clauses S of size i has expansion

$$\frac{|var(S)|}{|S|} < 1 + \delta,$$

where $i = 1, 2, \cdots, \varDelta n$. There are $\binom{\varDelta n}{i}$ such sets of clauses and $\binom{n}{(1+\delta)i}$ possible small vertex sets. The probability for a single edge to fall within the small vertex set is $\binom{(1+\delta)i}{3}/\binom{n}{3} \leq \left(\frac{(1+\delta)i}{n}\right)^3$. Thus

$$\Pr[A_i] \leq \binom{\varDelta n}{i}\binom{n}{(1+\delta)i}\left(\frac{(1+\delta)i}{n}\right)^{3i}.$$

We need to bound the probability $\Pr[A_i]$ in order to show that

$$\Pr[\bigvee_{i=1}^{\varDelta n} A_i] = o(1).$$

Since $\Pr[\bigvee_{i=1}^{\varDelta n} A_i] \leq \Pr[\bigvee_{i=1}^{\sqrt[3]{n}} A_i] + \Pr[\bigvee_{i=\sqrt[3]{n}+1}^{cn} A_i] + \Pr[\bigvee_{i=0.1n+1}^{\varDelta n} A_i]$, where c is a constant between 0 and $\varDelta$ to be determined later, it is enough to show that

$$\Pr[\bigvee_{i=1}^{\sqrt[3]{n}} A_i] = n^{-\varOmega(1)} = o(1) \tag{1}$$

$$\Pr[\bigvee_{i=\sqrt[3]{n}+1}^{cn} A_i] = n^{-\varOmega(\sqrt[3]{n})} = o(1) \tag{2}$$

$$\Pr[\bigvee_{i=cn+1}^{\varDelta n} A_i] = 2^{-\varOmega(n)} = o(1) \tag{3}$$

Using the estimation $\binom{a}{b} \leq (\frac{ea}{b})^b$, we get

$$\begin{aligned}\Pr[A_i] &\leq \left(\frac{e\varDelta n}{i}\right)^i \left(\frac{en}{(1+\delta)i}\right)^{(1+\delta)i} \left(\frac{(1+\delta)i}{n}\right)^{3i} \\ &= \left[\frac{e^{2+\delta} \cdot \varDelta \cdot (1+\delta)^{2-\delta} \cdot i^{1-\delta}}{n^{1-\delta}}\right]^i.\end{aligned}$$

Thus

$$\begin{aligned}\Pr[\bigvee_{i=1}^{\sqrt[3]{n}} A_i] &\leq \sqrt[3]{n} \cdot \frac{e^{2+\delta} \cdot \varDelta \cdot (1+\delta)^{2-\delta} \cdot \sqrt[3]{n}^{1-\delta}}{n^{1-\delta}} \\ &= e^{2+\delta} \cdot \varDelta \cdot (1+\delta)^{2-\delta} \cdot n^{\frac{2\delta-1}{3}}\end{aligned}$$

which tends to 0 as n tends to infinity. This implies (1).

In order to show (2) and (3), we need to use a better approximation for $\Pr[A_i]$.

$$\begin{aligned}\Pr[A_i] &\le \binom{\mathit{\Delta} n}{i}\binom{n}{(1+\delta)i}\left(\frac{(1+\delta)i}{n}\right)^{3i}\\ &= \frac{(\mathit{\Delta} n)!}{(\mathit{\Delta} n-i)!\cdot i!}\cdot\frac{n!}{(n-(1+\delta)i)!\cdot((1+\delta)i)!}\cdot\left(\frac{(1+\delta)i}{n}\right)^{3i}\end{aligned}$$

Using Stirling's approximation,

$$\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}\le n!\le en^{n+\frac{1}{2}}e^{-n}$$

we get

$$\begin{aligned}\Pr[A_i] &\le \frac{e(\mathit{\Delta} n)^{\mathit{\Delta} n+\frac{1}{2}}e^{-\mathit{\Delta} n}}{\sqrt{2\pi}(\mathit{\Delta} n-i)^{\mathit{\Delta} n-i+\frac{1}{2}}e^{-(\mathit{\Delta} n-i)}\sqrt{2\pi}i^{i+\frac{1}{2}}e^{-i}}\\ &\quad\cdot\frac{en^{n+\frac{1}{2}}e^{-n}}{\sqrt{2\pi}(n-(1+\delta)i)^{n-(1+\delta)i+\frac{1}{2}}e^{-(n-(1+\delta)i)}\sqrt{2\pi}((1+\delta)i)^{(1+\delta)i+\frac{1}{2}}e^{-((1+\delta)i)}}\cdot\left(\frac{(1+\delta)i}{n}\right)^{3i}\\ &\le \frac{(\mathit{\Delta} n)^{\mathit{\Delta} n+\frac{1}{2}}}{(\mathit{\Delta} n-i)^{\mathit{\Delta} n-i+\frac{1}{2}}\cdot i^{i+\frac{1}{2}}}\cdot\frac{n^{n+\frac{1}{2}}}{(n-(1+\delta)i)^{n-(1+\delta)i+\frac{1}{2}}\cdot((1+\delta)i)^{(1+\delta)i+\frac{1}{2}}}\cdot\left(\frac{(1+\delta)i}{n}\right)^{3i}=f(i).\end{aligned}$$

Then

$$\begin{aligned}\ln f(i) &= (\mathit{\Delta} n+\frac{1}{2})\ln(\mathit{\Delta} n)-(\mathit{\Delta} n-i+\frac{1}{2})\ln(\mathit{\Delta} n-i)-(i+\frac{1}{2})\ln i\\ &\quad+(n+\frac{1}{2})\ln n-(n-(1+\delta)i+\frac{1}{2})\ln(n-(1+\delta)i)-((1+\delta)i+\frac{1}{2})\ln((1+\delta)i)\\ &\quad+3i(\ln((1+\delta)i)-\ln n)=g(i),\\ \frac{\mathtt{d}g(i)}{\mathtt{d}i} &= \frac{\mathit{\Delta} n-i+1/2}{\mathit{\Delta} n-i}+\ln(\mathit{\Delta} n-i)-\frac{i+1/2}{i}-\ln i+\frac{(1+\delta)(n-(1+\delta)i+1/2)}{n-(1+\delta)i}\\ &\quad+(1+\delta)\ln(n-(1+\delta)i)-\frac{(1+\delta)i+1/2}{i}-(1+\delta)\ln((1+\delta)i)+3(\ln((1+\delta)i)-\ln n)+3\\ &= \ln(\mathit{\Delta} n-i)+(1-\delta)\ln i+(1+\delta)\ln(n-(1+\delta)i)-3\ln n\\ &\quad+\frac{1}{2(\mathit{\Delta} n-i)}-\frac{1}{i}+\frac{1+\delta}{2(n-(1+\delta)i)}+(2-\delta)\ln(1+\delta)+3\\ &= \ln\frac{(\mathit{\Delta} n-i)i^{1-\delta}(n-(1+\delta)i)^{1+\delta}}{n^3}+\frac{1}{2(\mathit{\Delta} n-i)}-\frac{1}{i}+\frac{1+\delta}{2(n-(1+\delta)i)}+(2-\delta)\ln(1+\delta)+3.\end{aligned}$$

When $i\in[\sqrt[3]{n},cn]$, for large enough n, we have

$$\begin{aligned}\frac{\mathtt{d}g(i)}{\mathtt{d}i} &\le \ln\frac{(\mathit{\Delta} n-i)i^{1-\delta}(n-(1+\delta)i)^{1+\delta}}{n^3}+4\\ &\le \ln\frac{(\mathit{\Delta} n)(cn)^{1-\delta}n^{1+\delta}}{n^3}+4\\ &= \ln(\mathit{\Delta} c^{1-\delta})+4.\end{aligned}$$

In order to make $\ln(\varDelta c^{1-\delta}) + 4 < 0$, it is enough to have $c = 0.01$. This means $g(i)$ is decreasing in interval $[\sqrt[3]{n}, cn]$, which implies $f(i)$ is also decreasing in the same interval. Thus

$$\begin{aligned}
\Pr[\bigvee_{i=\sqrt[3]{n}+1}^{cn} A_i] &\le (cn - \sqrt[3]{n}) f(\sqrt[3]{n}) \\
&\le cn \cdot \frac{(\varDelta n)^{\varDelta n + \frac{1}{2}}}{(\varDelta n - \sqrt[3]{n})^{\varDelta n - \sqrt[3]{n} + \frac{1}{2}} \cdot \sqrt[3]{n}^{\sqrt[3]{n} + \frac{1}{2}}} \\
&\quad \cdot \frac{n^{n+\frac{1}{2}}}{(n - (1+\delta)\sqrt[3]{n})^{n-(1+\delta)\sqrt[3]{n}+\frac{1}{2}} \cdot ((1+\delta)\sqrt[3]{n})^{(1+\delta)\sqrt[3]{n}+\frac{1}{2}}} \cdot \left(\frac{(1+\delta)\sqrt[3]{n}}{n}\right)^{3\sqrt[3]{n}} \\
&\le cn \cdot \left(\frac{\varDelta n}{\varDelta n - \sqrt[3]{n}}\right)^{\varDelta n + \frac{1}{2}} \cdot (\varDelta n)^{\sqrt[3]{n}} \cdot n^{-\sqrt[3]{n}/3 - 1/6} \\
&\quad \cdot \left(\frac{n}{n - (1+\delta)\sqrt[3]{n}}\right)^{n+\frac{1}{2}} \cdot n^{(1+\delta)\sqrt[3]{n}} \cdot (1+\delta)^{(2-\delta)\sqrt[3]{n} - 1/2} \cdot n^{-(7+\delta)\sqrt[3]{n}/3 - 1/6} \\
&= \frac{cn \cdot \left(\frac{\varDelta n}{\varDelta n - \sqrt[3]{n}}\right)^{\varDelta n + \frac{1}{2}} \cdot \varDelta^{\sqrt[3]{n}} \cdot \left(\frac{n}{n-(1+\delta)\sqrt[3]{n}}\right)^{n+\frac{1}{2}} \cdot (1+\delta)^{(2-\delta)\sqrt[3]{n} - 1/2}}{n^{(2-2\delta)\sqrt[3]{n}/3 + 1/3}} \\
&= o(1)
\end{aligned}$$

which proves (2).

In order to show (3), for the range $cn \le i \le \varDelta n$, let $i = tn$ where $c \le t \le \varDelta$. In this case,

$$\begin{aligned}
f(i) = f(tn) &= \frac{(\varDelta n)^{\varDelta n + \frac{1}{2}}}{(\varDelta n - tn)^{\varDelta n - tn + \frac{1}{2}} \cdot (tn)^{tn + \frac{1}{2}}} \\
&\quad \cdot \frac{n^{n+\frac{1}{2}}}{(n - (1+\delta)tn)^{n-(1+\delta)tn+\frac{1}{2}} \cdot ((1+\delta)tn)^{(1+\delta)tn+\frac{1}{2}}} \cdot \left(\frac{(1+\delta)tn}{n}\right)^{3tn} \\
&= \frac{\sqrt{\varDelta}}{nt\sqrt{(\varDelta - t) \cdot (1 - (1+\delta)t) \cdot (1+\delta)}} \cdot \left[\frac{\varDelta^{\varDelta} \cdot (1+\delta)^{(2-\delta)t} \cdot t^{(1-\delta)t}}{(\varDelta - t)^{(\varDelta - t)} \cdot (1 - (1+\delta)t)^{1-(1+\delta)t}}\right]^n
\end{aligned}$$

Define

$$h(t) = \frac{\varDelta^{\varDelta} \cdot (1+\delta)^{(2-\delta)t} \cdot t^{(1-\delta)t}}{(\varDelta - t)^{(\varDelta - t)} \cdot (1 - (1+\delta)t)^{1-(1+\delta)t}}.$$

Using Mathematica, one can show that there exists some constant $0 < \gamma < 0.970$ such that $h(t) \leq \gamma$ holds when $c \leq t \leq \Delta$. Therefore

$$\begin{aligned}
\Pr[\bigvee_{i=cn+1}^{\Delta n} A_i] &\leq \Pr[\bigvee_{i=cn+1}^{\Delta n-1} A_i] + \Pr[A_{\Delta n}] \\
&\leq \frac{(\Delta - c)\sqrt{\Delta n}}{c\sqrt{(1-(1+\delta)\Delta)\cdot(1+\delta)}} \cdot \gamma^n + \binom{n}{(1+\delta)\Delta n}\left(\frac{(1+\delta)\Delta n}{n}\right)^{3\Delta n} \\
&\leq \frac{(\Delta - c)\sqrt{\Delta n}}{c\sqrt{(1-(1+\delta)\Delta)\cdot(1+\delta)}} \cdot \gamma^n + \left(\frac{e^{1-(1+\delta)\Delta}((1+\delta)\Delta)^{3\Delta}}{(1-(1+\delta)\Delta)^{1-(1+\delta)\Delta}}\right)^n = o(1)
\end{aligned}$$

which shows (3).

Proof (of Lemma 11). Let $\epsilon = \lambda\delta$. By Lemma 3, we need to show that for any $S' \subseteq \psi$, $|S'| - |var(S')\backslash\Psi| \leq (\lambda - \epsilon)n$.

By the expansion of ψ and the fact that there is exactly one variable from Ψ in each clause from U, we have that $|var(S')\backslash\Psi| \geq (1+\delta)|S'| - \min(|S'|, \lambda n)$.

First suppose that $|S'| \geq \lambda n$. Then we have that

$$\begin{aligned}
|S'| - |var(S')\backslash\Psi| &\leq |S'| - ((1+\delta)|S'| - \lambda n) \\
&= \lambda n - \delta|S'| \\
&\leq \lambda n - \delta\lambda n \\
&= \lambda n(1-\delta) \\
&= (\lambda - \epsilon)n
\end{aligned}$$

Now suppose that $|S'| < \lambda n$. Then we have that

$$\begin{aligned}
|S'| - |var(S')\backslash\Psi| &\leq |S'| - ((1+\delta)|S'| - |S'|) \\
&= (1-\delta)|S'| \\
&\leq \lambda n(1-\delta) \\
&= (\lambda - \epsilon)n
\end{aligned}$$

Thus in either case $|S'| - |var(S')\backslash\Psi| \leq (\lambda - \epsilon)n$

Proof (of Theorem 7). Let $\Delta > 0.91$, and let $\phi = (C_1, C_2, \ldots, C_{\Delta n})$ be a random 3-XOR formula on n variables with clause density Δ.

By Theorem 5, with high probability ϕ will be unsatisfiable.

Let $\psi = (C_1, C_2, \ldots, C_{0.6n})$ be the 3-XOR formula that is the first $0.6n$ clauses of ϕ. We will show that there exists $\epsilon > 0$ such that w.h.p $\text{OBDD}_\pi(\psi) \geq 2^{\epsilon n}$ for *any* order π. This implies that $S^+(\phi) \geq 2^{\epsilon n}$, proving the theorem.

Suppose the conclusions of Lemma 9, Lemma 10, and Lemma 6 hold with respect to ψ (which occurs with high probability).

By Lemmas 9 and 10, there exists a δ such that ψ is a $(0.6, 1+\delta)$-expander, and for any order π there exists t_π such that more than δn clauses from ψ are split by t_π according to π. For a given π, let S be a set of δn clauses from ψ such that every clause in S is split by t_π according to π.

By Lemma 6, there exists $\gamma > 0$ and $T \subseteq S$, with $|T| = \gamma n$, such that the clauses of T are disjoint. Either $\frac{\gamma n}{2}$ of the clauses from T are left-split by t_π, or at least $\frac{\gamma n}{2}$ of the clauses from T are right-split by t_π. We now divide our proof into two cases depending on which of these is true.

First assume that there exists $U \subseteq T$, with $|U| \geq \frac{\gamma n}{2}$ such that every clause in U is left-split by t_π. By Lemma 11, there exists $\epsilon > 0$ and $V \subseteq U$, such that

1. $|V| = \epsilon n$
2. There exists an SDR σ on ψ such that no element of left_V is in the range of σ, where $\text{left}_V = var(V) \cap \text{left}_U$

(Here U and left_U correspond to U and Ψ from Lemma 11 respectively.)

Due to the properties of σ, any assignment of the variables of left_V can be extended to satisfy the formula ψ: Define a function $f : \{0,1\}^{|V|} \to \{0,1\}^n$ in such a way that $f(\mathbf{x}) = (\mathbf{x}, \mathbf{y}, \mathbf{z})$, where

1. $\mathbf{x}$ is an assignment to the variables of left_V.
2. $\mathbf{y}$ is an assignment to the variables $var_{t_\pi,\pi} \backslash \text{left}_V$, where $var_{t_\pi,\pi}$ are the first t_π variables of ψ according to the order π.
3. $\mathbf{z}$ is an assignment to $var(\psi) \backslash var_{t_\pi,\pi}$ (i.e. the last $n - t_\pi + 1$ variables according to the order π.)
4. $\psi(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 1$

For $\mathbf{x}, \mathbf{x}' \in \{0,1\}^{|V|}$ with $\mathbf{x} \neq \mathbf{x}'$, let $f(\mathbf{x}) = (\mathbf{x}, \mathbf{y}, \mathbf{z})$ and $f(\mathbf{x}') = (\mathbf{x}', \mathbf{y}', \mathbf{z}')$. We will show that $\psi(\mathbf{x}, \mathbf{y}, \mathbf{z}) \neq \psi(\mathbf{x}', \mathbf{y}', \mathbf{z})$. This implies that $|\{\psi|_{\pi,\mathbf{w}} : \mathbf{w} \in \{0,1\}^{t_\pi}\}| \geq 2^{|V|}$, which by Theorem 3 implies that $|\text{OBDD}_\pi| \geq 2^{|V|} = 2^{\epsilon n}$

By definition we have that $\psi(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 1$. $\mathbf{x}$ and $\mathbf{x}'$ differ on their assignments to some variable in left_V. Let $C \in V$ be the clause containing the variable in left_V on which $\mathbf{x}$ and $\mathbf{x}'$ differ. The other two variables in C are assigned equally in $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ and $(\mathbf{x}', \mathbf{y}', \mathbf{z})$, since their values are determined by the assignment $\mathbf{z}$. Therefore C is not satisfied as a 3-XOR clause by $(\mathbf{x}', \mathbf{y}', \mathbf{z})$, and $\psi(\mathbf{x}, \mathbf{y}, \mathbf{z}) \neq \psi(\mathbf{x}', \mathbf{y}', \mathbf{z})$.

Now we consider the second case, where there exists $U \subseteq T$, with $|U| \geq \frac{\gamma n}{2}$ such that every clause in U is right-split by t_π.

By Lemma 11, there exists $\epsilon > 0$ and $V \subseteq U$, such that

1. $|V| = \epsilon n$
2. There exists an SDR σ on ψ such that no element of right_V is in the range of σ, where $\text{right}_V = var(V) \cap \text{right}_U$

Due to the properties of σ, any assignment of the variables of right_V can be extended to satisfy the formula ψ: Define a function $f : \{0,1\}^{|V|} \to \{0,1\}^n$ in such a way that: $f(\mathbf{z}) = (\mathbf{x}, \mathbf{y}, \mathbf{z})$, where

1. $\mathbf{z}$ is an assignment to the variables of right_V.

2. $\mathbf{y}$ is an assignment to the variables $var(\psi)\backslash(var_{t_\pi,\pi} \cup \text{right}_V)$ (i.e the last $n - t_\pi + 1$ variables according to the order π, not including variables from right_V).
3. $\mathbf{x}$ is an assignment to the variables of $var_{t_\pi,\pi}$.
4. $\psi(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 1$

For $\mathbf{z}, \mathbf{z}' \in \{0,1\}^{|V|}$ with $\mathbf{z} \neq \mathbf{z}'$, let $f(\mathbf{z}) = (\mathbf{x}, \mathbf{y}, \mathbf{z})$ and $f(\mathbf{z}') = (\mathbf{x}', \mathbf{y}', \mathbf{z}')$. We will show that $\psi(\mathbf{x}, \mathbf{y}, \mathbf{z}) \neq \psi(\mathbf{x}', \mathbf{y}, \mathbf{z})$. This implies that $|\{\psi|_{\pi,\mathbf{w}} : \mathbf{w} \in \{0,1\}^{t_\pi}\}| \geq 2^{|V|}$, which by Theorem 3 implies that $|\text{OBDD}_\pi| \geq 2^{|V|} = 2^{\epsilon n}$

By definition we have that $\psi(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 1$. $\mathbf{z}$ and $\mathbf{z}'$ differ on their assignments to some variable in right_V. Let $C \in V$ be the clause containing the variable in right_V on which $\mathbf{z}$ and $\mathbf{z}'$ differ. By definition $\psi(\mathbf{x}', \mathbf{y}', \mathbf{z}') = 1$, so C is satisfied as a 3-XOR clause by the assignment $(\mathbf{x}', \mathbf{y}', \mathbf{z}')$. The only variable in C that is assigned by $\mathbf{y}'$ or $\mathbf{z}'$ is the variable right_C, so the clause C is not satisfied as a 3-XOR clause by $(\mathbf{x}', \mathbf{y}, \mathbf{z})$, and therefore $\psi(\mathbf{x}, \mathbf{y}, \mathbf{z}) \neq \psi(\mathbf{x}', \mathbf{y}, \mathbf{z})$.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de