Building Decision Procedures for Modal Logics from Propositional Decision Procedures: The Case Study of Modal $K(m)$

Fausto Giunchiglia
IRST, 38050 Povo (TN), Italy; and DISA, via Inama 5, 38100 Trento, Italy

and

Roberto Sebastiani
DISA, via Inama 5, 38100 Trento, Italy
E-mail: fausto@irst.itc.it, rseba@cs.unitn.it

The goal of this paper is to propose a new technique for developing decision procedures for propositional modal logics. The basic idea is that propositional modal decision procedures should be developed on top of propositional decision procedures. As a case study, we consider satisfiability in modal $K(m)$, that is modal $K$ with $m$ modalities, and develop an algorithm, called $K$sat, on top of an implementation of the Davis–Putnam–Longemann–Loveland procedure. $K$sat is thoroughly tested and compared with various procedures and in particular with the state-of-the-art tableau-based system KRIS. The experimental results show that $K$sat outperforms KRIS and the other systems of orders of magnitude, highlight an intrinsic weakness of tableau-based decision procedures, and provide partial evidence of a phase transition phenomenon for $K(m)$.

1. INTRODUCTION

The goal of this paper is to describe a new technique for developing decision procedures for propositional modal logics. Our approach is based on two basic intuitions. The first is that modal reasoning can be implemented as an appropriate...
composition of reasoning inside multiple propositional theories (or models, if one thinks of satisfiability). \[\text{GS94}\] shows how this can be done for provability in the most common normal modal logics; \[\text{GSGF93}\] extends these results to various nonnormal modal logics. Similar ideas are implicit, even if never spelled out as such, in the tableaux for normal modal logics (see, e.g., \[\text{Fit88, Mas94}\]). The second is that propositional reasoning can be performed very efficiently by using state-of-the-art propositional decision procedures, e.g., Davis–Putnam–Longemann–Loveland (DPLL) \[\text{DLL62}\], OBDD \[\text{Bry92}\], KE \[\text{DM94}\], or even partial decision procedures like GSAT \[\text{SLM92}\]. This allows us to exploit the huge amount of technology developed in this area, which is very advanced and well understood.

In this paper we concentrate on the satisfiability problem, not restricted to CNF, for modal \(K(m)\), that is \(K\) with \(m\) modalities, and use DPLL for testing propositional satisfiability (SAT).\(^2\) The reasons for this choice are manyfold. First, \(K(m)\) is an interesting logic per se; for instance it is well known that \(K(m)\) is a notational variant of the terminological logic \(\mathcal{L}^k\) \[\text{Sch91}\]. (This is actually the main motivation for this work.) Second, \(K\) is the smallest normal modal logic. The algorithm(s) described in this paper can be (more or less trivially) extended to the other normal modal logics, for instance along the lines of what is described in \[\text{Fit88, Mas94, GS94}\]. Third, DPLL is the most widely known and studied propositional decision procedure and also one of the most efficient \[\text{BB92}\] (but see also \[\text{US94}\]).

We have tested and compared various refinements of the algorithm we have developed, called \(\text{Ksat}\), among themselves and also against the decision procedures and systems for modal logics we have been able to acquire. It turns out that all these implementations are tableau-based. In this paper we consider two of them. The first is a straightforward implementation of the algorithm described in \[\text{HNSS90}\], due to B. Nebel and E. Franconi. This procedure is called \text{Tableau} from now on. The second is the state-of-the-art system \(\text{Kris}\) described in \[\text{HNSS90, BFH+94}\]. The testing confirms our original intuitions and also highlights some very interesting and unexpected phenomena. We can summarize our results as follows:

1. \(\text{Ksat}\) outperforms all the other implementations of orders of magnitude;
2. increasing the quality of the implementation produces an increase of performance. This increase is only quantitative and it does not change the shape of the performance curves;
3. tableau-based decision procedures are intrinsically less efficient than \(\text{Ksat}\). This difference is quantitative but also, and more importantly, qualitative. The efficiency of tableau-based decision procedures keeps decreasing with the increase of the length of the input formulas (normalized to the number of propositional variables), while the efficiency of \(\text{Ksat}\), after having decreased for a while, increases again;
4. \(\text{Ksat}\) produces what looks like a phase transition phenomenon \[\text{MSL92, WH94}\]. If the current (still partial) evidence is confirmed, this is the first time that

\[\text{AG93}\] and \[\text{Sch94}\] show how decision procedures for CNF formulas like DPLL and GSAT can be modified to work for non-CNF formulas.
this phenomenon, well known for SAT and other NP-hard problems, is found in modal logics.

This paper is structured as follows. In Section 2 we introduce formal framework, definitions, and notation and provide three simple but important results which motivate the algorithm, and directly imply its correctness and completeness. In Section 3 we describe the algorithm. The presentation is done incrementally, first by giving the algorithm implementing the basic idea and then by providing three enhancements, each improving on a specific aspect. In Section 4 we test the versions of \textsc{Ksat}, as described in Section 3, a further improved version of \textsc{Ksat}, called \textsc{Ksat}_s (where \textit{s} stands for smart), which implements some smart implementation tricks, and compare them with \textsc{Tableau} and \textsc{Kris}.\footnote{The various versions of \textsc{Ksat}, the test code, and all the results presented in this paper are available via anonymous FTP at ftp.mrg.dist.unige.it in pub/mrg-systems/\textsc{ksat}/. \textsc{Tableau} is available at ftp.mrg.dist.unige.it in pub/mrg-systems/\textsc{tableau}/. \textsc{Kris} is available at ftp.dfki.uni-sb.de in /pub/tacos/\textsc{Kris}/.} This allows us to show and discuss the first three results hinted above. Finally, in Section 5, we analyze in detail the efficiency curves of \textsc{Ksat}. This allows us to study the behavior of \textsc{Ksat} on \textsc{K}(\textit{m}) and, among other things, to clearly identify the phase transition phenomenon (fourth result) mentioned above. Section 6 provides some conclusive remarks and describes the directions for future work.

2. THE FORMAL FRAMEWORK

We use here the standard modal logics syntax and Kripke semantics (see, e.g., [HM92]). In particular we use \&, \lor, \neg, T, F, \rightarrow, \equiv, \square_1, \ldots, \square_m as logical symbols, and \textit{A}_1, \textit{A}_2, \ldots as propositional variables. We use Greek letters \textit{\alpha}, \textit{\beta}, \ldots to denote \textsc{K}(\textit{m}) formulas. We call depth of \textit{\varphi}, written \textit{depth}(\textit{\varphi}), the maximum number of nested modal operators in \textit{\varphi}.

Let us call \textit{atom} any formula that cannot be decomposed propositionally, that is, any formula whose main connective is not propositional. Examples of atoms are \textit{A}_1, \textit{g}_{11}(\textit{A}_1 \land \textit{A}_2), and \textit{g}_2(\square_1 \textit{A}_1 \lor \neg \textit{A}_2). A \textit{literal} is either an atom or its negation. Given a formula \textit{\varphi}, an atom [\textit{literal}] is a top-level atom [\textit{literal}] for \textit{\varphi} if and only if it occurs in \textit{\varphi} and under the scope of no boxes. \textit{TopAtoms}(\textit{\varphi}) is the set of the top level atoms of \textit{\varphi}.

A \textit{truth assignment} \textit{\mu} for a modal formula \textit{\varphi} is a truth value assignment to all the top-level atoms of \textit{\varphi}:

\[
\begin{align*}
\mu = \{ & \square_1 x_{11} = \text{True}, \ldots, \square_1 x_{1N_1} = \text{True}, \square_1 \beta_{11} = \text{False}, \ldots, \square_1 \beta_{1M_1} = \text{False}, \\
& \ldots \\
& \square_m x_{m1} = \text{True}, \ldots, \square_m x_{mN_m} = \text{True}, \square_m \beta_{m1} = \text{False}, \ldots \\
& A'_1 = \text{True}, \ldots, A'_R = \text{True}, A'_{R+1} = \text{False}, \ldots, A'_{S} = \text{False} \},
\end{align*}
\]

where, e.g., “\square_1 x_{11} = \text{True}” should be read as “assign \text{True} to \square_1 x_{11}”. \textit{A}' \in \{ \textit{A}_1, \textit{A}_2, \ldots \}, for every \textit{i}. Identical atoms are always assigned identical truth values. A crucial property
of truth assignments is that different atoms, e.g., $\Box g_2 (\neg \psi_1 \lor \psi_2)$ and $\Box g_2 (\psi_2 \lor \psi_1)$, or, even, $\Box_2 \psi_1$ and $\Box_2 (\psi_2 \land \psi_1)$, are treated differently and may thus be assigned different truth values. In this respect, notice that the $\psi$’s are not part of the language. This allows us to avoid assignments like, e.g., $\mu = \{ \Box_2 \psi = True, \Box_2 \neg \psi = True, ... \}$ which are intrinsically inconsistent.

A partial truth assignment $\mu$ for $\varphi$ is a truth value assignment to a proper subset of the top-level atoms of $\varphi$. If $\mu_2 \subseteq \mu_1$, then we say that $\mu_1$ extends $\mu_2$ and that $\mu_2$ subsumes $\mu_1$. A restricted truth assignment

$$\mu' = \{ \Box_2 \sigma_{\psi_1} = True, ..., \Box_2 \sigma_{\psi_M} = True, \Box_2 \beta_{\psi_1} = False, ..., \Box_2 \beta_{\psi_M} = False \}$$

is given by restricting $\mu$ to the set of atoms in the form $\Box_2 \psi$, where $1 \leq r \leq m$.

Notationally, we use the Greek letters $\mu, \eta$ to represent truth assignments. Furthermore, from now on we often write a truth assignment as a formula,

$$\mu = \bigwedge_i \Box_2 \sigma_{\psi_i} \land \bigwedge_j \neg \Box_2 \beta_{\psi_j} \land \cdots \land \bigwedge_i \Box_2 \sigma_{\mu_1} \land \bigwedge_j \neg \Box_2 \beta_{\mu_j} \land \gamma; \quad (1)$$

where the $\Box_2 \sigma_{\psi_i}$’s are the boxed atoms set to $True$, $\Box_2 \beta_{\psi_j}$’s are the boxed atoms set to $False$, and $\gamma = \bigwedge_{k=1}^m A_k \land \bigwedge_{k=1}^n \neg A_k$ is a conjunction of propositional literals. Notice that $\gamma$ is always consistent. Similarly, we represent restricted assignments as:

$$\mu' = \bigwedge_i \Box_2 \sigma_{\psi_i} \land \bigwedge_j \neg \Box_2 \beta_{\psi_j}. \quad (2)$$

Furthermore, we say that an assignment [restricted assignment] is $K(m)$-satisfiable meaning that its corresponding formula (1) [(2)] is $K(m)$-satisfiable.

**Example 1.** Consider the following $K(2)$ formula $\varphi$:

$$\varphi = \{ \neg \Box_1 (\neg A_3 \lor \neg A_1 \lor A_2) \lor A_1 \lor A_5 \} \land \{ A_2 \lor \neg A_5 \lor \Box_2 (\neg A_2 \lor \neg A_4 \lor \neg A_3) \} \land \{ A_1 \lor \Box_2 (\neg A_4 \lor A_5 \lor A_2) \lor A_2 \} \land \{ \neg \Box_2 (A_4 \lor \neg A_3 \lor A_1) \lor \neg \Box_1 (A_4 \lor \neg A_2 \lor A_3) \lor \neg A_5 \} \land \{ \neg A_3 \lor A_1 \lor \Box_2 (\neg A_4 \lor A_5 \lor A_3) \} \land \{ \Box_1 (\neg A_4 \lor A_4 \lor A_3) \lor \Box_1 (\neg A_1 \lor A_4 \lor A_3) \lor \neg A_1 \} \land \{ A_1 \lor \Box_2 (\neg A_2 \lor A_1 \lor A_4) \lor A_2 \}$$

Consider the following truth assignment $\mu$, which sets to $T$ the literals which are underlined:
\[ \mu = [\Box(A_5 \lor A_4 \lor A_3) \land [\Box((A_4 \lor A_1 \lor A_4))] ] \land \neg [\Box(A_3 \lor A_4 \lor A_2) \land \neg [\Box(A_4 \lor A_2 \lor A_3)]] \land \neg [\Box(A_4 \lor A_3 \lor A_2) \land [\Box((A_4 \lor A_2 \lor A_3))] ] \]

Notice that the two occurrences of \([\Box(A_4 \lor A_5 \lor A_2)]\) in rows 3 and 5 of \(\varphi\) are both assigned True. \(\mu\) gives rise to two restricted assignments \(\mu^1\) and \(\mu^2\):

\[ \mu^1 = [\Box(A_5 \lor A_4 \lor A_3) \land [\Box((A_4 \lor A_1 \lor A_4))] ] \land \neg [\Box(A_3 \lor A_4 \lor A_2) \land \neg [\Box(A_4 \lor A_2 \lor A_3)]] \land \neg [\Box(A_4 \lor A_3 \lor A_2) \land [\Box((A_4 \lor A_2 \lor A_3))] ] \]

\[ \mu^2 = [\Box(A_5 \lor A_4 \lor A_2) \land [\Box((A_4 \lor A_2 \lor A_3))] ] \land \neg [\Box(A_3 \lor A_4 \lor A_2) \land [\Box((A_4 \lor A_2 \lor A_3))] ] \land \neg [\Box(A_4 \lor A_3 \lor A_2) \land [\Box((A_4 \lor A_2 \lor A_3))] ] \]

A truth assignment \(\mu\) for \(\varphi\) propositionally satisfies \(\varphi\), written \(\mu \models_p \varphi\), if and only if it makes \(\varphi\) evaluate to True; that is, for all subformulas \(\varphi_1, \varphi_2\) of \(\varphi\):

\[ \mu \models_p \varphi_1, \varphi_1 \in \text{Top.Atoms}(\varphi) \iff \varphi_1 = \text{True} \in \mu; \]

\[ \mu \models_p \neg \varphi_1 \iff \mu \not\models_p \varphi_1; \]

\[ \mu \models_p \varphi_1 \land \varphi_2 \iff \mu \models_p \varphi_1 \text{ and } \mu \models_p \varphi_2. \]

We say that a partial truth assignment \(\mu\) propositionally satisfies \(\varphi\) if and only if all the assignments for \(\varphi\) which extend \(\mu\) propositionally satisfy \(\varphi\). For instance, if \(\varphi = [\Box, \varphi_1 \lor \neg [\Box, \varphi_2]_2\), then the partial assignment \(\mu = [\Box, \varphi_1 = \text{True}]\) is such that \(\mu \models_p \varphi\). In fact, both \(\{[\Box, \varphi_1 = \text{True}, [\Box, \varphi_2 = \text{True}]\}\) and \(\{[\Box, \varphi_1 = \text{True}, [\Box, \varphi_2 = \text{False}]\}\) propositionally satisfy \(\varphi\). From now on, if not otherwise specified, when dealing with propositional satisfiability we do not distinguish between assignments and partial assignments.

**Theorem 1.** A modal formula \(\varphi\) is \(K(m)\)-satisfiable if and only if there exists a \(K(m)\)-satisfiable truth assignment \(\mu\) such that \(\mu \models_p \varphi\).

Notice that this result is not committed to \(K(m)\), and it can be extended to any logic which gives a standard interpretation to the propositional connectives.

**Theorem 2.** The truth assignment \(\mu\) of Eq. (1) is \(K(m)\)-satisfiable if and only if the restricted truth assignment \(\mu'\) of Eq. (2.2) is \(K(m)\)-satisfiable, for all \(\Box, \land\).
Theorem 3. The restricted assignment $\mu'$ of Eq. (2) is $K(m)$-satisfiable if and only if, for every $\neg \square$, $\beta_j$ occurring in $\mu'$, the following wff is $K(m)$-satisfiable

$$\varphi' = \bigwedge_i x_i \land \neg \beta_j.$$  

(3)

Theorem 1 reduces the $K(m)$-satisfiability of a formula $\varphi$ to the $K(m)$-satisfiability of its truth assignments. Theorems 2 and 3 show how to reduce the latter to the $K(m)$-satisfiability of formulas of smaller depth. This process can be applied recursively, decreasing the depth of the formula considered at each iteration.

Example 2. In Example 1, consider the formula $\varphi$ and the assignments $\mu$, $\mu^1$, and $\mu^2$, $\mu$ propositionally satisfies $\varphi$, as it verifies one literal for every clause. Thus, for Theorem 1, $\mu$ is $K(m)$-satisfiable if and only if $\varphi$ is $K(m)$-satisfiable. For Theorem 2 $\mu$ is $K(m)$-satisfiable if and only if both $\mu^1$ and $\mu^2$ are. For Theorem 3, $\mu^2$ is trivially $K(m)$-satisfiable, as it contains no negated boxes, and $\mu^1$ is $K(m)$-satisfiable if and only if each of the wffs

$$\varphi^{11} = \bigwedge_i x_i \land \neg \beta_{11}$$

$$= (\neg A_5 \lor A_4 \lor A_3) \land (\neg A_2 \lor A_1 \lor A_4) \land A_3 \land A_1 \land \neg A_2,$$

$$\varphi^{12} = \bigwedge_i x_i \land \neg \beta_{12}$$

$$= (\neg A_5 \lor A_4 \lor A_3) \land (\neg A_2 \lor A_1 \lor A_4) \land \neg A_4 \land A_2 \land \neg A_3$$

are $K(m)$-satisfiable. As they both are, then $\varphi$ is $K(m)$-satisfiable. 

3. THE ALGORITHM(S)

3.1. Version 1: The Basic Algorithm

The basic version of $\text{KSAT}$ is reported in Fig. 1. $\text{KSAT}$ takes in input a modal propositional wff $\varphi$ and returns a truth value asserting whether $\varphi$ is $K(m)$-satisfiable or not. $\text{KSAT}$ invokes $\text{KSAT}_W$ (where $W$ stands for wff), passing as arguments $\varphi$ and the truth value $T$ (i.e., by (1), the empty truth assignment). $\text{KSAT}_W$ tries to build a $K(m)$-satisfiable truth assignment $\mu$ satisfying $\varphi$. This is done recursively, according to the following steps:

---

(base) If $\varphi = T$, then $\mu$ satisfies $\varphi$. Thus, if $\mu$ is $K(m)$-satisfiable, then $\varphi$ is $K(m)$-satisfiable. Therefore $\text{KSAT}_W$ invokes $\text{KSAT}_{A}(\mu)$ (where $A$ stands for (truth) assignment). $\text{KSAT}_{A}$ returns a truth value asserting whether $\mu$ is $K(m)$-satisfiable or not.

(backtrack) If $\varphi = F$, then $\mu$ cannot be a truth assignment for $\varphi$. Therefore $\text{KSAT}_W$ returns $\text{False}$. 

function \text{KSAT}(\varphi) \\
\text{return } \text{KSAT}_W(\varphi, T); \\

function \text{KSAT}_W(\varphi, \mu) \\
\text{if } \varphi \equiv T \\
\quad \text{then return } \text{KSAT}_A(\mu); \\
\text{/* base */} \\
\text{if } \varphi \equiv F \\
\quad \text{then return } \text{False}; \\
\text{/* backtrack */} \\
\text{if } \{ \text{a unit clause } l \text{ occurs in } \varphi \} \\
\quad \text{then return } \text{KSAT}_W(\text{assign}(l, \varphi), \mu \land l); \\
\text{/* unit */} \\
l := \text{choose-literal}(\varphi); \\
\text{if } \text{KSAT}_W(\text{assign}(l, \varphi), \mu \land l) \text{ or } \\
\quad \text{KSAT}_W(\text{assign}(\neg l, \varphi), \mu \land \neg l); \\
\text{/* split */} \\

function \text{KSAT}_A(\land_i \square_i \alpha_i \land \land_j \neg \square_i \beta_{ij} \land \ldots \land \land_m \square_m \alpha_m \land \land_j \neg \square_m \beta_{mj} \land \gamma) \\
\text{for each box index } r \text{ do} \\
\quad \text{if not } \text{KSAT}_{RA}(\land_i \square_r \alpha_i \land \land_j \neg \square_r \beta_{rj}) \\
\quad \quad \text{then return } \text{False}; \\
\text{return } \text{True}; \\

function \text{KSAT}_{RA}(\land_i \square_r \alpha_i \land \land_j \neg \square_r \beta_{rj}) \\
\text{for each conjunct } \neg \square_r \beta_{rj} \text{ do} \\
\quad \text{if not } \text{KSAT}(\land_i \alpha_i \land \neg \beta_{rj}) \\
\quad \quad \text{then return } \text{False}; \\
\text{return } \text{True}; \\

\text{FIG. 1. The basic version of KSAT algorithm.} \\

— (unit) If a literal \( l \) occurs in \( \varphi \) as a unit clause, then \( l \) must be assigned \( T \).\(^4\) 
To obtain this, \text{KSAT}_W is invoked recursively with arguments the wff returned by \text{assign}(l, \varphi) and the assignment obtained by adding \( l \) to \( \mu \). \text{assign}(l, \varphi) substitutes every occurrence of \( l \) in \( \varphi \) with \( T \) and evaluates the result. 

— (split) If none of the above situations occur, then \text{choose-literal}(\varphi) returns an unassigned literal \( l \) according to some heuristic criterion. Then \text{KSAT}_W is first invoked recursively with arguments \text{assign}(l, \varphi) and \( \mu \land l \). If the result is negative, then \text{KSAT}_W is invoked with arguments \text{assign}(\neg l, \varphi) and \( \mu \land \neg l \). 

\text{KSAT}_W is a variant of a non-CNF version of DPLL. Unlike DPLL (which returns True), whenever an assignment \( \mu \) has been found, \text{KSAT}_W invokes \text{KSAT}_A(\mu). Essentially, DPLL is used to generate truth assignments whose \( K(m) \)-satisfiability is recursively checked by \text{KSAT}_A. \text{KSAT}_A(\mu) invokes \text{KSAT}_{RA}(\mu_r) (where \( RA \) stands for restricted assignment) for any index \( r \) such that \( \square_r \) occurs in \( \mu \). This is repeated until either \text{KSAT}_{RA} returns a negative value (in which case \text{KSAT}_A(\mu) returns False) or no more \( \square_r \)'s are available (in which case \text{KSAT}_A(\mu) returns True). \text{KSAT}_{RA}(\mu_r) invokes \text{KSAT}(\varphi''') for any conjunct \( \neg \square_r \beta_{rj} \) occurring in \( \mu_r \). Again, this is repeated until either \text{KSAT} returns a negative value (in which case \text{KSAT}_{RA}(\mu_r) returns False) or no more \( \neg \square_r \beta_{rj} \)'s are available (in which case \text{KSAT}_{RA}(\mu_r) returns True). 
Notice that \text{KSAT}_W, \text{KSAT}_A, and \text{KSAT}_{RA} are a direct implementation of Theorems 1, 2, and 3, respectively. This guarantees their correctness and completeness. 

\(^4\) A notion of unit clause for non-CNF propositional wffs is given in [AG93].
3.2. Version 2: Sorting Modal Atoms

One of the main causes of inefficiency of Version 1 is the number of truth assignments found by $\text{Ksat}$, which can be very large. This is a direct consequence of the large number of distinct modal atoms which can occur inside a $K(m)$ wff. Generally speaking, a solution which allows for a drastic reduction of distinct modal atoms could be to treat logically equivalent modal atoms as the same atom. Unfortunately, this would have unacceptable computational costs. Nevertheless, some low-cost preprocessing can be performed which collapses together trivially equivalent modal atoms. For instance, all modal atoms can be (internally) sorted, according to some order on sub-wffs. (The specific ordering is irrelevant as long as there is one.) This avoids assigning different truth values to permutations of the same sub-wffs.

Example 3. Consider the modal atoms occurring in the wff $\phi$ in Example 1 (e.g., the atom $\Box_i(\neg A_1 \lor A_2)$ in the first row). For any atom in $\phi$ there may be up to $3! = 6$ equivalent permutations, which are all mapped into one atom (e.g., $\Box_i(\neg A_1 \lor A_2 \lor \neg A_3)$) if the modal atoms are sorted.

3.3. Version 3: Factorizing $\land_i x_i$

In Fig. 1, $\text{Ksat}_{RA}$ invokes repeatedly $\text{Ksat}$ passing as arguments wffs of the form $\land_i x_i \land \neg \beta_1, \ldots, \land_i x_i \land \neg \beta_M$. All these wffs have the conjuncts $\land_i x_i$ in common. At the $j$th call, $\text{Ksat}$ searches for a $K(m)$-satisfiable truth assignment $\eta_j$ satisfying $\land_i x_i \land \neg \beta_j$. This is done from scratch, i.e., without trying to reuse any of the previously computed assignments $\eta_1 \cdots \eta_{j-1}$ or their restrictions to $\land_i x_i$. The idea underlying Version 3 is to factorize the search of the truth assignments satisfying $\land_i x_i$. Given $\alpha = \land_i x_i$ and a nonempty set $B = \{ \beta_1, \ldots, \beta_M \}$, a propositional algorithm is used to find a sequence of truth assignments $\eta_1 \eta_2 \cdots$ satisfying $\alpha$. At the $k$th truth assignment $\eta_k$, all the wffs $\beta_j$'s compatible with $\eta_k$ (i.e., such that

```
function $\text{Ksat}_{RA}((\land_i \Box_x \alpha_1 \land \land_j \neg \Box_x \beta_{i+1}))$

if {there is no conjunct "\land_j \neg \Box_x \beta_{i+1}"} 
    then return True;

return $\text{Is-Empty}(\text{Incompatible-Subset}(\land_i \land \alpha, T, \{\beta_1, \ldots, \beta_M\});$
```

```
function $\text{Incompatible-Subset}(\alpha, \eta, B)$

if $\alpha = T$ /* base */
    then return $\{ \beta_j \in B \mid \text{Ksat}(\eta \land \neg \beta_j) = \text{False} \};$

if $\alpha = F$ /* backtrack */
    then return $\emptyset;$

if {a unit clause (l) is in $\alpha$} /* unit */
    then return $\text{Incompatible-Subset}(\text{assign}(l, \alpha), \eta \land l, B);$;

$l := \text{choose-literal}(\alpha);$ /* split */

$B' := \text{Incompatible-Subset}(\text{assign}(l, \alpha), \eta \land l, B);$;

if $\text{Is-Empty}(B')$
    then return $B'$;
else return $\text{Incompatible-Subset}(\text{assign}(\neg l, \alpha), \eta \land \neg l, B');$
```

FIG. 2. $\text{Ksat}$ Version 3: a new schema for $\text{Ksat}_{RA}$. 
\( \eta_k \land \neg \beta_j \) is \( K(m) \)-satisfiable) are discharged from \( B \). This is iterated till \( B \) is empty (\( \mu' \) is \( K(m) \)-satisfiable) or no more assignments \( \eta_k \) can be found (\( \mu' \) is not \( K(m) \)-satisfiable).

In Fig. 2 we present a revised version \( KSAT_{RA} \). As before, \( KSAT_{RA} \) takes in input a restricted truth assignment \( \mu' = \land_i \land_j \neg \square_i \land \neg \square_j \beta_j \), and returns a truth value asserting whether \( \mu' \) is \( K(m) \)-satisfiable or not. If no conjunct of the form \( \neg \square_i \beta_j \) occurs in \( \mu' \), then \( \mu' \) is \( K(m) \)-satisfiable, and thus \( KSAT_{RA} \) returns \( True \). Otherwise \( KSAT_{RA} \) invokes the function \( \text{Incompatible-Subset} \), passing as arguments the wff \( \chi = \land_i \chi_i \), an empty assignment \( T \), and the (always non-empty) wff set \( B = \{ \beta_{r1}, \ldots, \beta_{rk} \} \). \( \text{Incompatible-Subset}(\chi, T, B) \) returns the set of the wffs \( \beta_j \)'s in \( B \) which are not compatible with any truth assignment which satisfies \( \chi \). \( KSAT_{RA} \) returns \( True \) if and only if this set is empty. Similarly to \( KSAT_{W} \), \( \text{Incompatible-Subset} \) tries to build truth assignments \( \eta \)'s satisfying \( \chi \). Whenever it finds one, all the wffs \( \beta_j \)'s which are compatible with this assignment are discharged from \( B \). Again, this is done recursively according to the following steps:

- (base) If \( \chi = T \), then \( KSAT \) is invoked on all wffs in the form \( \eta \land \neg \beta_j \), for all \( \beta_j \)'s in \( B \). The set of the wffs \( \beta_j \)'s in \( B \) which are not compatible with \( \eta \) (i.e., \( KSat(\eta \land \neg \beta_j) \) returns \( False \)) is then returned.

- (backtrack) If \( \chi = F \), then \( \eta \) does not satisfy \( \chi \). Therefore \( \text{Incompatible-Subset} \) returns the whole set \( B \).

- (unit) If a literal \( l \) occurs in \( \chi \) as a unit clause (or equivalent form for non-CNF wffs) then \( l \) is added to \( \eta \) and \( \text{Incompatible-Subset} \) is invoked recursively with \( \text{assign}(l, \chi) \), \( \eta \land l \), and \( B \).

- (split) If none of the above situations occur, then \( \text{choose-literal}(\chi) \) returns an unassigned literal \( l \). Then \( \text{Incompatible-Subset} \) is first invoked with \( \text{assign}(l, \chi) \), \( \eta \land l \), and \( B \), and the set returned is stored in \( B' \). If \( B' \) is empty, then no more incompatible wffs are available, and thus \( B' \) is returned. Otherwise, \( \text{Incompatible-Subset} \) is invoked with \( \text{assign}(\neg l, \chi) \), \( \eta \land \neg l \), and \( B' \).

The reader may recognize in \( \text{Incompatible-Subset} \) a variant of DPLL. The main difference is that \( \text{Incompatible-Subset} \) returns a set of wffs instead of a truth value. Notice that the split step is asymmetric, as the second call is invoked with the smaller set \( B' \). As in \( KSAT_{W} \), the base step is modified to use DPLL for generating truth assignments.

**Example 4.** Consider \( \varphi \), \( \mu' \), \( \varphi^{11} \), and \( \varphi^{12} \) as in Examples 1 and 2. We have

\[
\bigwedge_i \chi_i = (\neg A_4 \lor A_4 \lor A_3) \land (\neg A_2 \lor A_1 \lor A_4),
\]

\[\neg \beta_{11} = A_3 \land A_1 \land \neg A_2,\]

\[\neg \beta_{12} = \neg A_4 \land A_2 \land \neg A_3.\]

Suppose \( \text{Incompatible-Subset} \) selects in sequence the literals \( \neg A_4 \) and \( \neg A_2 \), finding thus the assignment \( \eta_1 = \neg A_4 \land \neg A_2 \) which satisfies \( \chi \). Then \( \eta_1 \land \neg \beta_{11} \) is satisfiable but \( \eta_1 \land \neg \beta_{12} \) is not. For the base step, \( \text{Incompatible-Subset} \) returns \( B' = \{ \beta_{12} \} \).
Then \textit{Incompatible-Subset} splits on the literal $\neg A_2$, finding $\eta_2 = \neg A_1 \land A_2 \land A_1$. As $\eta_2 \land \neg \beta_{12}$ is satisfiable, \textit{Incompatible-Subset} returns the empty set. Therefore $\mu^1$ is $K(2)$-satisfiable.  

3.4. Version 4: Intermediate Assignment Checking

Despite the improvement brought by internally sorting modal atoms, the number of truth assignments found by $\text{Ksat}_W$ may still be too large. Version 4 starts from the empirical observation that most assignments found by $\text{Ksat}_W$ are trivially $K(m)$-unsatisfiable; that is, they will remain $K(m)$-unsatisfiable even after removing some of their conjuncts. If an incomplete\footnote{By incomplete assignment $\mu$ for $\varphi$ we mean that $\mu$ neither satisfies $\varphi$ nor falsifies it.} assignment $\mu'$ is $K(m)$-unsatisfiable, then all its extensions are $K(m)$-unsatisfiable. If the unsatisfiability of $\mu'$ is detected on time, then this prevents checking the $K(m)$-satisfiability of all the up to $2^{|\text{TopAtom}(\varphi)| - |\varphi'|}$ truth assignments which extend $\mu'$.

This suggests the introduction of an intermediate $K(m)$-satisfiability test on incomplete assignments just before the split. (Notice there is no need to introduce similar tests before unit propagation.) This can be done by introducing the three lines below in the function $\text{Ksat}_W$ of Fig. 1, just before the “split”:

\begin{verbatim}
if Likely-Unsatisfiable(\mu) \n  \{ /* intermediate assignment check */ \n    if not Ksat_A(\mu) \n    then return False; \n\}
\end{verbatim}

Let us temporarily ignore the test performed by \textit{Likely-Unsatisfiable}. $\text{Ksat}_A$ is invoked on the current incomplete assignment $\mu$. If $\text{Ksat}_A(\mu)$ returns $\text{False}$, then all possible extensions of $\mu$ are unsatisfiable, and therefore $\text{Ksat}_W$ returns $\text{False}$.

Example 5. Consider the formula $\varphi$ of Example 1. Suppose that, after three recursive calls, $\text{Ksat}_W$ build the incomplete assignment:

\[
\mu' = \Box(\neg A_1 \lor A_4 \lor A_3) \land \Box(\neg A_2 \lor A_1 \lor A_4) \land \neg \Box(A_4 \lor \neg A_2 \lor A_3)
\]

(rows 6, 7, and 4 of $\varphi$). If it is invoked on $\mu'$, $\text{Ksat}_A$ will check the $K(2)$-satisfiability of the single formula

\[
(\neg A_1 \lor A_4 \lor A_3) \land (\neg A_2 \lor A_1 \lor A_4) \land \neg A_4 \land A_2 \land \neg A_3,
\]

which is unsatisfiable. Therefore there will be no more need to select further literals, and $\text{Ksat}_W$ will backtrack. 

It may be argued that the introduction of an intermediate consistency check before every split could negatively affect the global worst-case performance. (In fact, in a binary tree the number of splitting internal nodes equals the number of leaves minus one.) In the hypothetical case in which no intermediate test caused backtracking, the number of $\text{Ksat}_A$ calls per $\text{Ksat}$ call could double, and the global
number of Ksat calls might increase of up to $2^{\text{depth}(\phi)}$. To avoid this, it is worth introducing an heuristic function \textit{Likely-Unsatisfiable}. The idea is that \textit{Likely-Unsatisfiable} estimates the possibility of $\mu$ being $K(m)$-satisfiable according to parameter values like, e.g., the number of conjuncts and the number of propositional variables in $\mu$. For instance, a simple heuristic could be to perform an intermediate check whenever the last literal added to $\mu$ is not propositional. More sophisticated heuristics could use previously tabulated satisfiability transition diagrams, like those we will see in next sections. Notice that, to make the intermediate consistency check worth doing, an average pruning of one single split per branch is sufficient.

4. DPLL-BASED VS. TABLEAU-BASED DECISION PROCEDURES

We organize this section as follows. We start by comparing a direct implementation of the various versions of Ksat, as described in Section 4.1, with Tableau (unless explicitly stated to the contrary, when we write Ksat, we mean Ksat Version 4). This allows us to compare the basic algorithms without introducing distortions (i.e., speed-ups) due to smart implementation techniques. In Section 4.2 we compare Tableau, Kris, Ksat, and Ksat. This allows us to analyze the effects of implementation improvements on efficiency. We conclude by providing an explanation of the main results presented (Section 4.3). All the testing described in this section is restricted to the case of one modality. This allows us to consider the simplest situation without losing in generality; the phenomena highlighted in this section are in fact confirmed by the exhaustive testing presented in Section 5 (which studies, among other things, the effects of varying the number of modalities).

4.1. Comparing Algorithms

We have implemented Ksat in Common Lisp on top of a DPLL procedure for non-CNF wffs previously developed [AG93]. With respect to the algorithm described in Section 3, we have made the following implementation choices: the function \textit{assign} (see Figs. 1 and 2) performs a (linear time) lazy evaluation; the function \textit{choose-literal}(\phi) (see Figs. 1 and 2) performs the simple heuristic: “choose the variable with most occurrences inside $\phi$”; in the implementation of Version 4, \textit{Likely-Unsatisfiable} is trivially implemented to always return “\textit{True}”. Both Tableau and Ksat have been compiled on AKCL 1.600 and executed under SunOS 4.1.3.

To perform our tests, we have adopted the modal 3-clause-length test method, as described in [GRS96]. A particular kind of modal CNF wffs, called 3CNF$_{K(m)}$, is randomly generated according to five parameters: (i) the modal depth $d$; (ii) the number of distinct boxes $m$; (iii) the number of top-level clauses $L$; (iv) the number of propositional variables $N$; (v) the probability $p$ with which any random 3CNF$_{K(m)}$ atom is propositional. For fixed $N, d, m$, and $p$, for increasing values of $L$, many (100, 500, 1000...) random 3CNF$_{K(m)}$ wffs are generated, internally sorted, and then

---

\footnote{A 3CNF$_{K(m)}$ wff is a conjunction of 3CNF$_{K(m)}$ clauses; a 3CNF$_{K(m)}$ clause is a disjunction of three 3CNF$_{K(m)}$ literals, i.e., 3CNF$_{K(m)}$ atoms or their negations; a 3CNF$_{K(m)}$ atom is either a propositional atom or a wff of the form $\Box_{C}$, where $C$ is a 3CNF$_{K(m)}$ clause.}
given in input to the procedure under test. Satisfiability percentages, mean/median CPU times or mean/median search space sizes are plotted against the $L/N$ ratio.

We have compared Tableau and Ksat on six groups of random 3CNF$_{K(m)}$ wffs, labeled (i) to (vi), with $m = 1$, $p = 0.5$, $d = 1$, $N = 2, 3, 4$ (groups (i) to (iii)) and $m = 1$, $p = 0.5$, $d = 2$, $N = 1, 2, 3$ (groups (iv) to (vi)). Each group has been organized into 40 subgroups, each corresponding to an integer value of $L/N$ ranging from 1 to 40. (As will be clear in Section 5, this range has been chosen empirically to cover the transition between 100% satisfiability to 100% unsatisfiability). The necessity to perform the Tableau tests in a reasonable time (e.g., the single higher-most point of the Tableau curve (v) has required 137 hours of CPU time on a SPARC10 machine!) has imposed various constraints on the testing methodology. First, we have run our tests on six SUN stations: (a) one SPARC10 SUPERSPARC 32M, (b) two twin SPARC2 SUN4/75 32M, (c) two twin SPARC ELC SUN4/25 16M, (d) one SPARC SLC SUN4/20 16M. However, for every problem group, Tableau and Ksat have always been run under the same configuration. Second, we have not exceeded the number of 100 samples/point (giving a total of 4000 random wffs). Finally, in all the tests we have stopped the execution of Tableau whenever the mean CPU time on 100 samples has exceeded the bound of 1000 s. The results are reported in Fig. 3. As it can be seen, Ksat outperforms Tableau in all the problems considered (notice the logarithmic scale in the vertical axis). All the Tableau mean CPU time plots present an exponential growth against the number of clauses, while the Ksat mean CPU time plots grow much slower. The Tableau curves reach the time bound of 1000 s after very few steps, $10^3 - 10^4$ times above the corresponding Ksat curves. The extrapolation of the Tableau curves suggests that the gap may reach several orders of magnitude for problems near the right-end side of the plots.

![Fig. 3. Tableau vs Ksat: mean CPU times (s); 100 samples/point. (First row) Problem groups (i) to (iii): $p = 0.5$, $d = 1$, $N = 2, 3, 4$, $L/N = 1 \cdots 40$. (Second row) Problem groups (iv) to (vi): $p = 0.5$, $d = 2$, $N = 1, 2, 3$, $L/N = 1 \cdots 40$. The labels (a) \cdots (d) indicate the machine configuration used for the test.](image-url)
In order to get an overall and comparative evaluation we have compared Versions 2, 3, and 4 of \textsc{Ksat} and \textsc{Tableau} on problem group (v). (We could not consider Version 1 as the input wffs were sorted.) In this test we have also analyzed the behavior of \textsc{Modified Tableau}, that is, \textsc{Tableau} modified to exploit lemma generation \cite{DM94}. More specifically, \textsc{Modified Tableau} is \textsc{Tableau} with the $\vee$-rule $\frac{\varphi \vee \psi}{\varphi, \psi}$ substituted with the rules $\frac{\varphi \vee \psi}{\varphi, \psi}$, $\frac{\varphi \vee \psi}{\varphi, \psi}$. (The reasons for this choice will become clear in Section 4.3.) The results are reported in Fig. 4. As can be seen, already Version 2 outperforms both \textsc{Tableau} and \textsc{Modified Tableau}. For instance, \textsc{Tableau} exceeds the time bound after eight steps, about $10^{2\cdot10^{3}}$ times above the corresponding value of Version 2. \textsc{Modified Tableau} outperforms \textsc{Tableau} (e.g., a $10^{2}$ factor at the 8th step). Despite this, the improvement introduced does not overcome the performance gap with Version 2. Second, adding the factorization of $\frac{\wedge, \psi}{\varphi}$ (Version 3) introduces a further improvement (about an order of magnitude around the 13th step). Finally, the biggest improvement is obtained by adding the incomplete assignment checking (Version 4). Again, the extrapolation of the curves suggests that the performance gaps might increase up to various orders of magnitude.

4.2. Comparing Algorithms and Systems

There are many tricks which can make an implementation more efficient. \textsc{Ksat}, has been obtained from \textsc{Ksat} by adding an initial phase of wff preprocessing and other relatively minor implementation variations (which can be understood simply by comparing the code of the two systems).\footnote{The implementation of \textsc{Ksat}, is still naive in many respects; e.g., it is in Lisp and it does not use fancy optimized data structures. However, the improvements allow us to get an idea of the effects of the quality of implementation on efficiency.} All the systems, that is \textsc{Tableau}, \textsc{Kris}, \textsc{Ksat}, and \textsc{Ksat}, have been compiled and run under Allegro CL 4.2 on a SUN SPARC10 32M workstation. This has allowed us to use the builtin Allegro timeout mechanism.

We have compared \textsc{Tableau}, \textsc{Kris}, \textsc{Ksat}, and \textsc{Ksat}, on a testbed similar to group (vi) of Fig. 3, that is, 4000 3CNF$_{K(m)}$ random formulas with $d=2$, $m=1$, 

![FIG. 4. \textsc{Tableau}, \textsc{Modified Tableau}, and \textsc{Ksat} Versions 2, 3, 4: mean CPU times (s). $d=2$, $p=0.5$, $N=2$, $L/N=1\cdots40$, 100 samples/point.](image-url)
$N = 3$, $p = 0.5$, $L/N \in \{1 \cdots 40\}$, with 100 samples/point. We have introduced some further improvements in the testing technique, again in order to minimize the testing time. First, we have introduced a timeout of 1000 s on each sample wff; any time the decision procedure under test has exceeded the timeout, the CPU time value has been conventionally set to 1000 s. Second, we have stopped running the test on a point whenever more than 50 samples have exceeded the timeout. Third, we have compared median values rather than mean values, as the former are much less sensitive to the noise introduced by outliers (see, e.g., [MSL92]). The results are presented in Fig. 5 (left).

Four observations can be made, given below in increasing order of importance. First, improving the quality of the implementation, e.g., from Tableau to Kris or from Ksat to Ksat*, introduces good quantitative performance improvements. In fact, Kris reaches the time bound at the 10th step, while Tableau reaches the time bound at the 7th step, about two orders of magnitude above the corresponding Kris value. Similarly, Ksat has a maximum at the 14th step, more than two orders of magnitude above the corresponding Ksat value. However, and this is the second observation, improving the quality of the implementation does not seem to affect the qualitative behavior of the procedures. In fact, as far as we can see, both the Tableau and the Kris curves present an exponential growth with the number of clauses, while the Ksat and Ksat* curves flatten when the number of clauses exceeds a certain value. Third, independent from the quality of implementation, Ksat and Ksat* quantitatively outperform Tableau and Kris. For instance, the performance gap between Ksat and Kris at the 10th step is about four orders of magnitude. Moreover, the extrapolation of the Kris curve suggests that its value, and the performance gap with Ksat*, would reach several orders of magnitude for problems at the right-end side of the plots. To support this consideration, we have run Kris on 100 samples of the same problem, for $L/N = 40$. No sample wff has been solved within the timeout. When releasing the timeout mechanism, Kris has not been able to end successfully the computation of the first sample wff after a run of one month. Fourth, and most important, independent of the quality of implementation, Ksat and Ksat* qualitatively outperform Tableau and Kris. In fact, while Tableau and Kris present an exponential growth against the number of clauses, the Ksat and Ksat* curves present a polynomial growth (for more on this see Section 5).

**FIG. 5.** (Left) $d = 2$, $m = 1$, $N = 3$, $p = 0.5$, $L = N \cdots 40/N$. Tableau, Kris, Ksat, and Ksat*. Median CPU time, 100 samples/point. (Right) Tableau, Kris, Ksat*, and Ksat Version 2 CPU times for $\varphi^5$ formulas.
To provide further evidence of the performance gap between DPLL-based procedures and tableau-based procedures, we have performed another, quite different, test, based on the class of wffs \( \varphi^K_d \) presented in [HM92]. This is a class of \( K(1) \)-satisfiable wffs, with depth \( d \) and \( 2d + 1 \) propositional variables. These wffs are paradigmatic for modal \( K \), as every Kripke structure satisfying \( \varphi^K_d \) has at least \( 2^{d+1} - 1 \) distinct states, while \( |\varphi^K_d| \) is \( O(d^2) \). From the results in [HM92] we can reasonably assume a minimum exponential growth factor of \( 2^d \) for any ordinary algorithm based on Kripke semantics. We have run Tableau, Kris, KSAT\(_s\), and KSAT Version 2, again compiled and run under Allegro CL 4.2 on a SUN SPARC10 32M workstation, on these formulas, for increasing values of \( d \). The results are plotted in Fig. 5 (right). The Tableau, Kris, KSAT\(_s\), and KSAT Version 2 curves grow exponentially, approximatively as \( (16.0)^d \), \( (12.7)^d \), \( (2.6)^d \), and \( (2.4)^d \), respectively, exceeding 1000 s for \( d = 6 \), \( d = 7 \), \( d = 11 \) and \( d = 12 \), respectively. The slight difference between KSAT\(_s\) and KSAT Version 2 is due to the overhead introduced by the \( \wedge, \pi \), factorization, which is useless with these formulas. It is worth observing that the result of tracing the global number of truth assignments \( \mu \), recursively found by both KSAT\(_s\) and KSAT Version 2, has given exactly \( 2^{d+1} - 1 \) for every \( d \), that is the minimum number of Kripke states. KSAT\(_s\) and KSAT Version 2 have found no redundant truth assignments.

### 4.3. An Explanation

The speed-up obtained by improving the quality of the implementation was to be expected. What is much more interesting is the quantitative and qualitative performance gap between the tableau-based procedures and the DPLL-based procedures. Let us concentrate on the basic algorithms and compare Tableau and KSAT. Both procedures work (i) by enumerating truth assignments which propositionally satisfy the input wff \( \varphi \) and (ii) by recursively checking the \( K(m) \)-satisfiability of the assignments found. Both algorithms perform the latter step in the same way. The key difference is in the way KSAT and Tableau handle propositional inference. Tableau propositional decision procedures have, with respect to DPLL, two weaknesses which make them intrinsically less efficient and whose effects get (up to) exponentially amplified when using them in modal inference. Let us consider them in turn.

#### 4.3.1. Syntactic vs semantic branching

In a tableau propositional decision procedure truth assignments are (implicitly) generated as branches of an analytic propositional tableau. Analytic propositional tableaux perform what we call syntactic branching, that is, a branching on the syntactic structure of \( \varphi \). In particular, as discussed in [DM94], an application of the \( \lor \)-rule (see Section 4.1) generates two subtrees which are not mutually inconsistent. The number of truth assignments generated grows exponentially with the number of disjunctions occurring positively in \( \varphi \) (in our tests, the number of clauses \( L \)). Therefore, the set of truth assignments enumerated by propositional tableau procedures is intrinsically redundant and may contain (up to exponentially many) duplicated and/or subsumed assignments.
Things get much worse in the modal case. When testing $K(m)$-satisfiability, unlike the propositional case where tableaux look for one assignment satisfying the input formula, the propositional tableaux are used to enumerate all the truth assignments, which must be recursively checked for $K(m)$-consistency. Notice that the number of assignments can be huge: up to many thousands in our tests. This requires checking recursively (possibly many) sub-wffs of the form $\land_i \alpha_i \land \beta_i$ of depth $d-1$, for which a propositional tableau will enumerate truth assignments, and so forth. Any redundant truth assignment enumerated at depth $d$ introduces a redundant modal search tree of depth $d$. Even worse, this propositional redundancy propagates up to exponentially with the depth $d$, following the analysis of the sub-wffs of decreasing depth.

In DPLL-based procedures, instead, truth assignments are generated one-shot by DPLL. DPLL-based procedures perform a search based on what we call semantic branching, that is, a branching on the truth value of the proper sub-wffs of $\varphi$. Every branching step generates two mutually inconsistent subtrees. Because of this, DPLL always generates nonredundant sets of assignments. This avoids any search duplication and, in the case of modal search, any recursive exponential propagation of inefficiency.

**Example 6.** Consider the simple wff $\varphi = (\alpha \lor \neg \beta) \land (\alpha \lor \beta) \land (\neg \alpha \lor \neg \beta)$, where $\alpha$ and $\beta$ are modal atoms, and let $d$ be the depth of $\varphi$. The only possible assignment satisfying $\varphi$ is $\mu = \alpha \land \neg \beta$. The $\lor$-rule is applied to the three clauses occurring in $\varphi$ in the order they are listed, and two distinct but identical open branches are generated, both representing the assignment $\mu$. Suppose now that $\mu$ is not $K(m)$-consistent. Then the tableau expands the two open branches in the same way, until it generates two identical (and possibly big) closed modal subtableaux $T$ of depth $d$, each proving the $K(m)$-unsatisfiability of $\mu$. This phenomenon may repeat itself at the lower level in each subtableaux $T$, and so forth. For instance, if $\alpha = \Box_1 (\alpha' \lor \neg \beta')$ and $\beta = \Box_1 (\alpha' \land \beta')$, then at the lower level we have a wff $\varphi'$ of depth $d-1$ analogous to $\varphi$. This propagates exponentially the redundancy with the depth $d$. Finally, notice that if we considered the wff $\varphi^K = \land_i \alpha_i \land \neg \beta_i$, the tableau would generate $2^k$ identical truth assignments $\mu^K = \land_i \alpha_i \land \neg \beta_i$, and things would get exponentially worse.

A DPLL-based procedure, instead, branches with $\alpha = \text{True}$ or $\alpha = \text{False}$. The first branch generates $\alpha \land \neg \beta$, while the second gives $\neg \alpha \land \neg \beta \land \beta$, which immediately closes. Therefore, only one instance of $\mu = \alpha \land \neg \beta$ is generated. The same applies recursively to $\mu^K$.

**4.3.2. Detecting constraint violations.** The explanation above leaves two questions unanswered. First, Fig. 4 shows a performance gap between Version 2 and Modified Tableau (more than a factor 30 for $L/N = 12$). This fact is still not explained, as Modified Tableau generates mutually inconsistent subtrees (see Section 4.1). Second, Fig. 5 (left) (but see also Fig. 6 in Section 5) shows that, for $L/N$ bigger than a certain value, the median time curves of KSAT decreases with the size of the formulas.

A propositional wff $\varphi$ can be seen as a set of constraints for the truth assignments which possibly satisfy it (see, e.g., [WH94]). For instance, a clause $A_1 \lor A_2$
constrains every assignment not to set both $A_1$ and $A_2$ to $F$. Unlike tableau, DPLL prunes branches as soon as they violate some constraint of the input wff. The more constrained the input wff is, the more likely a truth assignment violates some constraint. (For instance, the bigger is $L$ in a CNF wff, the more likely an assignment generates an empty clause.) Therefore, as $\varphi$ becomes highly constrained (e.g., when $L$ is big enough) the search tree is very heavily pruned. As a consequence, for $L$ bigger than a certain value, the size of the search tree decreases with $L$.

**Example 7.** Consider the wff $\varphi = (x \lor \varphi_1) \land (y \lor \varphi_2) \land \varphi_3 \land (\neg x \lor \neg y)$, being $x$ and $y$ atoms, $\varphi_1$, $\varphi_2$, and $\varphi_3$ sub-wffs, such that $x \land y \land \varphi_3$ is propositionally satisfiable. Again, assume the $\lor$-rule is applied in order, left to right. After two steps, the branch $x$, $y$ is generated, which violates the constraint imposed by the last clause $(\neg x \lor \neg y)$. A tableau-based procedure is not able to detect such a violation until it explicitly branches on that clause, that is, only after having generated the whole sub-tableau $T'$ for $x \land y \land \varphi_3$, which may be rather big. DPLL, instead, detects the violation and immediately prunes the branch. For instance, in KSAT this is done by the function assign.

5. KSAT ON $K(m)$

We have tested KSAT, on a total of 48,000 random 3CNF $K(m)$ wffs, organized in three experiments. As above, we have computed 100 samples/point and chosen the range $[1 \cdots 40]$ for $L/N$ to cover the “100% satisfiable–100% unsatisfiable” transition. The results are reported in Fig. 6. In each experiment we investigate the effects of varying one parameter while fixing the others. In Experiment 1 (left column) we fix $d = 2$, $m = 1$, $p = 0.5$ and plot different curves for increasing numbers of the variables $N = 3, 4, 5$. In Experiment 2 (center column) we fix $d = 2$, $N = 4$, $p = 0.5$ and plot different curves for increasing numbers of distinct modalities $m = 1, 2, 5, 10, 20$. In Experiment 3 (right column) we fix $m = 1$, $N = 3$, $p = 0.5$ and plot different curves for increasing modal depths $d = 2, 3, 4, 5$. For each experiment, we present three sets of curves, each corresponding to a distinct row. In the first (top row) we plot the percentage of satisfiable wffs evaluated by KSAT. This gives a coarse indication of the average level of constrainedness of the test wffs. In the second (middle row) we plot the median CPU time obtained by running both KSAT. This gives an overall picture of the KSAT, qualitative behavior. In the third (bottom row) we plot the KSAT, median number of recursive DPLL calls, that is, the size of the space effectively searched by KSAT. It can be noticed that the figures in the middle row report also the KRIS performance curves. These curves have been reported in order to back up the results and analysis in Section 4. As can be noticed, KSAT outperforms KRIS in all the testbeds, independently on the number of variables $N$, the number of modalities $m$, or the depth $d$ considered. Again, this is a quantitative and qualitative performance gap.
FIG. 6. The results of the three experiments.

space. Each variable may in fact assume distinct truth values inside distinct states/possible worlds, that is, each variable must be considered with an implicit multiplicity equal to the number of states of a potential Kripke model. The results of the second experiment (center column) present two interesting aspects. First, the complexity of the search monotonically decreases with the increase of the number \( m \) of modalities (middle and bottom box). At first sight it may sound like a surprise, but it should not be so. In fact, each truth assignment \( \mu \) is partitioned into \( m \) independent subassignments \( \mu_r \)'s, each restricted to a single \( \Box \), (see Eqs. (1) and (2)). This means dividing and conquering the search tree into \( m \) noninterfering search trees. Therefore, the bigger \( m \) is, the more partitioned the search space is, and the easier the problem is to solve. Second, a careful look reveals that the satisfiability percentage increases with \( m \). Again, there is no mutual dependency between the satisfiability of the distinct \( \mu_r \)'s. Therefore the bigger \( m \) is, the less constrained \( \mu \) is, and the more likely satisfiable \( \varphi \) is. The results of the third experiment (right column) provide evidence of the fact that the complexity increases with the modal depth \( d \). This is rather intuitive: the higher \( d \) is, the deeper the Kripke models to be searched are, and the higher the complexity of the search is.
Giving a global look at the middle and bottom rows it can be noticed that these curves show the existence of an easy-hard-easy component plus a linear component. The former represents the number of recursive DPLL calls, i.e., the size of the tree effectively searched, while the latter is due to the preprocessing and to the linear-time function assign, which is invoked at every DPLL recursive call. In fact, if we increase $N$ from 3 to 5 and $L$ accordingly (left column), the size of the search space has a relevant increase. Therefore, while for $N=3$ the linear component prevails, for $N=5$ the easy-hard-easy component dominates. Moreover, when varying the number of modalities (center column), the wff sizes are kept the same for all curves. Therefore, when the effect of the easy-hard-easy component vanishes ($L/N > 20$), the curves collapse together, as the linear component does not depend on the number of modalities $m$. Notice that the locations of the easy-hard-easy zones do not seem to vary significantly, with the number of variables $N$ (left column), with the number of modalities $m$ (center column), or with the depth $d$ (right column). Let us now consider the satisfiability plots (top row). Despite the noise and the approximations due to timeouts, it is easy to notice that the 50% satisfiability point is centered around $L/N = 15 \sim 20$ in all the experiments. Moreover, in the first experiment a careful look reveals that the satisfiability transition becomes steeper when increasing $N$ (e.g., compare the $N=3$ and $N=5$ plots). Finally, in all experiments, the curves representing the median number of DPLL calls (top row) locate the peaks inside the satisfiability transition, although they seem to anticipate a little the 50% crossover point.

From the above considerations we may conjecture (to be verified!) the existence for $K(m)$ of a phase transition phenomenon, similar to that already known for SAT and other NP-hard problems (see, e.g, [MSL92, WH94]). As far as we know, this is the first time this phenomenon is revealed with modal formulas. This conjecture is also backed up by the analysis given in Section 4.3 (discussion on pruning unsatisfiable assignments). In fact, if we add a new clause to an unsatisfiable formula, this causes extra pruning in the search. Therefore, when we are in the 100% unsatisfiable zone, the search space size decreases with $L/N$. On the other hand, if we drop a clause from a satisfiable wff, this monotonically increases the number of satisfiable branches and thus decreases the size of the search space visited to find a satisfiable branch. Therefore, when we are in the 100% satisfiable zone, the search space size increases with $L/N$. These two facts locate the peak inside the satisfiability transition zone. It is interesting to notice that, instead, the semantic branching has the effect of pushing down the performance curves for any value of $L/N$. In fact, no matter which satisfiability zone we are in, it always decreases the number of assignments considered.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed and tested $Ksat$, an algorithm built on top of DPLL which tests $K(m)$-satisfiability. $Ksat$ outperforms quantitatively and qualitatively the previous state-of-the-art tableau-based modal decision procedures and has allowed us, among other things, to reveal what looks like a phase transition phenomenon for $K(m)$. 
This work is only an instance of the more general idea that propositional modal decision procedures should be developed on top of propositional decision procedures. There are clearly many directions for future research. An interesting issue is to what extent the kind of propositional reasoning used, and its efficiency, influence the efficiency of modal reasoning. For instance, we conjecture that all the modal procedures developed on top of propositional procedures which perform semantic branching and pruning will show a phase transition phenomenon. Toward this direction we have already acquired and preliminarily tested what is considered one of the fastest implementation of DPLL, i.e., Max Böhm's system [BB92], and are in the process of acquiring an implementation of OBDDs.

The other obvious direction of research is the extension to other modal logics. We plan to consider other normal, nonnormal modal, temporal and dynamic logics. It is our conjecture that this technique will be even more successful when applied to nonnormal modal logics.

Received KKK; published online August 4, 2000

REFERENCES


