Journal on Satisfiability, Boolean Modeling and Computation 6 (2008) 33-51

HaifaSat: a SAT solver based on an Abstraction/Refinement model

Roman Gershman gershman@cs.technion.ac.il
Computer Science,
Technion, Haifa,
Israel
Ofer Strichman ofers@ie.technion.ac.il
Information Systems Engineering, IE,
Technion, Haifa,
Israel

Abstract

The popular abstraction/refinement model frequently used in verification, can also explain the success of a SAT decision heuristic like Berkmin. According to this model, conflict clauses are abstractions of the clauses from which they were derived. We suggest a clause-based decision heuristic called Clause-Move-To-Front (CMTF), which attempts to follow an abstraction/refinement strategy (based on the resolve-graph) rather than satisfying the clauses in the chronological order in which they were created, as done in Berkmin. We also show a resolution-based score function for choosing the variable from the selected clause and a similar function for choosing the sign. We implemented the suggested heuristics in our SAT solver HaifaSat. Experiments on hundreds of industrial benchmarks demonstrate the superiority of this method comparing to the Berkmin heuristic. HaifaSat won the 3rd place in the industrial-benchmarks category in the SAT competition of 2005, and did not compete or was developed since. We present experimental results with the benchmarks of the 2007 competition that show that it is about 32% slower than RSAT, the winner of 2007. Considering the time difference, it shows that it is rather robust. The abstraction/refinement theoretical model is still relevant, and there is still room for further research on how to exploit it better given a recent result that permits storing and manipulating the resolve graph in the main memory.

Keywords: *SAT-solver, abstraction*

Submitted January 2008; revised June 2008; published October 2008

1. Introduction

HaifaSat is a SAT solver that was developed during 2004-2005, and won the third place in the industrial benchmarks category of the SAT competition 2005.[1] Although it was not developed since, it is still being used and downloaded rather regularly (241 downloads since March 2005, from which 45 in 2007). The current paper extends an earlier proceedings version [8] that describes some of HaifaSat's features, mainly by presenting its results on the 2007 competition benchmarks. Although it is, as expected, not competitive with the

1. The first and second place winners, SateliteGTI and MiniSAT, are variations on the same code-base.

winner of 2007 (Rsat) [16], we find the theoretical model that it describes still useful in thinking about decision heuristics, and it still leaves room for future research, especially in light of some recent advances that permits maintaining and manipulating the resolve graph in the main memory [18]. A clause-based decision heuristic very similar to the one suggested in this paper was suggested independently by Dershowitz, Hanna and Nadel [6] with very similar conclusions, although without the abstraction/refinement model that we describe here. Their SAT solver, Eureka, which won the second place in the SAT 2006 competition, is based on this heuristic as well.

By now the view by which a SAT solver should not only be seen as a search engine based on *enumeration*, but rather also as a *proof engine* based on resolution, is prevalent. Traditionally the first view was dominant, hence the emphasis in designing SAT solvers and explaining their success was on pruning search spaces. Decision heuristics and learning schemes can all be interpreted as aiming at this goal. Yet the harder and larger the CNF instances are, pruning alone cannot account for the success of modern SAT solvers. It is their ability as proof engines that makes them succeed. This distinction has practical implications, too. For example, for many years decision heuristics gave higher priority to variables in shorter clauses, and to learning shorter conflict clauses. The reasoning was that such clauses can potentially prune larger search-spaces. Although this claim is true, all competitive decision heuristics ignore the length of the clauses, after reaching empirically the conclusion that there are more important considerations. Ryan experimented in his thesis [17] with first-UIP [20] and all-UIP [13] learning schemes, and although the latter generate on average shorter clauses, the former is empirically better. He hypothesized that the learning scheme should be geared towards resolution rather than for pruning. In this article we extend this approach by looking at clause-learning and the decision heuristic as one complete mechanism and refer to a SAT solver as a prover rather than as a search engine. It turns out, empirically, that when conflict clauses are effective, which is the case in all real-world instances, this is the right way to go.

Conflict clauses are derived through a process of resolution (see, for example, [20] and [2] for a more formal treatment of this subject). If a clause c is derived by resolution from a set of clauses $c_1 \dots c_n$ then

$$c_1 \wedge \dots \wedge c_n \rightarrow c$$

while the other direction does not hold. This means that c can be seen as an over-approximating abstraction of the resolving clauses $c_1 \dots c_n$. Attempting to satisfy c first, therefore, can be seen as an attempt to satisfy the abstract model first. Like any abstraction/refinement technique (also called *localization* techniques) [12, 5, 4, 9, 11, 1], a successful assignment to c is one that satisfies the concrete model (the $c_1 \dots c_n$ clauses) as well. And an unsuccessful assignment leads to a refinement step, or, in our case, to derivation of new conflict clauses which further constrain the abstract model. According to this model, Berkmin is only one of many possible strategies to refine the abstract model. In Sect. 3 we suggest one such alternative clause-based decision heuristic called Clause-Move-To-Front (CMTF), which attempts to follow the order of the clauses in the *resolve-graph* [21] rather than their chronological order in which they were created. In Sect. 4 we also show a resolution-based score function for choosing the variable from the selected clause and a similar function for choosing the sign. In Sect. 5 we report experimental results on hundreds of industrial benchmarks that prove the advantage of our approach.

Related work.

Decision heuristics are probably the most important and most studied aspect of SAT solving. Let us mention some prominent such heuristics from the last decade. GRASP [13] included several decision heuristics but used as default DLIS (Dynamic Largest Individual Sum), which at each decision point counts the number of unsatisfied clauses in which each literal appears, and decides on the literal with the highest count. This heuristic is computationally expensive as it requires traversing the entire clause database at each decision point. Chaff [15] introduced the VSIDS (Variable State Independent Decaying Sum) decision heuristics, which differ from DLIS in two aspects. First, VSIDS counts the number of clauses in which each literal occurs, but unlike DLIS it ignores the question whether these clauses are currently satisfied. Hence, the criterion less accurately predicts the immediate impact of the decision, but on the other hand is cheaper to compute: rather than traversing all clauses, it only requires updating counters once new clauses are added. Second, VSIDS periodically divides the literal counters by a constant number. This gives higher priority to literals that occurred in recently learned conflict clauses, since their contribution to the counter is divided less times. VSIDS was the first *conflict-driven* decision heuristic, in the sense that it was the first to attempt to focus the search near recent conflicts. The best solver to date, Minisat [7], uses a similar idea. Berkmin [10], which we describe in detail in Sect. 2.2, is even more extreme than VSIDS in focusing the search on conflicts: it adds the conflict clauses to a stack, and at decision points it chooses a variable from the most recent unsatisfied conflict clause. VMTF (Variable Move To Front) [17] is another very successful conflict-driven decision heuristic, which was implemented in the SAT solver Siege: when learning a new conflict clause, it pushes to a stack some constant number of literals from that clause. The top literal in this stack that does not yet have a value is the decision literal.

We mentioned earlier several references relating to abstraction refinement in model checking, some of which are based on SAT (e.g., [1]), although the abstraction-refinement loop is external to the SAT engine. An exception, perhaps, is [11], where abstraction-refinement was used to accelerate SAT solving: starting from a small subset of the clauses in the original formula, in each iteration more clauses from the original formula were added so as to block the current satisfying assignment. The context in that work was Bounded Model Checking [3] formulas, and correspondingly the abstraction refinement loop followed the traditional way of refining the model: the initial set of clauses correspond to the property, and in each iteration another 'layer' of state variables are added (i.e., the clauses that define the behavior of these state variables). The current work is more generic and is based on changing the decision heuristic itself. The abstraction is not due to the removal of clauses, rather due to resolution, as we will explain later on.

2. Background

The explanation of our methods and the analysis of various heuristics later on will require some basic definitions.

The Abstraction-refinement model: from structures to formulas

The classic use of the terms abstraction and refinement in the context of model-checking is the following. Let M be a Kripke structure, $P(M)$ the set of propositions labeling its states and $\mathcal{L}(M)$ the language defined by M (i.e., the set of traces in M). A model $\hat{M}$ such that $P(\hat{M}) \subseteq P(M)$ is an over-approximating abstraction of M if for every property φ it holds that

$$\hat{M} \models \varphi \rightarrow M \models \varphi. \tag{1}$$

Equivalently, for every trace s,

$$s \in \mathcal{L}(M) \rightarrow s \in \mathcal{L}(\hat{M}). \tag{2}$$

The inclusion relation is defined with respect to the alphabet of the language, e.g., $s \in \mathcal{L}(M)$ is defined with respect to the projection of s to $P(M)$.

M_1 is a *refinement* of $\hat{M}$ with respect to M, if for every trace s,

$$s \in \mathcal{L}(M) \rightarrow s \in \mathcal{L}(M_1), \tag{3}$$

and

$$s \in \mathcal{L}(M_1) \rightarrow s \in \mathcal{L}(\hat{M}). \tag{4}$$

Abstraction-Refinement is a process in which we find increasingly accurate models (closer to the concrete model M) until proving the property or, in the worst case, reaching the original model M.

We now wish to bridge between the terminology of models and traces on one hand, and the terminology of formulas and satisfying assignments on the other hand. Thus, consider now formulas rather than models.

For two formulas f and $\hat{f}$ such that $var(\hat{f}) \subseteq var(f)$, we can restate an implication of the form

$$f \rightarrow \hat{f}, \tag{5}$$

by saying that for every assignment α,

$$\alpha \models f \rightarrow \alpha \models \hat{f}. \tag{6}$$

As usual satisfaction is defined with respect to a projection of α to the variables of the formula.

Due to the resemblance to (2), we now say that $\hat{f}$ is a conservative abstraction (over-approximation) of f.

Further, for a formula f_1 such that $var(\hat{f}) \subseteq var(f_1) \subseteq var(f)$, we can restate

$$f \rightarrow f_1 \tag{7}$$

and

$$f_1 \rightarrow \hat{f} \tag{8}$$

by saying that for every assignment α,

$$\alpha \models f \rightarrow \alpha \models f_1, \tag{9}$$

and

$$\alpha \models f_1 \rightarrow \alpha \models \hat{f}. \tag{10}$$

Once again, due to the resemblance of (3) and (4) to (9) and (10), respectively, we now say that f_1 refines $\hat{f}$ with respect to f.

Continuing with this terminology, abstraction-refinement for formulas is an iterative process, in which one begins with some abstract formula $\hat{f}$ of a concrete formula f and gradually refines it through a series of formulas $\hat{f}_1, \ldots, \hat{f}_n$ until proving or disproving the desired property of f. Here again, in the worst case $\hat{f}_n = f$. Thus, there is a parallelism between abstraction refinement of structures, and the process described here for formulas.

2.1 Conflict clauses and resolution

The binary resolution inference rule is:

$$\frac{a_1 \vee \ldots \vee a_n \vee \beta \qquad b_1 \vee \ldots \vee b_m \vee (\neg\beta)}{a_1 \vee \ldots \vee a_n \vee b_1 \vee \ldots \vee b_m} \quad \text{(Bin-Res)}$$

where $a_1, \ldots a_n, b_1, \ldots b_m, \beta$ are literals. β is known as the *resolution variable* (also known as the *pivot variable*) of this derivation. Clauses $(a_1, \ldots, a_n, \beta)$ and $(b_1, \ldots, b_n, \overline{\beta})$ are called *resolving clauses* and the clause $(a_1, \ldots, a_n, b_1, \ldots b_n)$ is a *resolvent*. It follows by the soundness of the rule, that the resolvent is always implied by its resolving clauses and can therefore be thought of as an abstraction of the clauses that participated in the derivation.

We now show why the process of generating conflict clauses indeed can be seen as a sequence of resolution steps. Algorithm 1 shows a simple and efficient implementation of the First-UIP resolution scheme, which is implemented in most competitive SAT solvers, including our solver HaifaSat. We will refer to this algorithm simply as the *resolution* algorithm. First, a conflicting clause is set to be the current resolved clause. The main loop processes literals in the current clause. All literals from the previous decision levels are gathered into $NewClause$ at line 13 and marked. Literals from the current level are marked in order to resolve on them (i.e., use them as resolution variables) further. In every iteration a new marked (yet unprocessed) literal u is chosen in line 17. This literal must be from the current decision level. The algorithm resolves on u by setting $currentClause$ to be the antecedent clause without u.

$ResolveNum$ counts the number of the marked literals from the current decision level that still have to be processed. When $ResolveNum = 0$ at line 22, then u is the *FirstUIP* or the *asserted* literal. The negation of this literal is added to the $NewClause$ causing u's value to be flipped after backtracking. For more details on the resolution algorithm see [15, 17].

We will use the following definition in order to denote the initial state of $NewClause$:

Definition 1 (Asserting clause)**.** *Suppose a new conflict clause C was created in Alg. 1 with asserted literal u. Suppose also that the solver backtracks after the conflict to level dl. Then C becomes an* asserting *clause when it implies $\overline{u}$ for the first time at level dl, and stops being asserting when the solver backtracks from dl.*

It follows from the definition that every conflict clause becomes asserting exactly once.

Algorithm 1 The First-UIP resolution algorithm

```
    procedure AnalyzeConflict(Clause: conflict)
 2:     currentClause ← conflict;
        ResolveNum ← 0;
 4:     NewClause ← ∅;
        repeat
 6:         for each literal lit ∈ currentClause do
               v ← var(lit);
 8:            if v is not marked then
                   Mark v;
10:                if dlevel(v) = CurrentLevel then
                       ++ ResolveNum;
12:                else
                       NewClause ← NewClause ∪ {lit};
14:                end if
               end if
16:        end for
           u ← last marked literal on the assignment stack;
18:        Unmark var(u);
           −− ResolveNum;
20:        ResolveCl ← Antecedent(u);
           currentClause ← ResolveCl \ {u};
22:    until ResolveNum = 0;
       Unmark literals in NewClause;
24:    NewClause ← NewClause ∪ {ū};
       Add NewClause to the clause database;
26: end procedure
```

Example 1. *Consider the following partial implication graph [14] and set of clauses.*

Denote by $Resolve(s, t, x)$ *the binary resolution of clauses* s *and* t *with the resolution variable* x*. Then the conflict clause* $c_5 : (x_{10}, x_2, \neg x_4)$ *is computed through a series of binary resolutions, starting from the conflicting clause* c_4*, and going backwards on the implication graph until all literals in the conflict clause are either from previous decision levels or the* $firstUIP$*.*

$$Resolve(Resolve(Resolve(c_4, c_3, x_7)), c_2, x_6), c_1, x_5) = (x_{10}, x_2, \neg x_4)$$

Algorithm 1 implicitly performs these resolution steps while computing the conflict clause c_5*.*

NewClause is derived through a series of binary resolutions that can be seen as a tree: every time the solver reaches line 21, an intermediate clause (consisting of all marked literals) is resolved with the antecedent clause of the chosen resolution variable. We can treat this process as one atomic action of *Hyper-resolution* (resolution between more than two clauses). Since each conflict clause is derived from a set of other clauses, we can keep track

$c_1 = (\neg x_4 \vee x_2 \vee x_5)$
$c_2 = (\neg x_4 \vee x_{10} \vee x_6)$
$c_3 = (\neg x_5 \vee \neg x_6 \vee \neg x_7)$
$c_4 = (\neg x_6 \vee x_7)$

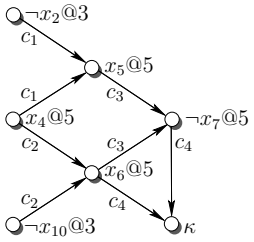

Figure 1. A partial implication graph and set of clauses demonstrate AnalyzeConflict. x_4 is the $FirstUIP$, and $\overline{x}_4$ is the asserted literal.

of this process with a *Resolve-Graph* [21]. Here we define a variation of the resolve-graph that distinguished between two types of resolutions:

Definition 2 (Colored Resolve Graph)**.** *A* Resolve Graph *is a DAG* $G(V, E_{as}, E_{nas})$*. V is the set of nodes, each of which corresponds to a clause; Both E_{as} and E_{nas} correspond to edges such that $(u, v) \in E_{as} \cup E_{nas}$ if and only if v participated in the Hyper-resolution of u. The distinction between these two nonintersecting sets is the following: For every edge $(u, v) \in E_{as} \cup E_{nas}$, it holds that $(u, v) \in E_{as}$ if v was an asserting clause during the resolution and $(u, v) \in E_{nas}$ otherwise.*

We use solid edges to denote elements of E_{as} edges, and dashed edges to denote elements of E_{nas}.

In (colored) resolve graphs, edges come from the resolvent to its resolving clauses, and hence the leafs of the graph correspond to the original clauses in the formula.[2.] Notice that since a conflict at level dl necessarily implies that the solver backtracks from dl and unassigns all the variables that were resolved on, any asserting clause which participated in the resolution will stop being asserting. Therefore for any conflict clause there can be at most one incoming solid edge. The original clauses do not have outgoing edges, and only dashed incoming edges.

Example 2. *Consider once again the implication graph in Fig. 1. Assuming that $c_1 \ldots c_4$ are original clauses, they are not asserting clauses at the time of resolving c_5 (or at any other time). The corresponding resolve-graph is thus as appears in Fig. 2(a).*

Now consider a similar case in which c_2 is not an original clause, and at the time when $x_4@5$ is asserted it does not yet exist (the notation $l@i$, adopted from [13, 14], means that literal l is asserted at decision level i). The implication graph at this stage appears in Fig. 3. Now assume that due to further decisions and implications in deeper decision levels a conflict is encountered, the solver creates the new conflict clause c_2, backtracks to decision level 5 and asserts $x_6@5$. This, in turn, completes the implication graph to the way it looks in Fig. 1. But now, since c_2 asserts x_6, its edge on the resolve-graph from c_5 belongs to E_{as}. Fig. 2(b) presents the corresponding colored resolve-graph.

2. This is somewhat counterintuitive because it goes in the reverse order of inference, but it is a common convention because it reflects the data structure maintained by most SAT solvers, which is necessary for traversing this graph.

Figure 2. Two (colored) resolve-graphs corresponding to the implication graph in Fig. 1. The left drawing corresponds to a case in which clauses $c_1, \ldots, c_4$ are original clauses and hence were not asserting at the time of resolving c_5; The right drawing corresponds to a case in which c_2 is a conflict clause that was asserting at the time of resolving c_5.

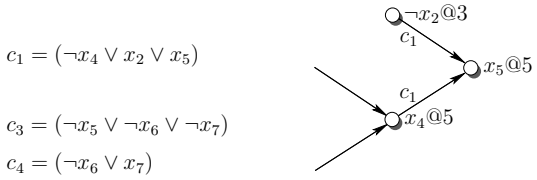

Figure 3. A partial implication graph corresponding to c_1, c_3, c_4 and the decision $x_4@5$.

The distinction between the two types of edges is important because a solid edge (u, v) indicates that the solver had to create v in order to later create u[3.]. We will use this distinction later on, when describing the scoring heuristic in Sect. 4.

2.2 The Berkmin Decision heuristic

We describe briefly the Berkmin Decision heuristic. Berkmin is a clause-based decision heuristic, like HaifaSat's heuristic, and therefore convenient for comparison. It is the decision heuristic of the Berkmin SAT solver and of Forklift, which won the 2003 competition [19] in the industrial category.

Berkmin [10] pushes every new conflict clause into a stack, and makes a decision by choosing an unassigned variable from the last unresolved conflict clause in this stack. It uses the VSIDS score system [15] to choose among such variables, and a similar scoring mechanism to choose its value (VSIDS counts the number of times each literal appears, and periodically divides these numbers so as to give priority to literals that occurred in recently added conflict clauses. The unassigned literal with the highest score is the next decision variable). If all the conflict clauses are satisfied, it uses the same scoring mechanisms to choose among all unassigned variables.

In Fig. 4(a) we show a sketch of the progress of Berkmin, which is helpful in understanding why this process can be seen as abstraction-refinement. Clauses $c_1, \ldots, c_{100}$ are conflict

3. By this we do not mean that this is the only way to create u.

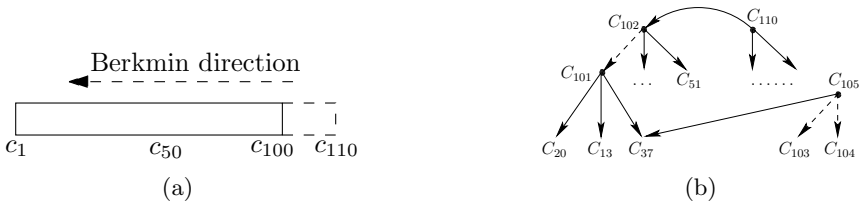

Figure 4. Berkmin's decision heuristic can be thought of as an abstraction-refinement, where a range of the conflict clauses from the right end until c_i represents an abstract model of the clauses on the left of c_i. (a) Berkmin clauses stack: after encountering a conflict, the new resolved clauses are added on the right end. By the time the solver returns to c_{50}, it will have a partial assignment that satisfies a refined model, i.e., the clauses $c_{51} \ldots c_{110}$ (b) The resolve sub-graph of some newly created clauses. Dashed edges are defined in Definition 2.

clauses ordered by their creation time (c_1 is first). Berkmin tries to satisfy these clauses from last to first, i.e., from right to left. Suppose that the clauses $c_{51} \ldots c_{100}$ are already satisfied, and now Berkmin focuses on c_{50}. We refer to $S = \{c_{51}, \ldots, c_{100}\}$ as our current abstract formula of the original formula φ (it is abstract because each of the clauses in S is derived by a resolution chain from the clauses of φ). Clauses in S must be currently satisfied, since the decision heuristic reached c_{50}. Berkmin now makes a decision on a variable from c_{50} which leads to a conflict and learning of a new clause. The decision heuristic backtracks to the clauses on the end of the list, until finally, through possibly additional iterations of conflicts and added clauses, it reaches c_{50} again while all the clauses to its right are satisfied. Denote by S' the clauses to the right of c_{50} at this point, e.g., $S' = \{c_{51} \ldots, c_{110}\}$. Clearly $S \subseteq S'$ and S' is an abstraction of φ. We can therefore say that S' is a refinement of S with respect to φ.

This view of the process possibly explains why a strategy of giving absolute priority to variables in a specific clause is empirically better than previous approaches like VSIDS that used only a score function.

Fig. 4(a) shows a 'linear' view of the conflict clauses in the order that they are added. This is also the order by which Berkmin considers them. Berkmin never tries to satisfy a clause before satisfying its resolvents and thus mimics a gradual process of refinement.

A different view of conflict clauses considers their partial order in the Resolve Graph. Fig.4(b) presents a possible Resolve sub-Graph corresponding to the same set of clauses. After the conflict, Berkmin starts from satisfying c_{110}. c_{102} is a resolving clause that can potentially refine the initial model, however Berkmin first passes through $c_{105}, c_{104}, c_{103}$ to which c_{110} is not connected at all. Therefore Berkmin is dispersed trying to refine several abstractions. Such unfocused behavior can lead to longer proofs. This problem is exactly what our decision heuristic CMTF attempts to solve, as we soon show.

Our SAT solver HaifaSat makes a decision in three steps: First, it chooses an unsatisfied clause according to the CMTF heuristic; Second, it chooses an unassigned variables from this clause, and, finally, it gives this variable a value. The next sections describe in detail these decision steps.

3. The Clause-Move-To-Front (CMTF) decision heuristic

The description above of Berkmin's decision heuristic, and the alternative view of the conflict clauses as being part of a resolve-graph, hints towards the process which is described in Alg. 2. In the first line $roots(ResolveGraph)$ refer to resolvent clauses that did not resolve other clauses. Note that in this general scheme a clause is processed only if at least one of its abstractions (its resolvent clauses) has already been processed. It is easy to see that Berkmin is an instantiation of this scheme. In fact, Berkmin is more strict and processes a clause only if *all* its abstractions are satisfied.

Algorithm 2 A Resolve-Graph Based decision heuristic

1: $S = roots(ResolveGraph)$;
2: **while** there exists $v \in S$ an unsatisfied clause (node) **do**
3: Process v; ▷ Processing a clause, among other things, satisfies it.
4: $S = S \cup children(v)$;
5: **end while**

CMTF instantiates this scheme in a different way. It causes the decision heuristic to be more focused on the current refinement path, i.e., to satisfy children of the currently satisfied clause s. It works as follows:

- All the conflict clauses are stored in a list.
- During the resolution in Alg 1, a bounded number of resolving conflict clauses which are processed at line 6 are moved to the front (front corresponds to the right end of Fig. 4(a)). The newly created clause $NewClause$ is also added to the list (can be done at line 25).
- Clauses are processed from right to left in the list, while ignoring satisfied clauses. If all the conflict clauses are satisfied then the solver reverts to some other heuristic (HaifaSat uses the VMTF strategy [17] in this case, a strategy by which a user-defined number of *variables* from the generated conflict clause are brought to the beginning of the list. Their value is determined by VSIDS.).

The idea of this strategy is to keep clauses that participate in resolution adjacent to their resolvents (at least until the next time they participate in a resolution, a case in which they can be moved to a new location).

CMTF shows an improvement on many industrial problems comparing to the Berkmin heuristic. Both are specific instantiation of the scheme showed above. The advantages of CMTF is its simplicity and the fact that the explicit storage of the resolve-graph is not required. However, it seems that there is still room for future research on how to use the general scheme. For example, classic AI search methods like best-first-search can be used to decide on the exploration order of nodes in S at line 2. It may happen that partial or full storage of the resolve-graph will improve the performance.

4. Resolution-based scoring

In the previous section we showed how HaifaSat decides which clause to satisfy first. Given a clause c there can still be several ways to satisfy it. HaifaSat computes dynamically an *activity score* for each variable and then chooses the variable with the maximal score. Then another *sign score* is used to determine its Boolean value. This scoring mechanism, which is simple to implement (only a few lines added to Algorithm 1) but somewhat hard to explain, is the subject of this section.

The idea, intuitively, is to give higher weights to variables that were frequently resolved on recently, while distinguishing between resolutions that were necessary for the progress of the solver, and those that were made due to the imperfection of the decision heuristic. We will need several definitions and lemmas to explain this heuristic more precisely.

Suppose that every time the solver makes a decision or processes a conflict it writes into a log the event $a_i = (dl, e)$ where dl is the decision level where the event occurred and $e \in Conflicts \cup Decisions$ is either a conflict event or a decision event. The global index i is incremented every time the event happens. We call the sequence $\{a_i\}_1^N$ the *flat log* of the solver's run. We will denote by $DL(a_i)$ the decision level of the event. We consider only the case in which $dl > 0$. All conflict events other, potentially, than the last one in an unsatisfiable instance are included by this definition. It must hold that for any conflict c there exists a decision d at the same level as c. In such a case, we say that d is refuted by c. More formally:

Definition 3 (Refuted decision by a conflict)**.** *Let $a_j = (dl, c)$ be a conflict event. Let $a_k = (dl, d)$, $k < j$, be the last decision event with decision level dl preceding a_j (note that for $i \in [k+1, j-1] : DL(a_i) > dl$). We say that d is the* refuted decision *of the conflict c, and write $D(a_j) = a_k$.*

Note that because of non-chronological backtracking the opposite direction does not hold: there are decisions that do not have conflicts on their levels that refute them.

For any conflict event a_j, the range $(D(a_j), a_j)$ defines a set of events that happened after $D(a_j)$ and led to the conflicts that were resolved into the conflict a_j which, in turn, refuted $D(a_j)$. These events necessarily occurred on levels deeper than $DL(D(a_j))$.

Definition 4 (Refutation Sequence and sub-tree events)**.** *Let a_j be a conflict event with $D(a_j) = a_k$. Then the (possibly empty) sequence of events $a_{k+1}, \ldots, a_{j-1}$ is called the* Refutation Sequence *of a_j and denoted by $RS(a_j)$. Any event $a_i \in RS(a_j)$ is called a* sub-tree event *of both a_j and a_k.*

Example 3. *Consider the conflict event $a_j := (27, c_{110})$ in Fig. 5. For every event a_i that follows decision $D(a_j)$ (the decision on x_{30} in level 27) until (but not including) the conflict c_{110} it holds that $a_i \in RS(a_j)$. Note that the solver can backtrack from deeper levels to level 27 as a result of conflict events. However no event between $D(a_j)$ and a_j occurred on levels smaller or equal to 27.*

The number of resolutions for each variable is bounded from above by the number of sub-tree conflicts that were resolved into the current conflict. However, not all sub-tree conflict clauses resolve into the current 'refuting' conflict. Some of them could be caused by the imperfection of the decision heuristic and are therefore not used at this point of the

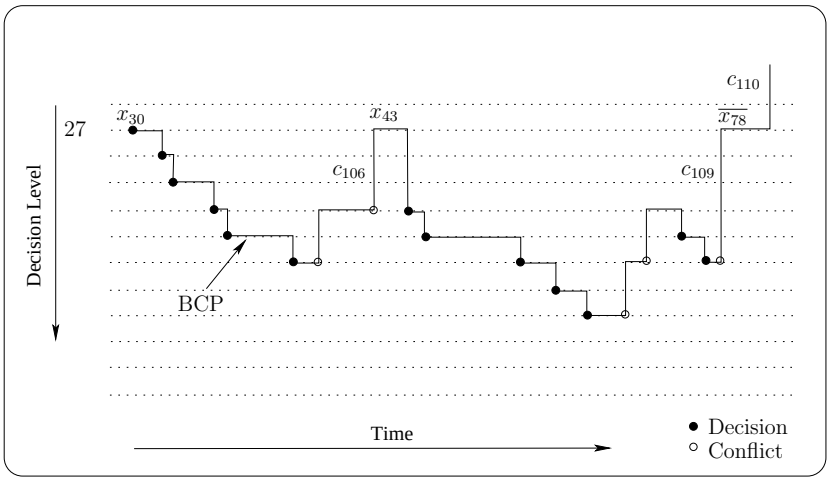

Figure 5. A possible scenario for the flow of the solver's run. After deciding x_{30} at decision level 27 the solver iteratively goes down to deeper decision levels and returns twice to level 27 with new asserted literals x_{43} and $\overline{x}_{78}$. The latter causes a conflict at level 27 and the solver backtracks to a higher decision level. Horizonal lines correspond to BCP. Some of the asserting clauses ($c_{106}, c_{109}, c_{110}$) are marked in the place in which they are created.

search. Our goal is to build a scoring system that is based solely on those conflicts that contribute to the resolution of the current conflict clause. In other words, we compute for each variable an *activity score* which reflects the *number of times it was resolved-on in the process of generating the relevant portion of the refutation sequences of recent conflicts*. We *hypothesize that this criterion for activity leads to faster solution times*.

The information in the colored resolve-graph can enable us to compute such a score.

Definition 5 (Asserting set)**.** *Let $G = (V, E_{as}, E_{nas})$ be a colored resolve-graph, and let $v \in V$ represent a conflict clause. The* Asserting set *$B(v) \subset V$ of v is the subset of (conflict) clauses that v has a solid path to them in G (i.e., a path made of E_{as} edges).*

The following theorem relates between a resolve-graph and sub-tree conflicts.

Theorem 1. *Let e_v be the conflict event that created the conflict clause v. Then the asserting set of v is contained in the refutation sequence of e_v, i.e., $B(v) \subseteq RS(e_v)$. In particular, since conflict events in $B(v)$ participate in the resolution of v by definition, they necessarily correspond to those sub-tree conflicts of e_v that participate in the resolution of v.*

Note that $B(v)$ does not necessarily include *all* the sub-tree conflicts that resolve into v, since the theorem guarantees containment in only one direction. Nevertheless, our heuristic is based on this theorem: it computes the size of the asserting set for each conflict.

In order to prove this theorem we will use the following lemmas.

Lemma 1. *Denote by stack(a_j) the stack of implied literals at the decision level $DL(a_j)$, where a_j is a decision event. Suppose that a literal t is asserted and entered into stack(a_j),*

where a_j is a decision event. Further, suppose that t is asserted by the conflict clause cl (cl is thus asserting at this point) which was created at event a_i. Then it holds that $j < i$, i.e cl was created after the decision event a_j occurred.

Proof. Right after the creation of cl, the DPLL algorithm backtracks to some level dl' with a decision event $a_k = (dl', d)$ and implies its asserted literal. It holds that $k < i$, because the solver backtracks to a decision level which already exists when cl is created. By the definition of an asserting clause, cl can be asserting exactly once, and since cl is asserting on level dl', it will never be asserting after the DPLL algorithm will backtrack from dl'. Therefore it must hold that $a_k = a_j$ $(k = j)$ and $dl' = dl$. □

Lemma 2 (Transitivity of RS)**.** *Suppose that a_i, a_j are conflict events such that $a_i \in RS(a_j)$. Then, for any event $a_k \in RS(a_i)$ it follows that $a_k \in RS(a_j)$.*

Proof. First, we will prove that $D(a_i) \in RS(a_j)$, or, in other words, that $D(a_i)$ occurred between $D(a_j)$ and a_j. Clearly, $D(a_i)$ occurred before a_i and, therefore, before a_j. Now, falsely assume that $D(a_i)$ occurred before $D(a_j)$. Then the order of events is $D(a_i), D(a_j), a_i$. However, this can not happen since $D(a_j)$ occurred on shallower (smaller) level than a_i and this contradicts the fact that all events between $D(a_i)$ and a_i occur on the deeper levels. Therefore, both $D(a_i)$ and a_i occurred between $D(a_j)$ and a_j. Now, since a_k happened between $D(a_i)$ and a_i it also happened between $D(a_j)$ and a_j and from this it holds that $a_k \in RS(a_j)$. □

Using this lemma we can now prove Theorem 1.

Proof. We need to show that any solid descendant of v is in $RS(e_v)$. By Lemma 2 it is enough to show it for the immediate solid descendants, since by transitivity of RS it then follows for any solid descendant. Now, suppose that there exists a solid edge (v, u) in the resolve-graph. By the definition of a solid edge, clause u was asserting during the resolution of v. On the one hand, u was resolved during the creation of v and, therefore, was created before v. On the other hand, by Lemma 1 it was created after $D(e_v)$. Therefore, $e_u \in RS(e_v)$. □

Definition 6 (Sub-tree weight of the conflict)**.** *Given a resolve-graph $G(V, E)$ we define for each clause v a state variable $W(v)$:*

$$W(v) = \begin{cases} \sum_{(v,u)\in E} W(u) + 1 & v \text{ is asserting} \\ 0 & \text{otherwise} \end{cases}$$

The function $W(v)$ is well-defined, since the resolve-graph is acyclic. Moreover, since the solid sub-graph rooted at v forms a tree (remember that any node has at most one incoming solid edge), $W(v)$ equals to $|B(v)| + 1$. Our recursive definition of $W(v)$ gives us a simple and convenient way to compute it as part of the resolution algorithm. Algorithm 3 is the same as Algorithm 1, with the addition of several lines: in line 5 we add $W \leftarrow 1$, at line 24 we add $W{+}{=}W(ResolveCl)$ and, finally, we set $W(NewClause) \leftarrow W$ at line 29. We need to guarantee that $W(C)$ is non-zero only when C is an asserting clause. Therefore, for any antecedent clause C, when its implied variable is unassigned we set $W(C) \leftarrow 0$.

Computing the scores of a variable

Given the earlier definitions, it is now left to show how activity score and sign score are actually computed, given that we do not have the resolve-graph in memory. For each variable v we keep two fields: $activity(v)$ and $sign\_score(v)$. At the beginning of the run $activity$ is initialized to $\max\{lit\_num(v), lit\_num(\overline{v})\}$ and $sign\_score$ to $lit\_num(v) - lit\_num(\overline{v})$. Alg. 3 shows the extended version of the resolution algorithm which computes the weights of the clauses and updates the scores. Recall that any clause weight is reset to zero when its implied variable is unassigned, so that any clause weight is contributed at most once. In order to give a priority to recent resolutions we occasionally divide both activities and sign scores by 2.

Algorithm 3 First-UIP Learning Scheme, including scoring

```
    procedure AnalyzeConflict(Clause: conflict)
 2:     currentClause ← conflict;
        ResolveNum ← 0;
 4:     NewClause ← ∅;
        wght ← 1;
 6:     repeat
            for each literal lit ∈ currentClause do
 8:             v ← var(lit);
                if v is not marked then
10:                 Mark v;
                    if dlevel(v) = CurrentLevel then
12:                     ++ ResolveNum;
                    else
14:                     NewClause ← NewClause ∪ {lit};
                    end if
16:             end if
            end for
18:         u ← last marked literal on the assignment stack;
            Unmark var(u);
20:         activity(var(u)) += wght;
            sign_score(var(u)) −= wght · sign(u);
22:         −− ResolveNum;
            ResolveCl ← Antecedent(u);
24:         wght+= W(ResolveCl);
            currentClause ← ResolveCl \ {u};
26:     until ResolveNum = 0;
        Unmark literals in NewClause;
28:     NewClause ← NewClause ∪ {ū};
        W(NewClause) ← wght ;
30:     Add NewClause to the clause database;
    end procedure
```

Our decision heuristic chooses a variable from the given clause with a biggest activity and then chooses its value according to the sign score: TRUE for the positive values and FALSE for the negative values of the sign score.

5. Experiments

We present several sets of experiments. The first set compares HaifaSat to Berkmin, and the second set compares HaifaSat to Rsat on the benchmarks from the 2007 competition.

HaifaSat vs. Berkmin.

Table 1 shows experiments on an Intel 2.5Ghz computer with 1GB memory running Linux, sorted according to the winning strategy, which is CMTF combined with the RBS scoring technique. The benchmark set is comprised of 165 industrial instances used in various SAT competitions. In particular, *fifo8, bmc2, CheckerInterchange, comb, f2clk, ip, fvp2, IBM02 and w08* are hard industrial benchmarks from SAT02; *hanoi* and *hanoi03* participated in SAT02 and SAT03; *pipe03* is from SAT03 and *01_rule, 11_rule_2, 22_rule, pipe-sat-1-1, sat02, vis-bmc, vliw_unsat_2.0* are from SAT04 each instance was set to 3000 seconds. If an instance could not be solved in this time limit, 3000 sec. were added as its solving time. All configurations are implemented on top of HaifaSat, which guarantees that the figures faithfully represent the quality of the various heuristics, as far as these benchmarks are representative. The results show that using CMTF instead of Berkmin's heuristic for choosing a clause leads to an average reduction of 10% in run time and 12-25% in the number of fails (depending on the score heuristic). It also shows a 23% reduction in run time when using RBS rather than VSIDS as a score system, and a corresponding 20-30% reduction in the number of fails. The differences in run times between HaifaSat running the berkmin heuristic and Berkmin561 are small: the latter solves these instances in 210793 sec. and 53 timeouts. We also ran zChaff2004.5.13 on these formulas: it solves them in 210395 sec, and has 53 timeouts.

HaifaSat vs. Rsat.

On the benchmarks listed in Table 1 Rsat performs very well: It fails only in 17 instances, and the overall runtime is 84371 sec., which is 43% less than the runtime of HaifaSat. The detailed results appear in Table 2.

Table 3 compares the results of HaifaSat and Rsat on the benchmarks from the 2007 competition with a timeout set to 1200 sec. HaifaSat is about 32% slower and timesout in 33% more cases. From the 108 instances that at least one of them solved before time out, Haifasat is better in 30 (27.7%). The experiments were run on the same machine as the first set of benchmarks.

6. Summary

We presented an abstraction/refinement model for analyzing and developing SAT decision heuristics. Satisfying a conflict clause before satisfying the clauses from which it was resolved, can be seen according to our model as satisfying an abstract model before satisfying

Table 1. A comparison of various configurations, showing separately the advantage of CMTF, the heuristic for choosing the next clause from which the decided variables will be chosen, and RBS, the heuristic for choosing the variable from this clause and its sign. The second column indicates the number of instances in each benchmark family.

		BERKMIN+RBS		BERKMIN+VSIDS		CMTF+RBS		CMTF+VSIDS	
Benchmark	#	time	fails	time	fails	time	fails	time	fails
hanoi	5	389	0	530	0	130	0	**74**	0
ip	4	**191**	0	395	0	203	0	324	0
hanoi03	4	1548	0	1342	0	426	0	**386**	0
CheckerI-C	4	1368	0	3323	0	**681**	0	3457	0
bmc2	6	1731	0	1030	0	1261	0	**1006**	0
pipe03	3	**845**	0	6459	2	1339	0	6160	1
fifo8	4	1877	0	3944	0	**1832**	0	3382	0
fvp2	22	**1385**	0	8638	1	1995	0	11233	3
w08	3	**2548**	0	5347	1	2680	0	4453	0
pipe-sat-1-1	10	**1743**	0	3881	0	3310	0	6053	0
IBM02	8	7083	1	9710	1	**3875**	0	7163	0
f2clk	3	4389	1	5135	1	**4058**	1	4538	1
comb	3	3915	1	**3681**	1	4131	1	4034	1
vis-bmc	8	15284	3	**7905**	1	13767	3	10119	2
sat02	9	17518	4	22785	5	**17329**	4	21262	4
01_rule	20	22742	4	33642	9	**19171**	2	23689	5
vliw_unsat_2	8	**16600**	4	24003	8	19425	5	22756	7
11_rule_2	20	31699	8	34006	10	**22974**	6	28358	6
22_rule	20	28844	8	33201	10	**27596**	8	30669	8
Total:	165	161706	34	208967	50	**146193**	30	189125	38

Table 2. Rsat's results on the same benchmark set listed in Table 1. Rsat's runtime is lower by 43% comparing to HaifaSat, and fails in 44% less cases.

		Rsat2.0	
Benchmark	#	time	fails
hanoi	5	79	0
ip	4	79	0
hanoi03	4	510	0
CheckerI-C	4	116	0
bmc2	6	258	0
pipe03	3	6668	2
fifo8	4	310	0
fvp2	22	5831	1
w08	3	720	0
pipe-sat-1-1	10	4705	1
IBM02	8	3392	0
f2clk	3	1291	0
comb	3	3319	1
vis-bmc	8	6979	1
sat02	9	11256	3
01_rule	20	1826	0
vliw_u_2.0	8	24000	8
11_rule_2	20	4358	0
22_rule	20	8664	0
Total:	164	84371	17

Table 3. A comparison of HaifaSat and Rsat on the industrial benchmarks from the 2007 competition, with a timeout of 1200 sec.

		HaifaSAT			Rsat		
Benchmark	#	Time	Timeouts	Memory	Time	Timeouts	Memory
anbulagan	60	43005	31	1	46698	37	0
babic	30	8098	4	0	224	0	0
crypto	10	11999	10	0	6202	2	0
fuhs	16	17857	14	0	15029	10	0
grieu	10	7722	6	0	6622	5	0
jarvisalo	7	6221	5	0	5521	4	0
manolios	10	11467	9	0	7180	5	0
narain	5	3311	2	1	2266	1	0
palacios	27	28852	19	0	15196	11	0
Total:	175	138535	100	2	104942	75	0

a more concrete version of it. Our Clause-Move-To-Front decision heuristic, according to this model, attempts to satisfy clauses in an order associated with the resolve-graph. CMTF does not require to maintain the resolve-graph in memory, however: it only exploits the connection between each conflict clause and its immediate neighbors on this graph. Perhaps future heuristics based on this graph will find a way to improve the balance between the memory consumption imposed by saving this graph and the quality of the decision order. A recent publication by Yorav and Shacham [18] show how to exploit the fact that most clauses are erased in order to maintain the entire resolve graph in memory. Perhaps this is the key for more sophisticated decision heuristics based on an analysis of this graph. We also presented a heuristic for choosing the next variable and sign from the clause chosen by CMTF. Our Resolution-Based-Scoring heuristic scores variables according to their involvement ('activity') in refuting recent decisions. Our experiments show that CMTF and RBS either separately or combined are better than Berkmin and the VSIDS decision heuristics. As a whole tool, HaifaSat, while not truly competitive anymore, holds reasonably well comparing to the currently best solver, Rsat.

7. Acknowledgments

We thank Maya Koifman for helpful comments on an earlier version of this paper.

References

[1] Nina Amla and Kenneth L. McMillan. Combining abstraction refinement and sat-based model checking. In *TACAS*, pages 405–419, 2007.

[2] P. Beame., H. Kautz, and A. Sabharwal. Towards understanding and harnessing the potential of clause learning. *Journal of Artificial Intelligence Research*, **22**:319–351, 2004.

[3] A. Biere, A. Cimatti, E. Clarke, and Y. Zhu. Symbolic model checking without BDDs. In *Proc. of the Workshop on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'99)*, LNCS, pages 193–207. Springer-Verlag, 1999.

[4] E.M. Clarke, O. Grumberg, S. Jha, Y. Lu, and H. Veith. Counterexample-guided abstraction refinement. *J. ACM*, **50**(5):752–794, 2003.

[5] E.M. Clarke, A. Gupta, J. Kukula, and O. Strichman. SAT based abstraction-refinement using ILP and machine-learning techniques. In E. Brinksma and K.G. Larsen, editors, *Proc.* 14^{th} *Intl. Conference on Computer Aided Verification (CAV'02)*, **2404** of *LNCS*, pages 265–279, Copenhagen, Denmark, July 2002. Springer-Verlag.

[6] Nachum Dershowitz, Ziyad Hanna, and Alexander Nadel. A clause-based heuristic for sat solvers. In *SAT*, pages 46–60, 2005.

[7] Niklas Eén and Niklas Sörensson. An extensible sat-solver. In *SAT*, pages 502–518, 2003.

[8] Roman Gershman and Ofer Strichman. Haifasat: A new robust SAT solver. In Yaron Wolfsthal Shmuel Ur, Eyal Bin, editor, *First International Haifa Verification Conference*, **3875** of *Lect. Notes in Comp. Sci.*, pages 76 – 89. Springer-Verlag, 2005.

[9] M. Glusman, G. Kamhi, S. Mador-Haim, R. Fraer, and M. Y. Vardi. Multiple-counterexample guided iterative abstraction refinement: An industrial evaluation. In *Tools and Algorithms for the Construction and Analysis of Systems, TACAS 2003,*, LNCS, pages 176–191, 2003.

[10] E. Goldberg and Y. Novikov. Berkmin: A fast and robust sat-solver. In *Design, Automation and Test in Europe Conference and Exhibition (DATE'02)*, page 142, Paris, 2002.

[11] Anubhav Gupta and Ofer Strichman. Abstraction refinement for bounded model checking. In K. Etessami and S. Rajamani, editors, *Proc.* 17^{th} *Intl. Conference on Computer Aided Verification (CAV'05)*, **3576** of *Lect. Notes in Comp. Sci.*, pages 112–124, Edinburgh, July 2005. Springer-Verlag.

[12] R. Kurshan. *Computer aided verification of coordinating processes*. Princeton University Press, 1994.

[13] João P. Marques-Silva and Karem A. Sakallah. GRASP - a new search algorithm for satisfiability. In *Proceedings of the 1996 International Conference on Computer-Aided Design (ICCAD '96)*, pages 220–227. IEEE Computer Society Press, 1996.

[14] João P. Marques-Silva and Karem A. Sakallah. GRASP: A search algorithm for propositional satisfiability. *IEEE Transactions on Computers*, **48**:506–516, 1999.

[15] M. Moskewicz, C. Madigan, Y. Zhao, L. Zhang, and S. Malik. Chaff: Engineering an efficient SAT solver. In *Proc. Design Automation Conference (DAC'01)*, 2001.

[16] Knot Pipatsrisawat and Adnan Darwiche. Rsat 2.0: Sat solver description. SAT competition'07, 2007.

[17] L. Ryan. Efficient algorithms for clause-learning SAT solvers. Master's thesis, Simon Fraser University, 2004.

[18] Ohad Shacham and Karen Yorav. On-the-fly resolve trace minimization. In *DAC*, pages 594–599, 2007.

[19] Laurent Simon and Daniel Le Berre. SAT competition 2003. http://www.satcompetition.org/2003/, 2003.

[20] L. Zhang, C. Madigan, M. Moskewicz, and S. Malik. Efficient conflict driven learning in a boolean satisfiability solver. In *ICCAD*, 2001.

[21] L. Zhang and S. Malik. Extracting small unsatisfiable cores from unsatisfiable boolean formulas. In *In Sixth International Conference on Theory and Applications of Satisfiability Testing (SAT2003), S. Margherita Ligure*, 2003.