Extended Resolution Clause Learning via Dual Implication Points

Sam Buss
University of California, San Diego, La Jolla, CA, USA

Jonathan Chung
Lorica Cybersecurity, Canada

Vijay Ganesh
Georgia Institute of Technology, Atlanta, USA

Albert Oliveras
Technical University of Catalonia, Barcelona

Abstract

We present a new extended resolution clause learning (ERCL) algorithm, implemented as part of a conflict-driven clause-learning (CDCL) SAT solver, wherein new variables are dynamically introduced as definitions for *Dual Implication Points* (DIPs) in the implication graph constructed by the solver at runtime. DIPs are generalizations of unique implication points and can be informally viewed as a pair of dominator nodes, from the decision variable at the highest decision level to the conflict node, in an implication graph. We perform extensive experimental evaluation to establish the efficacy of our ERCL method, implemented as part of the MapleLCM SAT solver and dubbed xMapleLCM, against several leading solvers including the baseline MapleLCM, as well as CDCL solvers such as Kissat 3.1.1, CryptoMiniSAT 5.11, and SBVA+CaDiCaL, the winner of SAT Competition 2023. We show that xMapleLCM outperforms these solvers on Tseitin and XORified formulas. We further compare xMapleLCM with GlucoseER, a system that implements extended resolution in a different way, and provide a detailed comparative analysis of their performance.

2012 ACM Subject Classification Theory of computation

Keywords and phrases SAT, Extended resolution, Conflict Analysis

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

Funding *Albert Oliveras*: Supported by grant PID2021-122830OB-C43, funded by MCIN/AEI/10.13039/501100011033 and by "ERDF: A way of making Europe" and by Barcelogic through research grant C-11423.

1 Introduction

Over the last several years, Conflict-Driven Clause-Learning (CDCL) SAT solvers have had a dramatic impact on many fields including software engineering [23], security [39], and AI [22, 21]. As solvers continue to be adopted in increasing complex settings, the demand for greater efficiency and reasoning power by users continues unabated.

While developers continue to improve CDCL SAT solvers, it is simultaneously true that these solvers are provably no more powerful than the relatively weak general resolution (Res) proof system [1, 32], and therefore are fundamentally limited. Hence, solver developers have been actively researching novel algorithms that implement stronger proof systems that go beyond Res. Examples of such algorithms include satisfaction-driven clause-learning (SDCL) SAT solvers [17, 31], bounded variable addition (BVA) [26], symmetry breaking [35], and extended resolution (ER) solvers such as GlucoseER [3].

Continuing this trend of strong proof system implementations, we present a new extended resolution clause learning (ERCL) algorithm, incorporated into a CDCL SAT solver, where new variables are dynamically introduced as definitions for *dual implication points* (DIPs) in the implication graph constructed by the solver at run time. The concept of a DIP is best understood as a generalization of unique implication points (UIPs). Informally, a UIP can be defined as a dominator node in an implication graph, corresponding to a variable at the highest decision level (DL), that *dominates* all paths from the decision variable node at the highest DL to the conflict node. By contrast, a DIP is a pair of dominator nodes in an implication graph such that any path from the decision variable node at the highest DL to the conflict node must pass through at least one node in the pair[1].

Implementation of the ERCL algorithm requires several additional methods. First, we need a method to identify DIPs, i.e., a technique that takes as input an implication graph and outputs a DIP and does so in time linear in the size of the input. Second, we need a technique replaces this DIP pair with a new variable and appropriately modifies the clause learning algorithm to learn new clauses involving DIPs. Third, we need an ER framework, built on top of a CDCL SAT solver, that enables new variable addition, ER clause addition and deletion, etc. Finally, we need heuristics that specialize the above mentioned methods in a variety of ways, such as clause learning and clause deletion policies that are based on different kinds of DIPs. We implement all of these methods as part of the MapleLCM solver [25], and refer to the resulting solver as xMapleLCM. In fact our proposed ERCL method, and its implementation, is very general and easily extensible thus encouraging future exploration and specialization efforts with a variety heuristics.

Contributions.

1. **DIP:** We introduce the concept of dual implication points (DIP), a generalization of UIPs in conflict graphs. We also came up with an algorithm that computes them in linear time. However, due to space limitations and to the non-trivial nature of the procedure, we discuss this in a separate paper [11].
2. **ERCL Algorithm:** We introduce a highly parameterizable DIP-based ERCL algorithm. The existence of a multitude of different DIPs in a single conflict graph allows us to derive a large variety of ERCL algorithms. This flexibility is crucial in adapting the procedure to different scenarios, unlike previous methods that couple CDCL with extended resolution.
3. **xMapleLCM:** We present a highly extensible and general ER framework as part of xMapleLCM, which allows developers to easily add their own new variable addition, ER clause learning/deletion, and branching policies. Given that DIP-based ERCL is highly flexible by nature, such a framework is mandatory to quickly prototype new procedures.
4. **Experiments:** We perform extensive empirical evaluation and ablation studies on 4 different classes of instances, namely, SAT Competition 2023 Main Track, random-k-XOR, Tseitin, and interval matching, and compare xMapleLCM against leading solvers such as Kissat 3.1.1 [8], Cryptominisat 5.11 [36], SBVA CaDiCaL [14], and the extended-resolution solver GlucoseER [4]. Results show that on the last 3 sets of hard combinatorial formulas, CDCL SAT solvers perform very poorly, whereas both xMapleLCM and GlucoseER excel on them. Hence, we can now finally state that extended resolution can be added to CDCL and improve the performance of these solvers, and that there are at least two completely different ways to achieve this. Moreover, a simple heuristic dynamically allows us to detect whether extended resolution is being helpful and turn back to standard CDCL in order to get the best of both worlds. This technique enables xMapleLCM to perform similarly to MapleLCM on the SAT Competition 2023 Main Track instances.

[1] While it is natural to generalize the concept of a DIP to K-Implication Points or k-IPs, we do not discuss them in this paper.

2 Related Work

The idea of using ER in SAT solving has been studied in various forms in the literature for nearly two decades. The closest approach to ours is GlucoseER [3], where extended variables are introduced dynamically during the CDCL search: whenever two consecutive learned lemmas are of the form $\neg l_1 \vee C$ and $\neg l_2 \vee C$, with l_1 and l_2 being their UIPs, then an extended variable $z \leftrightarrow l_1 \vee l_2$ is generated and any future lemma of the form $l_1 \vee l_2 \vee D$ is replaced by $z \vee D$.

Our method differs significantly from GlucoseER in the way extended variables are identified: we choose DIPs, which can be seen as pairs of variables for which adding a definition would create a better first UIP (often written as 1UIP) in the conflict graph, whereas GlucoseER definitions are constructed using already existing UIPs. Also, unlike in GlucoseER, our approach does not always learn the standard 1UIP clause, but multiple clauses might be learned that take into account the newly introduced variable. On the other hand, there are similarities at the heart of both procedures: a certain restriction of ER is considered and introduced variables are used to shorten subsequently found clauses.

More recent uses of ER in SAT solvers are Bounded Variable Addition (BVA) [26] and its structured version SBVA [14], that due to clever strategies are able to identify new extended variables whose introduction can reduce the number of clauses. Essentially, the ultimate goal of BVA techniques is to reduce the size of the formula by reducing the number of clauses at the cost of adding a new variable. One big difference with our work is that this process is done only in a preprocessing step. Finally, one additional direction that has been researched is the development of a BDD-based solver to generate ER proofs [10].

Other approaches aiming at improving CDCL solvers by allowing them to use a more powerful proof system are related to the Propagation Redundancy notion [16, 18], either via preprocessing steps [34] or via the use of the SDCL algorithm [19, 31]. While these methods implement proof systems that are stronger than Res, many of them (without new variable addition) are known to be weaker than ER.

3 Preliminaries

We assume that the reader is familiar with the satisfiability (SAT) problem and the CDCL algorithm, and we refer her to the Handbook of Satisfiability for an excellent overview of these topics [27]. Below we focus on conflict analysis, which is the most relevant ingredient from the CDCL algorithm for this paper. We do so by means of the following example.

▶ **Example 1.** Consider the following clauses

(1)	$y_1 \vee \neg x_1 \vee x_2$	(6)	$\neg x_5 \vee x_7$	(11)	$\neg x_{11} \vee x_{12}$
(2)	$\neg x_1 \vee \neg x_3$	(7)	$x_6 \vee \neg x_7 \vee x_8$	(12)	$x_{10} \vee \neg x_{11} \vee x_{13}$
(3)	$y_2 \vee \neg x_1 \vee x_4$	(8)	$\neg y_3 \vee \neg y_4 \vee \neg x_5 \vee \neg x_9$	(13)	$x_{12} \vee \neg x_{13}$
(4)	$\neg y_3 \vee \neg x_2 \vee x_3 \vee \neg x_4 \vee x_5$	(9)	$\neg y_4 \vee x_9 \vee \neg x_{10}$		
(5)	$y_1 \vee \neg x_5 \vee \neg x_6$	(10)	$\neg y_5 \vee y_6 \vee \neg x_8 \vee x_9 \vee x_{11}$		

Assume that CDCL has constructed an assignment that contains, among others, literals $\{\neg y_1, \neg y_2, y_3, y_4, y_5, \neg y_6\}$. Since no propagation is possible, it now decides to add the decision literal x_1. Due to clause (1), we can unit propagate literal x_2, being (1) the reason of x_2 and its antecedents $\{\neg y_1, x_1\}$. Similarly, $\neg x_3$ is propagated due to reason (2), with antecedents $\{x_1\}$, and x_4 due to reason (3) with antecedents $\{\neg y_2, x_1\}$. If we continue this process we eventually find that clause (13) is conflicting and we can construct the *conflict graph* in

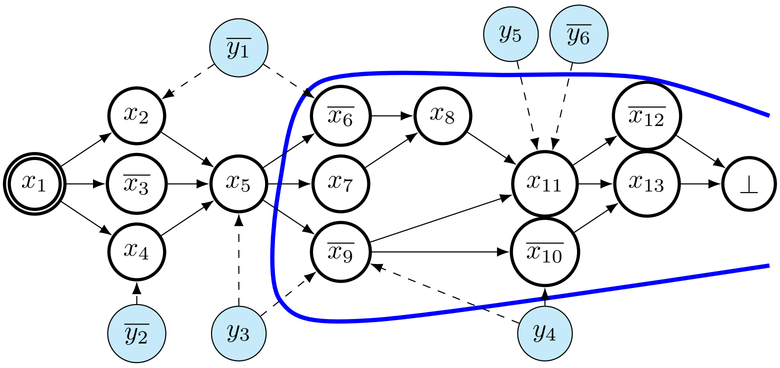


Figure 1 Conflict graph associated with Example 1. If 1UIP learning is applied, we generate lemma $\neg y_4 \vee \neg y_5 \vee y_6 \vee x_5$. White nodes belong to the current decision level, whereas blue ones are from previous decision levels.

Figure 1, where every literal in the current decision level has incoming edges corresponding to its antecedents (except of course the decision literal). A special conflict node $\perp$ with incoming edges $\{\overline{x_{12}}, x_{13}\}$ represents conflicting clause (13).

The graph clearly shows that if we set x_1 to true, together with the literals of previous decision levels (the y's), we obtain a conflict. However, the same happens with x_5, since any path from x_1 to the conflict necessarily goes through x_5. Literals with this property are called *Unique Implication Points* (UIPs), of which we only have x_1 and x_5. Since x_5 is the one closest to the conflict we call it *First Unique Implication Point* (1UIP) [40]. It is easy to see that if we set x_5 and the y literals that enter the cut delimited by the blue line, unit propagation derives the same conflict. Hence, since they cannot be simultaneously true, we can learn $\neg y_3 \vee \neg y_4 \vee \neg y_5 \vee y_6 \vee x_5$. The quality of a lemma can be assessed by its *Literal Block Distance (LBD)* [4]: the number of different decision levels of the literals in the lemma. The lower the LBD, the better the lemma. In our case, if we are at decision level 5, y_4 and y_6 belong to decision level 2, and y_5 to decision level 4, the LBD of the lemma is 3. ◻

Resolution. Given two clauses $l \vee C$ and $\neg l \vee D$, the *resolution* inference rule allows one to derive the logical consequence $C \vee D$. It is well known that the lemma derived in Example 1 can be obtained via a series of resolution steps that start with the conflicting clause, and resolve with reasons of literals in the reverse order in which they were added to the assignment. In fact, it has been proved [33, 2] that resolution and CDCL (with restarts) are polynomially equivalent, and hence classes of formulas that are hard for resolution; e.g., the pigeonhole principle - PHP [15], or Tseitin formulas [38]) are also hard for CDCL SAT solvers.

Extended Resolution (ER). Given two literals l_1 and l_2, the *extended resolution* [37] rule allows us to introduce clauses representing the definition $z \leftrightarrow l_1 \vee l_2$. ER can be substantially more powerful than resolution; for instance, it allows polynomial size proofs of PHP [13]. Incorporating ER to CDCL solvers could potentially enable them to solve such formulas in polynomial time. However, we lack good methods to incorporate ER into CDCL proof search. This is precisely the aim of this paper, namely, incorporating a restricted version of ER into CDCL.

4 Dual Implication Points

As discussed in the previous section, the unique implication points (UIPs) are crucial for conflict clause learning in the CDCL algorithm. We now introduce a new concept of a *Dual*

Dual Implication Point (DIP)	Extension Variable	Post-DIP Learned Clause	Pre-DIP Learned Clause
$\overline{x_{12}}, x_{13}$	$z \leftrightarrow (\overline{x_{12}} \wedge x_{13})$	$\neg z$	$\neg x_5 \vee y_1 \vee \neg y_3 \vee \neg y_4 \vee \neg y_5 \vee y_6 \vee z$
x_{11}, x_{13}	$z \leftrightarrow (x_{11} \wedge x_{13})$	$\neg z$	$\neg x_5 \vee y_1 \vee \neg y_3 \vee \neg x_4 \vee \neg y_5 \vee y_6 \vee z$
$\overline{x_{10}}, x_{11}$	$z \leftrightarrow (\overline{x_{10}} \wedge x_{11})$	$\neg z$	$\neg x_5 \vee y_1 \vee \neg y_3 \vee \neg y_4 \vee \neg y_5 \vee y_6 \vee z$
$\overline{x_9}, x_{11}$	$z \leftrightarrow (\overline{x_9} \wedge x_{11})$	$\neg z \vee \neg y_4$	$\neg x_5 \vee y_1 \vee \neg y_3 \vee \neg y_4 \vee \neg y_5 \vee y_6 \vee z$
$x_8, \overline{x_9}$	$z \leftrightarrow (x_8 \wedge \overline{x_9})$	$\neg z \vee \neg y_4 \vee \neg y_5 \vee y_6$	$\neg x_5 \vee y_1 \vee \neg y_3 \vee \neg y_4 \vee z$

Figure 2 The complete list of DIPs for the conflict graph of Figure 1. The DIP-learnable clauses involve the new extension variable z that can be introduced for that DIP. (The extension clauses defining z must also be learned; e.g., for the first line, the extension clauses express that $z \leftrightarrow \overline{x_{12}} \wedge x_{13}$.)

Implication Point (DIP) that gives a tool for analyzing the conflict graph. Its applications include discovering new implied 2-clauses, introducing new variables by extension, and learning clauses involving the extension variables. The idea behind a dual implication point is that it consists of a pair of vertices (literals) in the conflict graph that disconnects or "separates" the decision literal from the contradiction.[2] More precisely, a DIP is defined to be a pair $\{x, y\}$ of literals such that all paths in the conflict graph to the vertex $\perp$ pass through at least one of x and y and such that neither x nor y is a UIP. In contrast, UIP is a *single* literal that separates the decision literal from the conflict.

We use the example in Figure 1 to illustrate the concept of DIPs and their potential applications. Recall that x_5 is the first UIP. In our applications, we are seeking DIPs between the first UIP and the conflict node $\perp$. An obvious DIP is the pair $\overline{x_{10}}$ and x_{11}, since it is immediately clear that any path from x_5 (or from x_1) must pass through one of $\overline{x_{10}}$ or x_{11}. On the other hand, the pair $\overline{x_{10}}$ and x_8 is not a DIP since there are paths from x_1 to $\perp$ that avoid these two literals; namely, any path that includes the edge from $\overline{x_9}$ to x_{11}. There are several other DIPs in Figure 1: a complete list is given in Figure 2.

Figure 2 also shows how a DIP pair can be used to introduce a variable z via extension, and the associated clauses that can be learned. For example, in the third line, the new extension variable z is introduced with the three clauses $\neg z \vee \neg x_{10}$, $\neg z \vee x_{11}$ and $x_{10} \vee \neg x_{11} \vee \neg z$ which express the condition $z \leftrightarrow (\neg x_{10} \wedge x_{11})$. From the conflict graph, this allows inferring the clauses $\neg z$ (the "post-DIP" learned clause) and $\neg x_5 \vee y_1 \vee \neg y_3 \vee \neg y_4 \vee \neg y_5 \vee y_6 \vee z$ (the "pre-DIP" learned clause). Since the post-DIP learned clause does not have any variables from lower decision levels, we can also infer the 2-clause $x_{10} \vee \neg x_{11}$ either instead of or in addition to introducing z and the pre- and post-DIP clauses. Introducing 2-clauses in this way might be helpful for CDCL solvers that do special processing of 2-clauses; for instance, in the work of Bacchus et al. [5] or the recent work of Biere et al. [9] or Buss et al. [12].

In general, for any literals a, b that form a DIP in the above fashion, we may introduce an extension variable $z \leftrightarrow a \wedge b$, and learn (1) a pre-DIP clause of the form $\neg f \vee \neg C \vee z$ (i.e. $f \wedge C \rightarrow z$), where f is the first UIP and C the set of literals from previous decision levels that have an edge in the conflict graph to any literal appearing after the first UIP and no later than the DIP pair; and (2) a post-DIP clause of the form $\neg z \vee \neg D$ (i.e. $z \wedge D \rightarrow \perp$), where D contains those literals from previous decision levels with an edge to any literal appearing strictly after the DIP pair.

[2] Perhaps "Dual Implication Pair" would be a better name than "Dual Implication Point', since a DIP is a pair of literals. We use "Point" however to match the terminology of "Unique Implication Points".

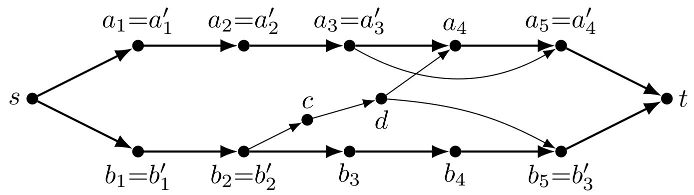


Figure 3 Nodes a_4, b_3 and b_4 are bypassed and thus not in any TVB. The path b_2, c, d, a_4 is a crossing separator that prevents any of a_1, a_2 and a_3 from being paired with b_5 to form a TVB pair. All other pairs $\{a'_i, b'_j\}$ are valid TVB pairs.

For example, the last line of the table in Figure 2 shows the case $z \leftrightarrow (x_8 \wedge \overline{x_9})$, where f is x_5, and C and D are $\neg y_1 \wedge y_3 \wedge y_4$ and $y_4 \wedge y_5 \wedge \neg y_6$, respectively. In Section 5.2 we discuss many possible ways that DIP extension variables and clauses may be introduced into CDCL solvers. The rest of this section describes how to find DIPs.

4.1 An Algorithm for Finding DIPs

A conventional CDCL algorithm maintains a trail of the literals set true, in the order they were set, and this allows finding the UIP very quickly. Finding the DIPs is much more complex than finding the UIP, and requires several traversals of the conflict graph between the first-UIP and the conflict node. Nonetheless, it is possible to find all DIPs very quickly, even in linear time.

We recast the DIP-finding problem in terms of a general directed graph G. Let $G = (V, E)$ be a directed graph with two distinguished vertices s and t. We assume that s is 2-connected to t in that there is a pair of vertex-disjoint paths π_1 and π_2 from s to t that share no vertices apart from s and t. By the vertex-version of Menger's theorem [28], either there are in fact *three* vertex-disjoint paths from s to t or there is at least one pair of vertices $\{a, b\}$ with neither a nor b in $\{s, t\}$ so that every path from s to t passes through at least one of a or b. Such a pair $\{a, b\}$, if it exists, is called a *Two Vertex Bottleneck* (TVB).

Our goal is to find all possible TVBs efficiently and in linear time. An algorithm for this is discussed in our companion paper [11]; for space reasons, we give only an abbreviated discussion of the algorithm here. The first step is to find two vertex disjoint paths from s to t: this is done by greedily finding a path from s to t and then using an augmenting path construction to find two vertex-disjoint paths from s to t. An example is shown in Figure 3, where the two paths are $a_0, \ldots, a_\ell$ and $b_0, \ldots, b_k$ where $a_0 = b_0 = s$ and $a_\ell = b_k = t$. These paths are called π_a and π_b, respectively. Henceforth, a *path* is a directed path without any repeated nodes. The *internal* vertices of a path π are the vertices on π other than the first and last vertices. Two paths are said to be *vertex-disjoint* if they have no internal vertices in common. A path π *avoids* π_a *and* π_b if it is vertex-disjoint from both paths.

Once the two vertex-disjoint paths are fixed, we can state the following definitions and theorems:

▶ **Definition 2.** [11] *A node a_i on π_a is* bypassed *if there are $j < i < j'$ and a path π from a_j to $a_{j'}$ such that π avoids π_a and π_b. A node b_i being bypassed is defined similarly.*

▶ **Definition 3.** *Two nodes a_i and b_j have a* crossing separator *if there are nodes $a_{i'}$ and $b_{j'}$ joined by a path π that avoids both π_a and π_b such that either (a) $i' < i$ and $j' > j$ and π is a path from $a_{i'}$ to $b_{j'}$, or (b) $i' > i$ and $j' < j$ and π is a path from $b_{j'}$ to $a_{i'}$.*

▶ **Theorem 4.** [11] *For $0 < i < \ell$ and $0 < j < k$, the two nodes a_i and b_j form a two-vertex bottleneck (TVB) if and only if a_i and b_j do not have a crossing separator and neither a_i*

nor b_j is bypassed.

Theorem 4 is proved in [11]. The theorem holds for both acyclic and cyclic directed graphs; however, for our applications to CDCL we are interested only in acyclic graphs since the conflict graph is always acyclic.

A consequence of Theorem 4 is that the set of all TVBs can be compactly represented in linear size, even though there can be quadratically many TVBs. Let $a'_1, \ldots a'_{\ell'}$ be the subsequence of the internal nodes $a_1 \ldots, a_{\ell-1}$ of path π_a that, according to the conditions of Theorem 4, are in at least one TVB pair. Let $b'_1, \ldots, b'_{k'}$ be the corresponding subsequence of the internal nodes of π_b. Then, for each a'_i there are $m \le n$ such that a'_i forms a TVB pair with each b'_j with $m \le j \le n$. Dually, for each b'_i there are $m \le n$ such that b'_i forms a TVB pair with each a'_j with $m \le j \le n$. (See Figure 4.)

▶ **Example 5.** In the conflict graph of Figure 1, consider the portion of the graph between the first UIP x_5 and the contradiction $\perp$. We can take path π_a to be $x_5, x_7, x_8, x_{11}, \overline{x_{12}}, \perp$ and path π_b to be $x_5, \overline{x_9}, \overline{x_{10}}, x_{13}, \perp$. The node x_7 is bypassed by the path $x_5, \overline{x_6}, x_8$, so it cannot be part of a TVB. There are two crossing separator paths: the first is the edge from $\overline{x_9}$ to x_{11}; the second is the edge from x_{11} to x_{13}. Therefore, the possible TVB pairs can be described in a table as:

Node on π_a	forms a TVB pair with	Node on π_b	forms a TVB pair with
x_8	$\overline{x_9}$	$\overline{x_9}$	x_8, x_{11}
x_{11}	$\overline{x_9}, \overline{x}_{10}, x_{13}$	$\overline{x_{10}}$	x_{11}
$\overline{x_{12}}$	x_{13}	x_{13}	$x_{11}, \overline{x}_{12}$

In this example, one node, x_7, was bypassed. It is also possible that non-bypassed nodes are eliminated just by the crossing separators. For example, if there were an additional edge from x_7 to $\overline{x_{10}}$, then the crossing separator condition would imply that x_8 and x_{11} are not part of any TVB pair. In this case, x_8 and x_{11} would not be included among the a'_i nodes.

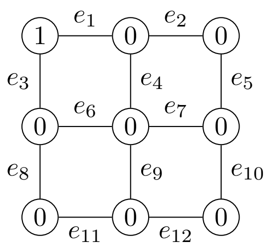
Theorem 4 is used in [11] to give an efficient, linear time algorithm for finding all TVBs. The algorithm has five phases. The first two phases find two vertex-disjoint paths from s to t; the next phase scans the graph from t to s to discover all relevant paths that avoid π_a and π_b; the fourth phase uses this to discard bypassed nodes and collect information on crossing separators; finally the fifth phase computes the compressed representation of all possible TVBs. Full details are given [11], which is available online in preprint form. In our experiments with the implication graphs constructed by the underlying CDCL solver, the time overhead in finding the TVBs is negligible.

5 Extension Variables from Dual Implication Points

This section discusses how DIPs can be used for introducing extension variables and implement the ERCL method for learning ER clauses in CDCL solver. We first present an example.

5.1 An example with a grid Tseitin principle

Given a graph where every vertex has a *charge*, a number which is 0 or 1, a Tseitin formula [37] is created by considering one variable per edge, and adding one constraint per vertex v expressing that the sum of the variables of all edges incident to v modulo 2 is equal to its charge. The CNF version that we consider converts each *xor* constraint into clauses by simple



$$
\begin{aligned}
e_1 \oplus e_3 &= 1 & e_5 \oplus e_7 \oplus e_{10} &= 0\\
e_1 \oplus e_2 \oplus e_4 &= 0 & e_8 \oplus e_{11} &= 0\\
e_2 \oplus e_5 &= 0 & e_9 \oplus e_{11} \oplus e_{12} &= 0\\
e_3 \oplus e_6 \oplus e_8 &= 0 & e_{10} \oplus e_{12} &= 0\\
e_4 \oplus e_6 \oplus e_7 \oplus e_9 &= 0 & &
\end{aligned}
$$

Figure 4 An instance of the 3×3-grid Tseitin principle. The nodes are assigned a polarity in $\{0, 1\}$. The edges are labeled with variables.

enumeration of falsifying assignments. It is easy to see that the formula is unsatisfiable if and only if the sum of all charges is even. Here we consider as a graph the 3×3-grid depicted in Figure 4 and only one vertex has charge 1, so the clauses are unsatisfiable.

The first steps of the CDCL solver are as follows: First e_1 is set true as a decision literal, and $\overline{e_3}$ is (unit) propagated. Second, e_2 is set true as a decision literal, and $\overline{e_4}$ and e_5 are propagated. Third, e_6 is set true as a decision literal, and e_8 and e_{11} are propagated. Fourth, e_7 is set true as a decision literal, and $\overline{e_9}$, $\overline{e_{10}}$, and e_{12} are propagated. This gives a contradiction, since the clause $e_{10} \vee \overline{e_{12}}$ (one of the two clauses from $e_{10} \oplus e_{12} = 0$) is falsified. Figure 5(a) shows the complete conflict graph at this point.

Examining the conflict graph at decision level 4, the first UIP is e_7 and there are two DIPs available, $\{\overline{e_9}, \overline{e_{10}}\}$ and $\{\overline{e_{10}}, e_{12}\}$. Selecting the former DIP, we introduce a new variable x by extension as $x \leftrightarrow \overline{e_9} \wedge \overline{e_{10}}$. We can learn the additional pre- and post-DIP clauses: $e_4 \vee \overline{e_5} \vee \overline{e_6} \vee \overline{e_7} \vee x$ and $\overline{x} \vee \overline{e_{11}}$

We next backtrack to decision level 3, unsetting e_7, e_9, e_{10} and e_{12}. Unit propagation at decision level 3 sets the new literal x and the first UIP $\overline{e_7}$ false and then sets literals e_9, e_{10}, and $\overline{e_{12}}$ true. This yields a contradiction with the clause $\overline{e_{10}} \vee e_{12}$. In the conflict graph at decision level 3, the first UIP is e_6 and there are two DIPs available, namely $\{\overline{e_7}, \overline{e_{12}}\}$ and $e_{10}, \overline{e_{12}}\}$. If we select the first one, then we introduce a new literal y defined by extension as $y \leftrightarrow \overline{e_7} \wedge \overline{e_{12}}$ and can in addition learn the clauses $e_3 \vee e_4 \vee \overline{e_5} \vee \overline{e_6} \vee y$ and $\overline{e_5} \vee \overline{y}$.

We do not carry this example further, but note that our experiments show that Tseitin tautologies (not just on grid graphs) are examples where our experiments show the DIP clause learning method is especially effective.

It is interesting to relate the Figure 5(b) to Theorem 4 on DIPs. In this example, there is only one way to choose the two vertex-disjoint paths. Namely, to let π_a and π_b be the paths $e_6, e_8, e_{11}, \overline{x}, \overline{e_7}, e_{10}, \perp$ and $e_6, e_9, \overline{e_{12}}, \perp$. The edge from e_6 to $\overline{e_7}$ bypasses e_8, e_{11} and $\overline{x}$; and the edge from e_{11} to $\overline{e_{12}}$ is a crossing separator. (The edge from $\overline{e_7}$ to e_9 is a "vacuous" crossing separator that does not actually remove any possible DIP pairs.)

It should be evident from this example that many conflicts have DIPs; indeed our experiments reported below show that approximately 2/3 of the conflicts have at least one DIP and, very frequently there are quite a few choices for DIPs.

5.2 Extending CDCL with Dual Implication Points

The use of DIPs in conflict analysis opens a large spectrum of possibilities. This section discusses some of them, with particular attention to the techniques that we have implemented. Our present implementation, xMapleLCM, is a flexible framework that allows one to implement extended-resolution based techniques in a simple way. It offers a set of clearly-specified functions that facilitate determining which extension variables to add,

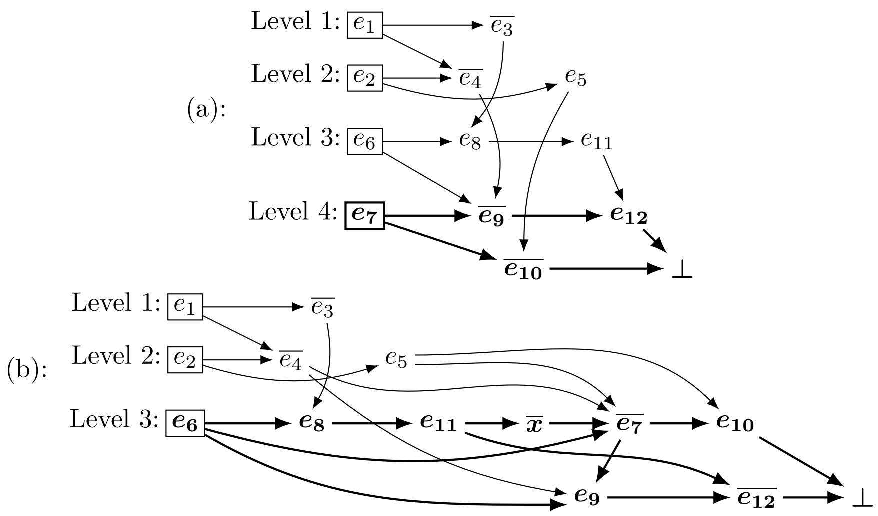


Figure 5 The complete conflict graph from the example; (a) at the first conflict and (b) at the second conflict. Boxed literals are decision literals. Implications at the top decision level are in bold.

performing the corresponding addition, the possible posterior deletion or replacing definitions in clauses. Additionally, a set of heuristic choices to control when and how these steps are performed are provided, and replacing them by custom ones is a smooth task. We want to remark that there are many more possible strategies for using DIPs than could be discussed. We believe that in this paper we merely scratched the surface of heuristics for exploiting DIPs, which is an indication of the potential of this approach.

Choice of DIP. As mentioned, there is possibly a quadratic number of DIPs. Even though we could learn multiple DIPs at every conflict, with their corresponding lemmas, we decided to choose only one. The first possibility we considered is to learn the DIP that is closest to the conflict, as we do with UIPs. This may often create a short post-DIP conflict but a long pre-DIP clause. Therefore, we considered the possibility of choosing a DIP that splits the conflict graph into two balanced regions. Ideally, that would result in two equally long pre and post-DIP clauses. Finally, we also implemented choosing a random DIP, to check whether any of the other two schemes could outperform a random strategy. These heuristics are referred to as **closest**, **middle**, and **random**, respectively.

Filtering out bad-quality DIPs. Learning a DIP whenever we find one would be too prolific and overwhelm the solver. In order to determine whether the DIP chosen in the previous step had to be discarded, the first possibility we explored was again inspired by 1-UIP learning, where learning glue clauses, i.e. having LBD equal to 2, is a desired situation. The first filtering mechanism we implemented discarded all DIPs that did not have a glue post-DIP clause. Another possibility we considered is to wait for a DIP to occur a certain number of times before using it in DIP-based learning. In our implementation, we tried 20 and 5 occurrences, as the minimum threshold to introduce a DIP. A third possibility is to use the activity-based heuristic of the literals in the DIP to assess its quality: DIPs whose literals have high decision-heuristic scores should be prioritized. In our implementation, we check whether the activity level of the current DIP is higher than the average activity level of the 20 most recently encountered DIPs; if so, the current DIP is a candidate for DIP-learning, otherwise it is discarded. We refer to these various techniques as **glue**, **occ5** and **act**, respectively.

Learning pre-DIP and post-DIP clauses. In our implementation, we only considered two variants: one that always learns both the pre- and the post-DIP clauses (**2-clause**) and one that only learns the post-DIP clause (**1-clause**).

Backjumping and asserting clauses. Recall that when a new DIP extension variable z is introduced, it is possible to learn a pre-DIP clause of the form $\neg f \vee \neg C \vee z$ and a post-DIP clause of the form $\neg z \vee \neg D$, where f is the first UIP and where C and D are conjunctions of literals that were set true at lower levels. (Note that C and D may have literals in common.) Letting ℓ_C and ℓ_D be the maximum of the levels at which literals in C and D (respectively) were set, our implementation always backtracks to level ℓ_D. This makes the post-DIP clause asserting, so $\neg z$ is set at decision level ℓ_D. Furthermore, if $\ell_C \leq \ell_D$ and the pre-DIP clause is learned, then it is asserting and $\neg f$ is set by unit propagation at decision level ℓ_D, as it should be.

It would make no sense to backtrack to level ℓ_C when $\ell_C < \ell_D$ since then neither $\neg z$ nor $\neg f$ would be propagated. However, another possible strategy would be to backtrack to the maximum of the decision levels ℓ_C and ℓ_D. This would mean $\neg z$ and $\neg f$ are both propagated. The disadvantage of this when $\ell_C > \ell_D$ is that it would mean $\neg z$ is asserted by the post-DIP clause at level ℓ_C, whereas it could have been propagated at the previous decision level ℓ_D. This breaks a usual invariant for CDCL solvers. This would only be possible in a solver that permits chronological backtracking [30, 29]; xMapleLCM, however, does not support this.

The previous reasoning needs some clarification for the case when the extension variable z with definition $z \leftrightarrow l_1 \wedge l_2$ we want to introduce already exists. If z is undefined or defined at the current decision level nothing changes. We know it cannot be true at some previous level, because otherwise l_1 and l_2 would have been propagated at that level, and not in the current one as all literals belonging to a DIP do. If it is false at some previous level, then we have no guarantee that the pre or the post-DIP clause is asserting at any decision level. Fortunately, that rarely happens in practice. However, we can always perform standard 1UIP learning or try to apply DIP-based conflict analysis starting with the clause $z \vee \neg l_1 \vee \neg l_2$ that we can guarantee is conflicting at the current decision level.

Replacing literals by extended variables. In xMapleLCM, every time a new lemma is learned, we try to replace some of its literals by extended variables. An extended variable $z \leftrightarrow l_1 \vee l_2$, allows one to replace a lemma of the form $l_1 \vee l_2 \vee C$ by $z \vee C$. This is done by checking, for all pairs of literals in the lemma that appear in some extended variable definition whether they are part of the same definition. Too long lemmas or lemmas that have large LBD are discarded to mitigate the cost of this operation.

If one wants to introduce literal $\neg z$ and obtain some reduction in formula size, the lemma should be of the form $\neg l_1 \vee C$ and there should be another clause of the form $\neg l_2 \vee C$. In this case, they can be replaced by a single clause $\neg z \vee C$. However, this situation can be expensive to detect since one has to traverse the whole database looking for a certain clause, and this operation should be repeated for every literal in the lemma. For this reason, xMapleLCM does not implement this.

Deleting extended variables. As it happens with lemma learning, where keeping too many lemmas slows done unit propagation, managing too many extended variables might also be counterproductive. Again following the analogy with learned lemmas, which are useful at some point of the search but might become inactive after a while, it is natural to think that extended variables follow the same behavior. All in all, it seems mandatory to consider the deletion of extended variables.

Deletion of variables in xMapleLCM is scheduled to be performed every 1000 conflicts.

At that point, several strategies are possible: delete all variables, delete the ones with a minimum decision-heuristic activity, or delete the worst $k\%$ variables according to some criterion (e.g. their decision-heuristic activity). In our implementation, we do the latter with $k = 50$. Note that variables appearing in the right-hand side of an extended-variable definition cannot be deleted. This is addressed by maintaining a counter for every variable that corresponds to the number of definitions where it is involved.

6 Experimental Evaluation

We have implemented the DIP-based clause learning schemes described in Section 5.2 on top of the xMapleLCM ER framework[3]. We started our experimental evaluation running a variant of DIP-based learning on all benchmarks of the 2023 SAT Competition [6] and comparing it to MapleLCM [25], the CDCL SAT solver on which it is based. Even though there did not seem to be a systematic improvement on all benchmarks, a few families with important speedups were identified. We start this section by analyzing the impact of DIP-based learning on these families and then move to final considerations about the performance on the overall 2023 SAT competition benchmarks.

6.1 Performance Analysis Methodology

For each family, we start by describing the problem they encode. After that, we report on the performance a variety of state-of-the-art solvers, each with some distinguished characteristic: Kissat 3.1.1 [8] (an extremely efficient CDCL solver), Cryptominisat 5.11 [36] (support for XOR reasoning), SBVA+CaDiCaL [14] (introduction of new variables via SBVA and winner of the main track of the 2023 SAT Competition), GlucoseER [3] (extended-resolution based CDCL solver) and a variant of xMapleLCM-DIP, which we refer to as the baseline, that we describe next.

Finally, we evaluate the impact of the different techniques explained in Section 5.2. In particular, we first consider as a **baseline** a version that (i) finds DIPs in the middle of the conflict graph, (ii) only adds a DIP if it has occurred at least 20 times and (iii) always learns the both pre- and post-DIP learning clauses. Different variants can be obtained by changing only one of the three previous design decisions at a time. Regarding the type of DIP used, we analyze the performance of **closest** and **random**, that is, the systems whose only difference w.r.t. **baseline** is that the type of DIP is changed. Regarding the criterion used to discard a DIP, we analyze the **glue**, **act** and **occ5** configurations, where the latter discards any DIP that has not occurred at least 5 times. Finally, the **1-clause** variant only learns the post-DIP clause. Our goal is to understand why certain configurations do not work well on particular families.

6.1.1 Matching of Properly Intersecting Intervals

We are given a sequence of numbers $(a_1, \ldots, a_n)$, initially all set to zero, and a set of operations, each consisting of assigning a certain number to a contiguous subsequence of the a_i's, defined by an interval. The goal is to perform each operation exactly once while maximizing the number of pairs of consecutive numbers (a_i, a_{i+1}) that are different at the end of the process. Given some additional conditions, this can be formulated as maximum

[3] All sources used for this evaluation can be found in `https://github.com/chjon/xMapleSAT/tree/main`.

	Baseline	1-Clause	Closest	Rand	Glue	Occ5	Act
PROPERLY INTERSECTING INTERVALS FORMULAS							
SOLVED	6	5	**7**	**7**	0	0	0
BEST	1	1	2	**3**	0	0	0
Median (secs.)	2777	2566	2358	**2189**	3600	3600	3600
TSEITIN FORMULAS							
SOLVED	11	13	13	11	0	**15**	3
BEST	2	5	0	0	0	**8**	0
Median (secs.)	897	628	2707	2782	3600	**425**	3600
(X)ORIFIED RANDOM k-xor FORMULAS							
SOLVED	10	8	**11**	9	0	9	7
BEST	**5**	2	1	2	0	1	0
Median (secs.)	**430**	803	1080	820	3600	936	3112

Table 1 Performance of DIP-learning variants on three sets of formulas. Median running times are in seconds and, for each row, the best performing system is in bold.

bipartite matching problem on a certain graph. By removing some edges from this graph, an unsatisfiable problem is generated. A more detailed description can be found in [6].

We downloaded the 23 instances submitted to the 2023 SAT Competition, using the Global Benchmark Database [20], and observed that no system without extended-resolution reasoning could solve any benchmark in a time limit of 1 hour. On the other hand, there are 14 benchmarks that GlucoseER could solve, of which our DIP-based variants could solve 7. Detailed results for our variants can be seen in Table 1.

For this particular type of formulas, the **rand** variant performs the best, but detailed results indicate that we cannot claim it is systematically faster. One clear conclusion we can infer is that **glue**, **occ5** and **act** exhibit very poor performance. A more detailed analysis on the runs of these systems reveal that, at least in these benchmarks, both **act** and **occ5** apply DIP learning in many more conflicts that the other variants, whereas the distinctive characteristic of **glue** is that the percentage of decisions on extended variables was extremely low. As can be seen at the end of this section, the latter is a situation than tends to indicate the DIP learning is not helping the solver in this context.

6.1.2 Tseitin Formulas

These formulas have already been described in Section 5.1. In this section, we consider unsatisfiable formulas generated by CNFgen [24]. We started generating instances with grids of size $n \times n$ and ran all systems with a time limit of 300 seconds in order to analyze how different solvers scale. The performance of all systems can be seen on the left plot in the first row of Figure 6 (note the logarithmic scale on the y axis). A point (x, y) in a solver line indicates that the solver took y seconds to solver the Tseitin formula on grid size $x \times x$.

These formulas are easy for Cryptominisat, since the application of Gaussian elimination in a preprocessing step solves them. Regarding the rest of the systems, only our DIP-based tool and GlucoseER have good scaling properties. We want to remark that this behavior is not unique to Tseitin formulas on the grid. We generated Tseitin formulas over random 4-regular and 6-regular graphs and the conclusions are the same as can be seen in the other two plots in the first row of Figure 6. These results start suggesting that, even though GlucoseER and our DIP-based learning are quite different approaches to integrating ER into

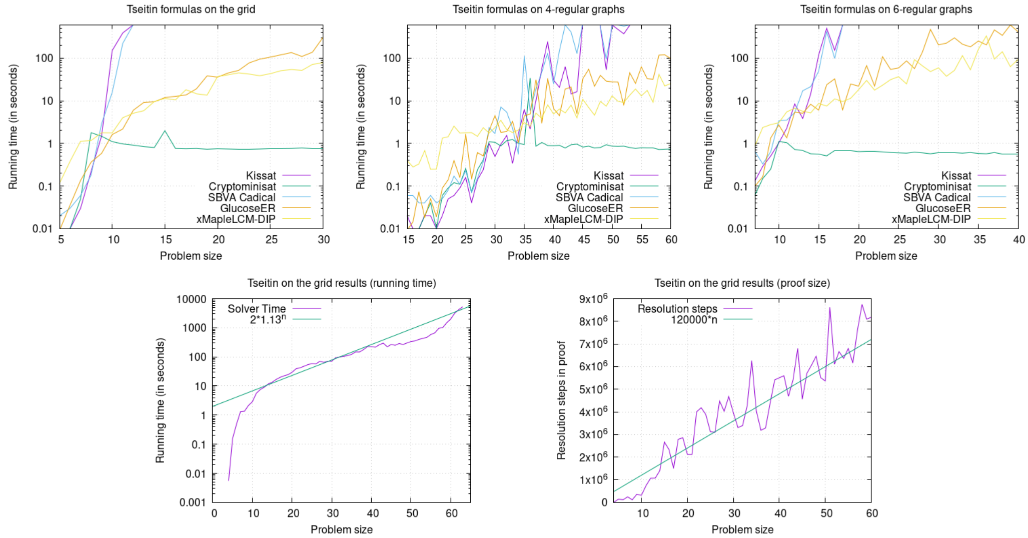


Figure 6 Performance analysis of different solvers on Tseitin formulas.

CDCL solvers, they seem to perform well on the same sets of benchmarks.

In order to understand whether the behavior of xMapleLCM-DIP was polynomial, we generated more challenging instances. Results can be seen in the left plot on the second row of Figure 6, with logarithmic scale on the y axis. There is little doubt that the runtime of our solver ends up being exponential w.r.t. the size of the problem. However, we went one step further and studied how large were the generated DRAT proofs. In the same row, but on the right, we show the number of resolution steps in the proof. Despite we have no theoretical support for that, it seems that although the solver takes exponential time in finding a proof, its size might be polynomial w.r.t. the problem size. That would indicate that our problem is having search heuristics that are not good enough to quickly find a short proof.

Finally, as we did in the previous section, we want to study the effect of the possible variants of DIP-based learning. For that purpose, we selected 5 challenging benchmarks of each type (grid, 4-regular, 6-regular) and evaluated the 7 DIP variants that we want to examine. Results can again be found in Table 1.

For Tseitin formulas, **occ5** performs much better than the rest. Its only difference is that the minimum number of occurrences for a DIP to be introduced is smaller; this suggests that, in these instances, being more aggressive in the introduction of DIP extension variables provides a competitive advantage. On the other hand, **glue** solves no instance. We realized that DIP-learning was hardly ever applied, that is, requiring glue clauses was too restrictive here. Regarding **act**, whose performance is similarly poor, the main noticeable difference in its behavior is exhibiting a very large number of conflicts where DIP learning is applied, between 20 and 50%, which is around ten times more than the variants that performed best.

6.1.3 (X)Orified Random k-XOR Formulas

On last family where DIP-based systems perform very well are random k-xor formulas, where orification or xorification [7] has been applied, i.e., replacing variables by (x)ors of fresh variables. We used CNFgen to obtain 4 sets of formulas: xor constraints of length 3 applying orification with 2 and 3 variables, and also with xorification of 2 and 3 variables. All formulas we consider are unsatisfiable, the number of variables before (x)orification is equal to the number of clauses, and we increase such number to get progressively more difficult formulas.

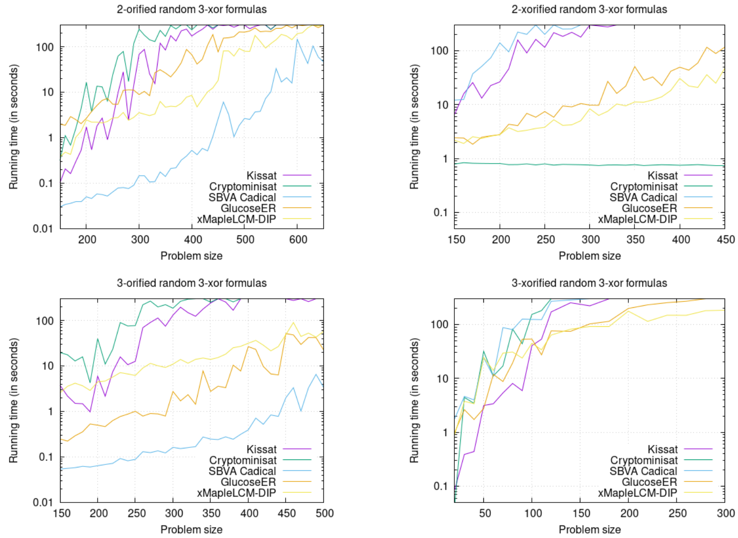


Figure 7 Performance analysis of different solvers on (x)orified random k-xor formulas.

We studied the scaling properties of all systems by running them on increasingly larger benchmarks with a time limit of 300 seconds. Results can be seen in Figure 7, where the y axis has logarithmic scale. Maybe surprisingly, Cryptominisat can only benefit from its preprocessing step in one of the families (*2-xorified*). Also, SBVA-CaDiCaL performs very well in two of them (*2 and 3-orified*). Again, both GlucoseER and our DIP-based implementations show the best average overall performance over all competitors.

Finally, an analysis of the impact of the different DIP-learning variants can be found in Table 1. Our baseline implementation seems to perform the best, but there are several other configurations that perform very well. On the negative side, **glue** did not solve any such formula within the time limit. The identifying characteristic of that version is that the percentage of conflicts with DIP-learning is very low compared to the others (around 0.1%).

6.2 Performance Analysis on 2023 SAT Competition Benchmarks

Our experimental evaluation concludes with the lessons we have learned from executing our DIP-variants[4] on all benchmarks from the 2023 SAT Competition. We first report on the overhead caused by the DIP detection algorithm and the subsequent additional work to retrieve the clauses to be learned. For our baseline DIP-based system, where two clauses are learned and hence is the most computationally demanding method, in 4.5% of the benchmarks the DIP-related work represented between 10 and 15% of the total runtime; in 11.5% of the benchmarks between 5 and 10%; in 35.5% between 2.5 and in 48.5% less than 2.5%. These data show that DIP computation does not significantly slow down the solver. Another interesting information concerns the percentage of conflicts where there is at least one DIP, which was on average 63%. This implies that we do need to have filtering mechanisms to discard some of them. Otherwise, the search would be totally dominated by DIPs.

A cactus plot with the results of executing all our DIP variants on the 2023 SAT Competition benchmarks with a time limit of 1 hours is in the left plot of Figure 8. Three

[4] We do not present a comparison with state-of-the-art solvers like Kissat or CaDiCaL since there are not many lessons to be learned. In a few words, our MapleLCM baseline CDCL solver is slower that those systems, and adding DIP-reasoning to it does not bridge this gap.

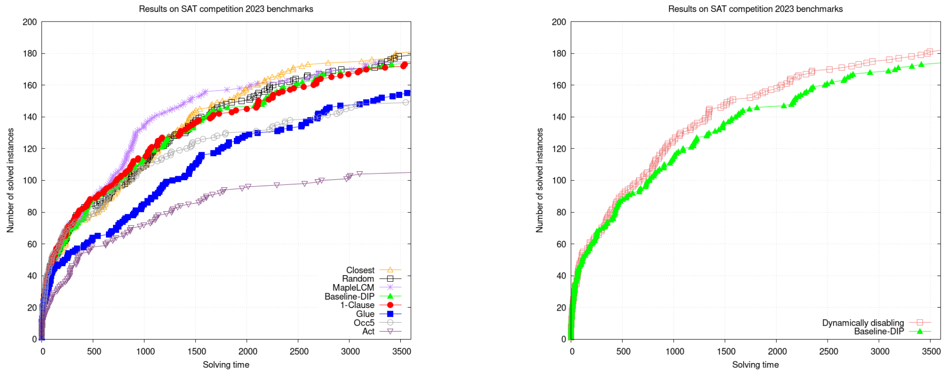


Figure 8 Performance of DIP variants on 2023 SAT Competition benchmarks.

variants are much worse than the rest: **act**, **occ5** and **glue**, whereas the other are relatively similar. These three configurations have a much larger percentage of conflicts where DIP-based learning is applied than the rest, being **act** around 30% and the other two around 5%. In the best performing variant this is around 1%, which seems to be a very low number and one might wonder whether this is enough to have any effect at all. This can be answered by analyzing the percentage of decisions on extended variables. For all benchmarks where DIP-based systems outperform the baseline CDCL solver, this number is surprisingly large, being usually over 10% of the decisions, whereas for most of the benchmarks is very low (in 70% of the benchmarks less than 1% of decisions are on extended variables). Moreover, this does not only happen in problems with a very low number of initial variables, where the introduction of a few extended variables could quickly dominate the decision heuristic.

This is remarkable since it shows that already existing decision heuristics like VSIDS or LRB somehow infer whether the newly introduced variables improve the behavior of the system. This is why we implemented, on top of our baseline DIP system, a procedure that computes the percentage of decisions on extended variables. If after a given number of conflicts it is still lower than 3%, it discontinues DIP learning and performs 1-UIP learning from that moment onwards. This is a very preliminary step in the direction of trying to automate the decision of whether to apply DIP reasoning or not, but the outcome, which is found in the right plot of Figure 8, is very positive.

7 Conclusions and Future Work

We introduce a novel extended resolution clause learning (ERCL) algorithm, that when implemented on top of a CDCL solver, turns out to be beneficial for a variety of problems, in particular Tseitin, random k-xor and interval matching formulas. This is remarkable since automating powerful proof systems is well-known to be a difficult task. Somewhat surprisingly, the only previously existing attempt to incorporate ER into CDCL performs similarly on the same instances. Considering the different nature of the two methods, this clearly deserves further study. Further, we also introduce a new heuristic that allows our ERCL solver xMapleSAT to perform similarly to the baseline CDCL solver MapleLCM, thus being able to get the best of both worlds, i.e., the benefit of ERCL, without sacrificing performance of the CDCL solver on, say, SAT Competition Main Track 2023 instances. As future work, we plan to investigate a variety of machine learning based heuristics (e.g., branching) specialized for the ERCL method. Also, the use of k-IPs with $k > 2$ is part of our next steps. On the theoretical side, challenging questions like determining whether DIP- or k-IP based ER simulates unrestricted ER are going to be central to our research efforts.

References

1 Albert Atserias, Johannes Klaus Fichte, and Marc Thurley. Clause-learning algorithms with many restarts and bounded-width resolution. *Journal of Artificial Intelligence Research*, 40:353–373, January 2011. Preliminary version in *SAT '09*.

2 Albert Atserias, Johannes Klaus Fichte, and Marc Thurley. Clause-learning algorithms with many restarts and bounded-width resolution. *J. Artif. Intell. Res.*, 40:353–373, 2011. doi:10.1613/jair.3152.

3 Gilles Audemard, George Katsirelos, and Laurent Simon. A restriction of extended resolution for clause learning sat solvers. In *AAAI*, 2010.

4 Gilles Audemard and Laurent Simon. Predicting learnt clauses quality in modern SAT solvers. In Craig Boutilier, editor, *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009*, pages 399–404, 2009. URL: http://ijcai.org/Proceedings/09/Papers/074.pdf.

5 Fahiem Bacchus. Enhancing Davis Putnam with extended binary clause learning. In *Proc. 18th Annual Conference on Artificial Intelligence and 15th Conference on Innovative Applications of Artificial Intelligence*, pages 613–619. AAAI Press and MIT Press, 2002.

6 Tomas Balyo, Marijn Heule, Markus Iser, Matti Järvisalo, and Martin Suda, editors. *Proceedings of SAT Competition 2023: Solver, Benchmark and Proof Checker Descriptions*. Department of Computer Science Series of Publications B. Department of Computer Science, University of Helsinki, Finland, 2023.

7 Eli Ben-Sasson. Size space tradeoffs for resolution. *SIAM Journal on Computing*, 38(6):2511–2525, 2009.

8 Armin Biere, Katalin Fazekas, Mathias Fleury, and Maximillian Heisinger. CaDiCaL, Kissat, Paracooba, Plingeling and Treengeling entering the SAT Competition 2020. In Tomas Balyo, Nils Froleyks, Marijn Heule, Markus Iser, Matti Järvisalo, and Martin Suda, editors, *Proc. of SAT Competition 2020 – Solver and Benchmark Descriptions*, volume B-2020-1 of *Department of Computer Science Report Series B*, pages 51–53. University of Helsinki, 2020.

9 Armin Biere, Nils Froleyks, and Wenzi Wang. CadiBack: Extracting backbones with CaDiCal. In *Proc. 26th Intl. Conf. on Theory and Applications of Satisfiability Testing SAT*, pages 3:1–12. Leibniz-Zentrum für Informatik, 2023.

10 Randal E. Bryant and Marijn J. H. Heule. Generating extended resolution proofs with a bdd-based SAT solver. *ACM Trans. Comput. Log.*, 24(4):31:1–31:28, 2023. doi:10.1145/3595295.

11 Sam Buss, Vijay Ganesh, and Albert Oliveras. A linear time algorithm for two-vertex bottlenecks. In preparation. Preliminary version available at https://math.ucsd.edu/~sbuss/ResearchWeb/TwoVertBottlenecks, March 2024.

12 Sam Buss, Oliver Kullmann, and Virginia Vassilevska Williams. Dual depth first search for binary clause reasoning. Preliminary version, March 2024.

13 Stephen A. Cook. A short proof of the pigeon hole principle using extended resolution. *SIGACT News*, 8(4):28–32, 1976. doi:10.1145/1008335.1008338.

14 Andrew Haberlandt, Harrison Green, and Marijn J. H. Heule. Effective auxiliary variables via structured reencoding. In Meena Mahajan and Friedrich Slivovsky, editors, *26th International Conference on Theory and Applications of Satisfiability Testing, SAT 2023, July 4-8, 2023, Alghero, Italy*, volume 271 of *LIPIcs*, pages 11:1–11:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023. URL: https://doi.org/10.4230/LIPIcs.SAT.2023.11, doi:10.4230/LIPICS.SAT.2023.11.

15 Armin Haken. The intractability of resolution. *Theoretical computer science*, 39:297–308, 1985.

16 Marijn J. H. Heule, Benjamin Kiesl, and Armin Biere. Short proofs without new variables. In Leonardo de Moura, editor, *Automated Deduction - CADE 26 - 26th International Conference on Automated Deduction, Gothenburg, Sweden, August 6-11, 2017, Proceedings*, volume 10395 of *Lecture Notes in Computer Science*, pages 130–147. Springer, 2017. doi:10.1007/978-3-319-63046-5_9.

17 Marijn J. H. Heule, Benjamin Kiesl, and Armin Biere. Encoding redundancy for satisfaction-driven clause learning. In *Tools and Algorithms for the Construction and Analysis of Systems - 25th International Conference (TACAS 2019), Part I*, Lecture Notes in Computer Science 11427, pages 41–58. Springer Verlag, 2019.

18 Marijn J. H. Heule, Benjamin Kiesl, and Armin Biere. Strong extension-free proof systems. *J. Autom. Reason.*, 64(3):533–554, 2020. `doi:10.1007/s10817-019-09516-0`.

19 Marijn J. H. Heule, Benjamin Kiesl, Martina Seidl, and Armin Biere. Pruning through satisfaction. In Ofer Strichman and Rachel Tzoref-Brill, editors, *Hardware and Software: Verification and Testing - 13th International Haifa Verification Conference, HVC 2017, Haifa, Israel, November 13-15, 2017, Proceedings*, volume 10629 of *Lecture Notes in Computer Science*, pages 179–194. Springer, 2017. `doi:10.1007/978-3-319-70389-3\_12`.

20 Markus Iser and Carsten Sinz. A problem meta-data library for research in sat. In Daniel Le Berre and Matti J\"arvisalo, editors, *Proceedings of Pragmatics of SAT 2015 and 2018*, volume 59 of *EPiC Series in Computing*, pages 144–152. EasyChair, 2019. URL: `https://easychair.org/publications/paper/jQXv`, `doi:10.29007/gdbb`.

21 Henry A. Kautz. Deconstructing planning as satisfiability. In *Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA*, pages 1524–1526. AAAI Press, 2006. URL: `http://www.aaai.org/Library/AAAI/2006/aaai06-241.php`.

22 Henry A. Kautz and Bart Selman. Planning as satisfiability. In Bernd Neumann, editor, *10th European Conference on Artificial Intelligence, ECAI 92, Vienna, Austria, August 3-7, 1992. Proceedings*, pages 359–363. John Wiley and Sons, 1992.

23 Daniel Kroening. Software verification. In Armin Biere, Marijn Heule, Hans van Maaren, and Toby Walsh, editors, *Handbook of Satisfiability - Second Edition*, volume 336 of *Frontiers in Artificial Intelligence and Applications*, pages 791–818. IOS Press, 2021. `doi:10.3233/FAIA201004`.

24 Massimo Lauria, Jan Elffers, Jakob Nordström, and Marc Vinyals. Cnfgen: A generator of crafted benchmarks. In Serge Gaspers and Toby Walsh, editors, *Theory and Applications of Satisfiability Testing - SAT 2017 - 20th International Conference, Melbourne, VIC, Australia, August 28 - September 1, 2017, Proceedings*, volume 10491 of *Lecture Notes in Computer Science*, pages 464–473. Springer, 2017. `doi:10.1007/978-3-319-66263-3\_30`.

25 Mao Luo, Chu-Min Li, Fan Xiao, Felip Manyà, and Zhipeng Lü. An effective learnt clause minimization approach for CDCL SAT solvers. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 703–711. ijcai.org, 2017. URL: `https://doi.org/10.24963/ijcai.2017/98`, `doi:10.24963/IJCAI.2017/98`.

26 Norbert Manthey, Marijn Heule, and Armin Biere. Automated reencoding of boolean formulas. In Armin Biere, Amir Nahir, and Tanja E. J. Vos, editors, *Hardware and Software: Verification and Testing - 8th International Haifa Verification Conference, HVC 2012, Haifa, Israel, November 6-8, 2012. Revised Selected Papers*, volume 7857 of *Lecture Notes in Computer Science*, pages 102–117. Springer, 2012. `doi:10.1007/978-3-642-39611-3\_14`.

27 João Marques-Silva, Inês Lynce, and Sharad Malik. Conflict-driven clause learning SAT solvers. In Armin Biere, Marijn Heule, Hans van Maaren, and Toby Walsh, editors, *Handbook of Satisfiability - Second Edition*, volume 336 of *Frontiers in Artificial Intelligence and Applications*, pages 133–182. IOS Press, 2021. `doi:10.3233/FAIA200987`.

28 Karl Menger. Zur algemeinen Kurventheorie. *Fundamenta Mathematicae*, 10:96–115, 1927.

29 Sibyle Möhle and Armin Biere. Backing backtracking. In *22nd Intl Conf. on Theory and Applications of Satisfiability Testing (SAT)*, Lecture Notes in Computer Science 11628, pages 250–266. Springer, 2019.

30 Alexander Nadel and Vadim Ryvchin. Chronological backtracking. In *21nd Intl Conf. on Theory and Applications of Satisfiability Testing (SAT)*, Lecture Notes in Computer Science 10929, pages 111–121. Springer, 2018.

31 Albert Oliveras, Chunxiao Li, Darryl Wu, Jonathan Chung, and Vijay Ganesh. Learning shorter redundant clauses in SDCL using maxsat. In Meena Mahajan and Friedrich Slivovsky, editors, *26th International Conference on Theory and Applications of Satisfiability Testing, SAT 2023, July 4-8, 2023, Alghero, Italy*, volume 271 of *LIPIcs*, pages 18:1–18:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023. URL: https://doi.org/10.4230/LIPIcs.SAT.2023.18, doi:10.4230/LIPICS.SAT.2023.18.

32 Knot Pipatsrisawat and Adnan Darwiche. On modern clause-learning satisfiability solvers. *Journal of Automated Reasoning*, 44(3):277–301, 2010.

33 Knot Pipatsrisawat and Adnan Darwiche. On the power of clause-learning SAT solvers as resolution engines. *Artif. Intell.*, 175(2):512–525, 2011. doi:10.1016/j.artint.2010.10.002.

34 Joseph E. Reeves, Marijn J. H. Heule, and Randal E. Bryant. Preprocessing of propagation redundant clauses. In Jasmin Blanchette, Laura Kovács, and Dirk Pattinson, editors, *Automated Reasoning - 11th International Joint Conference, IJCAR 2022, Haifa, Israel, August 8-10, 2022, Proceedings*, volume 13385 of *Lecture Notes in Computer Science*, pages 106–124. Springer, 2022. doi:10.1007/978-3-031-10769-6_8.

35 Karem A. Sakallah. Symmetry and satisfiability. In Armin Biere, Marijn Heule, Hans van Maaren, and Toby Walsh, editors, *Handbook of Satisfiability - Second Edition*, volume 336 of *Frontiers in Artificial Intelligence and Applications*, pages 509–570. IOS Press, 2021. doi:10.3233/FAIA200996.

36 Mate Soos, Karsten Nohl, and Claude Castelluccia. Extending SAT solvers to cryptographic problems. In Oliver Kullmann, editor, *Theory and Applications of Satisfiability Testing - SAT 2009, 12th International Conference, SAT 2009, Swansea, UK, June 30 - July 3, 2009. Proceedings*, volume 5584 of *Lecture Notes in Computer Science*, pages 244–257. Springer, 2009. doi:10.1007/978-3-642-02777-2_24.

37 Grigori S Tseitin. On the complexity of derivation in propositional calculus. *Automation of reasoning: 2: Classical papers on computational logic 1967–1970*, pages 466–483, 1983.

38 Alasdair Urquhart. Hard examples for resolution. *J. ACM*, 34(1):209–219, 1987. doi:10.1145/7531.8928.

39 Yichen Xie and Alexander Aiken. Saturn: A SAT-Based Tool for Bug Detection. In *Proceedings of the 17th International Conference on Computer Aided Verification, CAV 2005*, pages 139–143, 2005. doi:10.1007/11513988_13.

40 Lintao Zhang, Conor F. Madigan, Matthew W. Moskewicz, and Sharad Malik. Efficient conflict driven learning in boolean satisfiability solver. In Rolf Ernst, editor, *Proceedings of the 2001 IEEE/ACM International Conference on Computer-Aided Design, ICCAD 2001, San Jose, CA, USA, November 4-8, 2001*, pages 279–285. IEEE Computer Society, 2001. doi:10.1109/ICCAD.2001.968634.