Exploiting Resolution Proofs to Speed Up LTL Vacuity Detection for BMC

Jocelyn Simmonds*, Jessica Davies*, Arie Gurfinkel† and Marsha Chechik*
*University of Toronto, Toronto, ON M5S 3G4, Canada.
Email: {jsimmond,jdavies,chechik}@cs.toronto.edu
†SEI at Carnegie Mellon University, Pittsburgh, PA 15213-2612, USA.
Email: arie@sei.cmu.edu

***Abstract*—When model-checking reports that a property holds on a model, *vacuity detection* increases user confidence in this result by checking that the property is satisfied in the intended way. While vacuity detection is effective, it is a relatively expensive technique requiring many additional model-checking runs. We address the problem of efficient vacuity detection for Bounded Model Checking (BMC) of LTL properties, presenting three partial vacuity detection methods based on the efficient analysis of the resolution proof produced by a successful BMC run. In particular, we define a characteristic of resolution proofs – *peripherality* – and prove that if a variable is a source of vacuity, then there exists a resolution proof in which this variable is peripheral. Our vacuity detection tool, *VaqTree*, uses these methods to detect vacuous variables, decreasing the total number of model-checking runs required to detect all sources of vacuity.**

I. Introduction

Model-checking [1] is a widely-used automated technique for verification of both hardware and software artifacts that checks whether a temporal logic property is satisfied by a finite-state model of the artifact. If the model does not satisfy the property, a counterexample, which can aid in debugging, is produced. If the model *does* satisfy the property, no information about why it does so is provided by the model-checker alone. A positive answer without any additional information can be misleading, since a property may be satisfied in a way that was not intended. For instance, a property "every request is eventually acknowledged" is satisfied in an environment that never generates requests.

Vacuity detection [2]–[5] is an automatic sanity check that can be applied after a positive model-checking run in order to gain confidence that the model and the property capture the desired behaviours. Informally, a property is said to be vacuous if it has a subformula which is not relevant to its satisfaction, or if the property itself is a tautology. Conversely, a property is satisfied non-vacuously if every part of the formula is important – even a slight change to the formula affects its satisfaction.

In this paper, we focus on vacuity detection for SAT-based Bounded Model Checking (BMC). Given a BMC problem with a particular bound, we wish to determine if the property holds vacuously on the model up to this bound. In this context, a naive method for detecting vacuity is to replace subformulas of the temporal logic property with unconstrained boolean variables and run BMC for each such substitution. If the property with some substitution still holds on the model, the property is vacuous. This naive approach is expensive, since in the worst case it requires as many model-checking runs as there are subformulas in the property. Our goal is to reduce the number of model-checking runs required to detect vacuity. We do this by detecting some vacuity through novel and inexpensive techniques reported in this paper, and complete the method by running the naive algorithm on the remaining atomic subformulas. The key to our technique is that SAT-based BMC can automatically provide useful information (a resolution proof) beyond a decision whether the property holds on the model; we exploit such proofs for partial vacuity detection.

In SAT-based BMC, the property and the behavior of the model are encoded in a propositional theory, such that the theory is satisfiable if and only if the formula does not hold. When the property does hold, a DPLL-based SAT solver can produce a resolution proof that derives false from a subset of the clauses in the theory called the UNSAT core. Intuitively, the resolution proof provides an explanation why the property is not falsified by the model, and the UNSAT core determines the relevant parts of the model and the property [6].

In this paper, we develop three methods of increasing precision (*irrelevance*, *local irrelevance* and *peripherality*) to analyze the resolution proof to achieve partial vacuity detection. These algorithms are used by our vacuity detection tool, ***VaqTree***, in order to reduce the number of model-checking runs required to find all sources of vacuity, thus reducing execution times. Irrelevance and local irrelevance detect vacuity based on which variables appear in the UNSAT core, and in which locations. However, as these methods only examine the UNSAT core, their precision is limited. The peripherality algorithm examines the *structure* of the resolution proof, identifying as vacuous those variables that are not necessary or central to the derivation of false. This method is as precise as can be achieved through analyzing a single resolution proof, and its running time is linear in the size of the resolution proof and the number of variables in the property. Our experience shows that local irrelevance is the ideal candidate for replacing naive vacuity.

The remainder of the paper is organized as follows. Sec. II presents some required background, followed, in Sec. III by our definition of vacuity, the naive algorithm for LTL vacuity detection using BMC, and an overview of work in the vacuity detection field. Sec. IV presents the three algorithms that detect vacuity by analyzing a resolution proof. Our experimental

results are presented in Sec. V. We conclude with a summary, additional related work, and suggestions for future work in Sec. VI.

II. Background

In this section, we review bounded model-checking and resolution proofs.

A. Bounded Model-Checking

Bounded model-checking (BMC) [7] is a method for determining whether a linear temporal logic (LTL) formula φ holds on a finite state system represented by a Kripke structure K up to a finite number of steps. An instance of a BMC problem, denoted by $BMC_k(K, \varphi)$, is whether $K \models_k \varphi$, where $\models_k$ is the *k-depth satisfaction relation*. An informal description of LTL formulas, Kripke structures and BMC is given in [8], and detailed definitions can be found in [1], [7].

To determine whether $K \models_k \varphi$, the problem is converted to a propositional formula Φ (see [7], [9], [10]) which is satisfiable if and only if there exists a length-k counterexample to $K \models_k \varphi$. Φ is then given to a SAT solver which decides its satisfiability. The propositional encoding represents the behavior of K up to k steps with a *path constraint* CL_K, and encodes all counterexamples to φ of length k in an *error constraint* CL_e. Therefore, if the theory $CL_K \cup CL_e$ is satisfiable, there is a path through K which obeys the transition relation and falsifies φ. The value of each variable v of K at each time step is represented using new boolean variables v_i ($0 \leq i \leq k$), called *timed variables*.

The transition relation can be represented symbolically by a propositional formula over the variables V and primed variables V' (which represent the variables in the next state). For example, in the model in Fig. 1(a), the transition relation is represented by the formula $R = (p \wedge \neg q \wedge \neg p' \wedge q') \vee (\neg p \wedge q \wedge \neg p' \wedge q')$. The path constraint is obtained by substituting the timed variables V_i for V in R, and replacing V' by the timed variables for the next step, V_{i+1}. This is repeated for each $0 \leq i < k$, and the resulting propositional formulas are conjoined along with a formula representing the initial state [7]. In Fig. 1(a), if $k = 1$,

$$CL_K = (p_0 \wedge \neg q_0) \wedge ((p_0 \wedge \neg q_0 \wedge \neg p_1 \wedge q_1) \vee (\neg p_0 \wedge q_0 \wedge \neg p_1 \wedge q_1))$$

CL_e is encoded according to a recursive procedure which removes the temporal and logical operators from the property [7], e.g., the algorithm encodes $\varphi = \mathbf{G}p$, where p is a propositional variable, expanded up to k = 2, by the formula $\neg p_0 \vee \neg p_1 \vee \neg p_2$.

After the boolean formulas for the path and error constraints are calculated, they are converted to *Conjunctive Normal Form* (CNF) before being passed to a SAT solver. If the solver reports that $CL_K \cup CL_e$ is unsatisfiable, it means that there is no length-k counterexample to φ; otherwise, a satisfying assignment is returned. When a DPLL-based SAT solver processes an unsatisfiable theory, a resolution derivation of false (or the empty clause) is implicitly constructed [11], [12]. This resolution proof is used to verify that false can indeed be derived from $CL_K \cup CL_e$ [13].

B. Resolution Proofs

Resolution is an inference rule that is applied to propositional clauses to produce logical consequences. A *clause* is a disjunction of boolean variables and their negations. For example, $(v_1 \vee \neg v_2 \vee v_5)$ is a clause stating that at least one of $v_1, \neg v_2$ or v_5 must be true. The resolution rule takes two clauses, where one contains a variable v and the other – its negation $\neg v$, and produces a clause containing the union of the two clauses minus v and $\neg v$. For example, resolving $(v_1 \vee \neg v_2 \vee v_5)$ and $(v_2 \vee v_6)$ produces the *resolvent* $(v_1 \vee v_5 \vee v_6)$.

A *resolution proof* Π is a directed acyclic graph whose nodes are labeled by propositional clauses. Π represents a tree of resolutions between the clauses labeling its nodes. Its *roots* are the nodes with no parents; otherwise, all nodes have exactly two parents. The nodes with no children are called the *leaves*. For example, the roots of resolution proof Π in Fig. 1(b) are $Roots(\Pi) = \{(\neg r_0), (r_0 \vee p_0), (\neg p_0 \vee q_0), (\neg p_0 \vee \neg q_0), (p_0)\}$, and the leaf of Π is the empty clause,i.e., $Leaf(\Pi) =$ false. Given a non-root node labeled by the clause c, and the labels of its parents, c_1 and c_2, c is the resolvent since it has been produced by resolving c_1 and c_2 on some variable v. A resolution proof Π is a *proof of unsatisfiability* of a set of clauses A if and only if all roots of Π belong to A, and one of the leaves of Π is the empty clause. For example, Fig. 1(b) shows a resolution proof of the unsatisfiability of $Roots(\Pi)$. If a propositional theory in CNF is unsatisfiable, an *UNSAT core* is an unsatisfiable subset of its clauses.

Given two disjoint sets of clauses A and B, a variable v is said to be *local* to A if and only if v appears in A but does not appear in B, and v is said to be *global* if it appears in both A and B. In Fig. 1(b), if $Roots(\Pi) = A \cup B$, where $A = \{(\neg r_0), (r_0 \vee p_0), (\neg p_0 \vee q_0)\}$ and $B = \{(\neg p_0 \vee \neg q_0), (p_0)\}$, then r_0 is local to A, and the rest are global.

III. Defining Vacuity

This paper uses the following definition of vacuity.

Definition 1 *Let K be a Kripke structure, φ be a formula satisfied by K (i.e., $K \models \varphi$), and p be a variable. Then, φ is p-vacuous in K iff $\varphi[p \leftarrow x]$ is satisfied by K, where x is a variable not occurring in K or in φ.*

We use $\varphi[p \leftarrow x]$ to indicate that all occurrences of p in φ are replaced by x.

Similarily, it is possible to define vacuity in the BMC setting.

Definition 2 *Let K be a Kripke structure, φ be a formula s.t. $K \models_k \varphi$, and p be a variable. φ is k-step p-vacuous iff $K \models_k \varphi[p \leftarrow x]$, where x is a variable not occurring in K or in φ.*

If φ is k-step p-vacuous, we call p a *k-step vacuous variable*. A property φ is *k-step vacuous* if and only if φ contains a k-step vacuous variable. Therefore, our techniques aim to find the k-step vacuous variables of φ. The qualifier "k-step" is omitted in the remainder of the paper but should be understood implicitly in the BMC context.

In the remainder of the paper, we avoid referring to k-vacuity, focusing instead on those variables p that are used to prove that a property is k-vacuous. When we say that a property φ is p-vacuous in $BMC_k(K, \varphi)$, it means that φ is k-vacuous, and p is

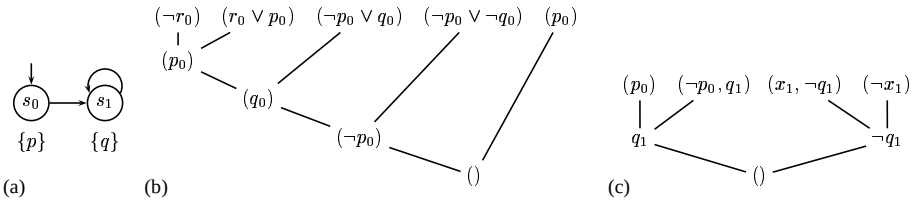

Fig. 1. (a) A Kripke structure; (b) A resolution proof for EXAMPLE 2; (c) A resolution proof for EXAMPLE 1.

such that $K \models_k \varphi[p \leftarrow x]$, where x is a new unconstrained variable of K.

Def. 1 suggests a sound and complete algorithm for vacuity detection: for each propositional variable p in φ, run BMC on $\varphi[p \leftarrow x]$, where x is a variable that does not appear in K and φ. If $K \models_k \varphi[p \leftarrow x]$ for some p, then φ is k-step vacuous. We refer to this algorithm as *naive*. Its drawback is that it may require as many model-checking runs as there are propositional variables in φ. Defs. 1 and 2 can be generalized to vacuity in arbitrary (not necessarily atomic) subformulas. This follows from the fact that a subformula is vacuous iff it is *mutually vacuous* in all of its atomic propositions [14, Th. 9], and that the definitions can be easily extended to mutual vacuity. For example, if φ contains subformula $\theta = p \wedge q$, and p and q are mutually vacuous, then we can deduce that θ is vacuous as well.

We now review some of the alternative definitions of vacuity and their algorithms. The first attempt to formulate and automate vacuity detection is due to Beer et al. [2]. They consider a property φ to be vacuous if φ contains a subformula ψ such that replacing ψ by any other formula does not affect the satisfaction of φ. Applying this definition directly would require an infinite number of subformula replacements, precluding a practical implementation. However, Beer et al. show that to detect vacuity w.r.t. a *single* occurrence of a subformula ψ in w-ACTL, it is sufficient to replace ψ with only true and false. This was later extended to CTL* by Kupferman and Vardi [3]. Purandare and Somenzi [4] showed how to speed up subformula vacuity by analyzing the parse tree of a CTL property.

Armoni et al. [5] generalized the above syntactic definition of vacuity by introducing universal quantification, i.e., $\forall x \cdot \varphi[\psi \leftarrow x]$. Based on the domain of x, three notions of vacuity are obtained, the most robust of which being *trace* vacuity. Gurfinkel and Chechik [15] extended Armoni's definition of vacuity to CTL*, thus uniformly capturing CTL and LTL. Armoni et al. also analyzed the syntactic structure of the property in order to avoid checking the operands of subformulas that are known to be vacuous. Such optimizations complement our techniques, which focus on detecting vacuous *atomic* subformulas.

Namjoshi [16] defines a somewhat different notion of vacuity, also based on a proof derived from a successful model checking run. According to Namjoshi, a property should only be considered vacuous if every proof of why it holds on the model exhibits vacuity. This definition of vacuity coincides with the definition of [5], [15] for a subset of LTL. Our methods efficiently examine proofs derived from model-checking runs, but are able to detect vacuity as defined by [2], [5], [15], [17]. Finally, we cannot empirically compare our techniques, since no experimental results are provided in [16].

Our definition of vacuity is syntactic, and in this respect, it is similar to the original definition of Beer et al. [2]. However, Def. 1 is stronger, and is equivalent to the semantic definition of Armoni et al. [5], as shown by Gurfinkel and Chechik [15].

IV. EXPLOITING RESOLUTION PROOFS

In Sec. III, we discussed the existence of a sound and complete vacuity detection algorithm for BMC, which requires as many model-checking runs as there are propositional variables in the property being checked. We propose a new vacuity detection strategy: first detect partial vacuity using inexpensive techniques and then complete the analysis using extra model-checking runs. Since we are interested in replacing expensive model-checking runs by inexpensive partial vacuity detection methods, we limit ourselves to considering the output of the original model-checking run on $BMC_k(K, \varphi)$, i.e.,$CL_K \cup CL_e$. This run provides us with a single resolution proof to analyze, but in general, there may be many ways to derive the empty clause from different subsets of $BMC_k(K, \varphi)$. Any method that only examines one of these derivations is inherently incomplete, in the sense that a property may be p-vacuous but there is no way of determining this based on a given resolution proof. For example, consider a model that is composed of two completely disjoint sub-models, running in parallel, i.e., $K = K_1 \parallel K_2$. Suppose that K_1 satisfies $\mathbf{G}p$, K_2 satisfies $\mathbf{G}q$, and that both do so non-vacuously. Then the property $\varphi = \mathbf{G}p \vee \mathbf{G}q$ holds on K p-vacuously and q-vacuously. However, one of the possible resolution proofs showing that φ holds proves that $\mathbf{G}p$ holds non-vacuously on K_1. Thus, it is impossible to determine that φ is vacuous in p from this proof. Any method based on examining only one resolution proof cannot prove the absence of vacuity, since another resolution proof, showing the property to be vacuous, might exist.

In this section, we introduce three algorithms of increasing precision for partial vacuity detection, based on examining the UNSAT core (irrelevance and local irrelevance) and the resolution proof produced by BMC (peripherality).

A. Examining UNSAT cores

Given a resolution proof that $BMC_k(K, \varphi)$ is unsatisfiable, we can sometimes cheaply determine that the similar theory $BMC_k(K, \varphi[p \leftarrow x])$ is also unsatisfiable, and therefore, that the property is p-vacuous. In this section, we consider how to determine that $BMC_k(K, \varphi[p \leftarrow x])$ is unsatisfiable given that $BMC_k(K, \varphi)$ is unsatisfiable, using only an UNSAT core.

1) Irrelevance: Intuitively, any variable that does not appear in the UNSAT core does not contribute to the reason why φ holds on K, so it can be considered *irrelevant*.

Definition 3 *Let K be a model, and φ an LTL formula. Assume that Π is an UNSAT core of BMC$_k$(K, φ) witnessing that $K \models_k \varphi$. Then, p is* irrelevant *with respect to BMC$_k$(K, φ) and Π iff p_i does not appear in Π for any time instance i.*

If a variable is irrelevant, it is also vacuous, as shown by the following theorem. The proofs of this and other theorems are given in Appendix A.

Theorem 1 *If p is irrelevant with respect to BMC$_k$(K, φ) and Π, then φ is k-step p-vacuous.*

Def. 3 provides an algorithm to detect some vacuous variables. However, a variable can appear in the UNSAT core and still be vacuous, as demonstrated by the following example.

EXAMPLE 1. Consider a Kripke structure K with variables p and q given by the constraints $Init = p \wedge q$, $R = p \Rightarrow q'$, which mean that the initial state is labeled by $\{p, q\}$, and the transition relation is expressed by the propositional formula $p \Rightarrow q'$ over unprimed and primed variables. Let $\varphi = X(p \vee q)$ be the property to check. φ is p-vacuous since it is satisfied simply because q is true in any successor of the initial state. The CNF encoding of the one-step BMC problem is $CL_K = \{(p_0 \wedge q_0), (p_0 \Rightarrow q_1)\} = \{(p_0), (q_0), (\neg p_0, q_1)\}$, $CL_e = \{(\neg p_1), (p_1, \neg q_1)\}$. In this case, the *unique* minimal UNSAT core contains all of the clauses of the problem except for (q_0). Thus, all p_i appear in the UNSAT core, and p cannot be determined vacuous using irrelevance. □

This example shows that even if we are to look at every UNSAT core of a BMC problem, irrelevance is still unable to detect existing vacuity.

2) Local Irrelevance: Variables which do not appear in the UNSAT core are vacuous. The converse is not true: vacuous variables may also appear in the UNSAT core. Intuitively, these variables are not the central reason why φ holds on K. For example, the clauses of CL_K may resolve against each other, representing some simplification and unification of parts of the model, before resolutions with CL_e clauses are performed. If a variable is resolved upon using only the CL_K clauses or only the CL_e clauses, it is potentially vacuous. By looking at the UNSAT core, it is possible to anticipate whether a variable will not be involved in resolutions between CL_K and CL_e using the following definition.

Definition 4 *Let K be a model, and φ an LTL formula. Assume that Π is an UNSAT core of BMC$_k$(K, φ) witnessing $K \models_k \varphi$. Then, p is* locally irrelevant *with respect to BMC$_k$(K, φ) and Π iff for each time instance i, either p_i does not appear in Π or p_i is local to either $CL_e \cap \Pi$ or $CL_K \cap \Pi$.*

In Example 1, p is locally irrelevant since p_1 only occurs in the clauses of U taken from CL_e, while p_0 only appears in U within CL_K clauses. Moreover, the UNSAT core of the original problem can be converted to an UNSAT core of the new theory, thus proving that p is vacuous. Specifically, $U = \{(p_0), (\neg p_0, q_1), (\neg p_1), (p_1, \neg q_1)\}$ is the UNSAT core of the original problem, so substituting x for p in the clauses of U that came from CL_e gives $U' = \{(p_0), (\neg p_0, q_1), (\neg x_1), (x_1, \neg q_1)\}$. This is a subset of BMC$_1$(K, $\varphi[p \leftarrow x]$) $= \{(p_0), (q_0), (\neg p_0, q_1), (\neg x_1), (x_1, \neg q_1)\}$, so it is a candidate for the new UNSAT core. The substitution may have prevented the resolutions necessary to derive the empty clause. However, Fig. 1(c) shows a proof that U' is also unsatisfiable. In this case, it was possible to substitute x_i for p_i in the clauses coming from CL_e in the original UNSAT core and create an UNSAT core for BMC$_k$(K, $\varphi[p \leftarrow x]$). In fact, this observation applies to all cases of local irrelevance by Theorem 2. Therefore, Def. 4 specifies an algorithm to detect some vacuous variables.

Theorem 2 *If p is locally irrelevant with respect to BMC$_k$(K, φ) and Π, then φ is k-step p-vacuous.*

Unfortunately, if a variable p is not locally irrelevant in an UNSAT core, the formula can still be p-vacuous, as shown by the following example.

EXAMPLE 2. Consider a Kripke structure with atomic propositions r, p and q whose initial state is given by the constraint: $Init = \neg r \wedge p \wedge q$. The formula $\varphi = \neg p \vee q$ is p-vacuous in the initial state. Let us assume that the zero-step BMC problem is encoded in CNF as follows:

$$CL_K = (\neg r_0)(r_0 \vee p_0)(\neg p_0 \vee q_0)$$
$$CL_e = (p_0)(\neg p_0 \vee \neg q_0)$$

There are several resolution proofs that can establish unsatisfiability of $CL_K \cup CL_e$; one such proof is shown in Fig. 1(b). In none of the proofs is p locally irrelevant with respect to CL_e and CL_K. □

The problem with local irrelevance is that it is impossible to tell if a variable is going to be used in a resolution joining CL_K and CL_e clauses based on the UNSAT core alone.

B. Peripherality

In Sec. IV-A, two vacuity detection methods based on examining the variables in the UNSAT core were found to fall short of completeness. It was seen that even if every possible resolution proof could be analyzed, irrelevance and local irrelevance still might fail to detect existing vacuity. Here, we extend the analysis to the resolution proof's structure. The resulting peripherality algorithm is superior, since it guarantees vacuity will be found if all possible resolution proofs are considered.

The limitations of detecting vacuity based only on the UNSAT core were demonstrated in Example 2. By examining the resolution proof in Fig. 1(b), we see that although p_0 appears both in CL_K clauses and in CL_e clauses, it is always resolved "locally". That is, if we resolve two clauses $c_1 = (..., p_i, ...)$ and $c_2 = (..., \neg p_i, ...)$, p_i and $\neg p_i$ must have been preserved from their original source in some set of root clauses. If all the originating root clauses belong to CL_K or all belong to CL_e, then p_i is being resolved on locally. In this case, we can replace p_i in either set of clauses without affecting their unsatisfiability. For example, in Fig. 1(b), p_0 can be replaced in CL_e by a new unconstrained variable x_0. This intuition is formalized below.

Given a resolution proof Π, a variable l, and a clause c, we denote by $S(l, c)$ the set of all root clauses that have contributed the variable l to c. $S(l, c)$ is defined recursively as shown in Fig. 3. A root clause c_r is an element of $S(l, c)$ if it contains

$L(c)$ **: clause** c**, variable** $p \rightarrow$ {**'∅', 'A', 'B', 'AB'**}

- **if** $c \in Roots(\Pi)$ **then**

$$L(c) \quad = \quad \begin{cases} \text{'}\emptyset\text{'} & \text{if } p \notin c \\ \text{'A'} & \text{if } p \in c \wedge c \in A \\ \text{'B'} & \text{if } p \in c \wedge c \in B \end{cases}$$

- **else if** c is a clause resulting from resolving c_1 and c_2 on variable v, i.e., $c = \exists v \cdot c_1 \wedge c_2$, **then**
  - **if** $v \neq p$, **then**

$$L(c) \quad = \quad \begin{cases} \text{'}\emptyset\text{'} & \text{if } L(c_1) = L(c_2) = \text{'}\emptyset\text{'} \\ \text{'A'} & \text{if } \exists i, j \cdot L(c_i) = \text{'A'} \wedge L(c_j) \subseteq \{\text{'A'}, \text{'}\emptyset\text{'}\} \\ \text{'B'} & \text{if } \exists i, j \cdot L(c_i) = \text{'B'} \wedge L(c_j) \subseteq \{\text{'B'}, \text{'}\emptyset\text{'}\} \\ \text{'AB'} & \text{otherwise} \end{cases}$$

  - **else if** $v = p$, **then**

$$L(c) \quad = \quad \begin{cases} \text{'}\emptyset\text{'} & \text{if } L(c_1) = L(c_2) \\ \text{'AB'} & \text{otherwise} \end{cases}$$

Fig. 2. Labeling function for the peripherality algorithm.

$$S(l, c) \quad = \quad \begin{cases} \emptyset & \text{if } l \notin c \\ c & \text{if } c \in Roots(\Pi) \wedge l \in c \\ S(l, c_1) \cup S(l, c_2) & \text{if } c_1 \text{ and } c_2 \text{ are parents of } c \ \wedge \ l \in c \end{cases}$$

Fig. 3. Definition of $S(l, c)$.

a variable l and there exists a path from c_r to c that does not contain a resolution on l. We can now define *peripherality* of variables, which captures the conditions when a global variable may not be central to the reason why φ holds on K.

Definition 5 *Let A and B be disjoint sets of clauses such that $C = A \cup B$ is unsatisfiable, and Π be a resolution proof establishing unsatisfiability of C. Then a variable l is* peripheral *with respect to A and B iff for every resolution on l between clauses c_1 and c_2, $S(l, c_1) \cup S(l, c_2) \subseteq A$ or $S(l, c_1) \cup S(l, c_2) \subseteq B$.*

Within the BMC setting, we have the following definition:

Definition 6 *Let K be a model, φ be an LTL formula, $BMC_k(K, \varphi)$ be a CNF encoding of a BMC problem for $K \models_k \varphi$, and Π be a proof of unsatisfiability of $BMC_k(K, \varphi)$. p is* peripheral *in φ iff for each time instance i, p_i is peripheral in Π with respect to CL_e and CL_K.*

If a variable is peripheral, it is vacuous by Theorem 3.

Theorem 3 *Let Π be a proof of unsatisfiability of $BMC_k(K, \varphi)$. If a variable p of φ is peripheral in Π, then φ is k-step p-vacuous.*

In Fig. 1(b), although p is not locally irrelevant in φ, it is peripheral, and therefore φ is p-vacuous. This also demonstrates that peripherality is a strictly stronger notion than local irrelevance. Theorem 4 shows that under our constraints this is the strongest result that we can hope to establish.

Theorem 4 *Assume φ is k-step p-vacuous in K. Then, there exists a resolution proof Π of unsatisfiability of $\mathrm{BMC}_k(K, \varphi)$ such that p is peripheral in Π.*

This is one of the main contributions of this paper: if a variable appears in all proofs, but is detected as peripheral in at least one of these proofs, it is vacuous. Conversely, if a variable appears in all proofs but is not peripheral in any of them, it is definitively not vacuous.

Peripherality of a variable can be detected by traversing the resolution proof from the roots to the leaf, keeping track of the source of the variable in each clause. If Π is a resolution proof whose root clauses are divided into two disjoint sets, $A \cup B$, then the labeling function L is defined recursively as shown in Fig. 2, where c is used to represent a clause. This labeling function defines an algorithm for detecting peripherality.

A CNF variable v is peripheral iff the label of the empty clause is not 'AB'. Thus, to detect whether a formula φ is p-vacuous, we need to check that all CNF variables p_i corresponding to p (see Sec. II) are peripheral. This can be done by applying the labeling function described in Fig. 2 with $A = CL_K$, and $B = CL_e$ for each p_i (for details, see [8]). It is also possible to simultaneously keep track of the labels for all CNF variables so that only a single pass through Π is needed. The time complexity of the peripherality algorithm is linear in the size of the resolution proof.

Theorem 5 *For a resolution proof Π that $BMC_k(K, \varphi)$ is unsatisfiable, determining which variables of φ are peripheral can be done in time linear in the size of Π.*

In this section, we defined three methods of detecting vacuity based on examining the UNSAT core and the resolution proof produced by BMC. Our evaluation of these algorithms w.r.t. precision and execution times can be found in Sec. V.

V. Practical Experience

The techniques reported in this paper have been implemented in a tool called ***VaqTree*** (see [8] for a description of this tool). The inputs to ***VaqTree*** are a model (encoded using the language of NuSMV [10]) and an LTL property. The tool returns the vacuity status of each variable in the property. Vacuity detection in ***VaqTree*** proceeds in two phases: a "partial pass" that applies one of our methods, and a "model-checking pass" that completes the analysis using additional model-checking runs.

We have run ***VaqTree*** on two benchmark suites. To evaluate the overall performance of the tool and the effectiveness of our partial vacuity detection methods, we have created a benchmark suite $\mathcal{S}_\mathcal{A}$ using various models and properties from the NuSMV distribution. To evaluate the scalability of the tool to industrial models, we have created a benchmark suite $\mathcal{S}_\mathcal{B}$ from the models in the IBM Formal Verification Benchmarks Library [18].

These models came with rather simple properties (one temporal operator), and (as expected from an industrial benchmark) did not exhibit a high degree of vacuity. Thus, we used this suite to measure the "worst-case" behavior of the tool, i.e., the amount of overhead incurred by our methods when no vacuity is found.

In the benchmarks, each test case consists of a model M, a property φ, and a bound k such that $M \models_k \varphi$. Note that finding an appropriate bound k is orthogonal to k-vacuity detection, which explains why our evaluation does not consider the time needed to find k. The experiments were performed on a Linux machine with a 2.8GHz P4 CPU, and 1GB of RAM, with up to 700MB of RAM available to each process. Currently, ***VaqTree*** is limited to proofs with up to 2.5 million resolutions. In $\mathcal{S}_\mathcal{A}$, this corresponds to a test case from the asynchronous **abp4** model (roughly 30 boolean variables, with $k = 19$). A sample of our experimental data is available in Appendix B, and the full results – in [19]. Below, we discuss results obtained with each benchmark individually.

A. Results obtained with $\mathcal{S}_\mathcal{A}$

This benchmark suite consists of 5 models: **abp4**, **msi_wtrans**, **pci**, and **prod-cell** from the NUSMV distribution (107 properties) and **toyFGS04** from [20] (14 properties). On average, the properties in the suite have 2 temporal operators (from the set G, F, U and X), with a maximum of 4 operators, and include both liveness and safety. 99 of the properties exhibit vacuity, and 22 do not.

Scatter plots in Fig. 4 compare the execution times of ***VaqTree*** (parametrized with irrelevance, local irrelevance, and peripherality), with naive detection for this benchmark. Execution times for naive detection include CNF theory generation and satisfiability testing for each variable of the property. Execution times for ***VaqTree*** include the time for the partial pass and the subsequent model-checking pass. Each point in the plot represents a single test case. The X-axis represents the time (in seconds) taken by naive detection. The Y-axis represents the time (in seconds) taken by ***VaqTree*** when parameterized by each of our methods. Points below the diagonal indicate where ***VaqTree*** was faster than naive detection; points near the diagonal indicate cases where the partial pass found a small percentage of the vacuous variables.

Fig. 5 shows that on $\mathcal{S}_\mathcal{A}$, ***VaqTree*** with irrelevance finds the fewest vacuous variables among our partial methods, as expected from the discussion in Section IV. Although Fig. 4(b) and (c) look similar, the numbers (see Appendix B and [19]) show that local irrelevance is faster than peripherality in 96% of the cases. This is consistent with the additional work peripherality must perform to analyze the proof tree. A detailed comparison of local irrelevance and naive detection shows that ***VaqTree*** with local irrelevance was faster or comparable to naive detection in 95% of the test cases. ***VaqTree*** with local irrelevance was faster than naive detection in 70 (58%) of the test cases, out of which 30 cases were twice as fast, and 20 cases were faster by an order of magnitude. In the remaining 51 cases, local irrelevance was at most 3% slower in 86% of these cases.

There are 10 cases where ***VaqTree*** with peripherality took much longer than naive detection. All of these cases are from the **abp4** model, and while they have the largest resolution proofs of the benchmark suite (between 300,000 and 2M clauses), other 300,000-clause test cases did not yield poor performance. We conjecture that the poor performance is due to a low clause/variable ratio [21] which favours naive detection in cases where vacuity is not present. Intuitively, a low ratio indicates that the SAT instance is underconstrained, and so a solution (if it exists) can be found quickly. On the other hand, finding a proof of *unsatisfiability* in a model with few constraints can be more difficult. Naive detection on a non-vacuous property requires solving satisfiable SAT instances, since replacing variables falsifies the property. However, peripherality on a non-vacuous property requires time linear in the size of the resolution proof obtained from the original model-checking run. If all of these SAT instances have a low clause/variable ratio, naive detection can be much faster than peripherality. This situation was only observed on the **abp4** model, with clause/variable ratio of 1.5-1.8 – significantly lower than any other test case with large proofs and without vacuity.

We now turn to measuring the effectiveness of our methods, using the number of vacuous variables found during the partial pass as a metric (see the scatter plots in Fig. 5). This number indicates how many additional model-checking runs are needed to complete vacuity detection. Since our partial methods can be ordered by increasing precision, Fig. 5(a) compares irrelevance and local irrelevance, Fig. 5(b) – local irrelevance and peripherality, and Fig. 5(c) – peripherality and naive detection. Each point in the plot represents a set of test cases – a larger point means a larger set. The axes show the number of vacuous variables detected by each method. Points below the diagonal indicate where the X-axis method detects more vacuous variables than the Y-axis method. The plots show that local irrelevance is clearly more effective than irrelevance. Contrary to our expectations, peripherality performed exactly as local irrelevance in all but 5 cases. Thus, local irrelevance appears to be more cost-effective. Fig. 5(c) shows that our techniques are effective when compared with naive detection: peripherality reduced the number of extra model-checking runs by 40% in 54 out of 99 cases that exhibited vacuity.

B. Results obtained with $\mathcal{S}_\mathcal{B}$

This benchmark suite consists of 13 models from the IBM Formal Verification Benchmarks Library [22] (18 properties). The properties have a single temporal operator (G or F), and include both safety and liveness. 12 of the properties exhibit vacuity, and 6 do not. evaluate the scalability of ***VaqTree*** to industrial models, we must first bound such that $M \models_k \varphi$. For this benchmark, we picked depth $k = 20$, which is in line with the bounds used for analyzing these models in [22, Sec. 2]. At this depth, only 13 models from the benchmark were suitable for our experiments. We report on the experiments below. At this k, some of the models where too large to analyze using VaqTree, and some of the properties did not hold. This is why we only report data for 13 models from this benchmark.

Table I, which includes full results for $\mathcal{S}_\mathcal{B}$, shows that proof sizes for this benchmark can be handled by ***VaqTree***. Interest-

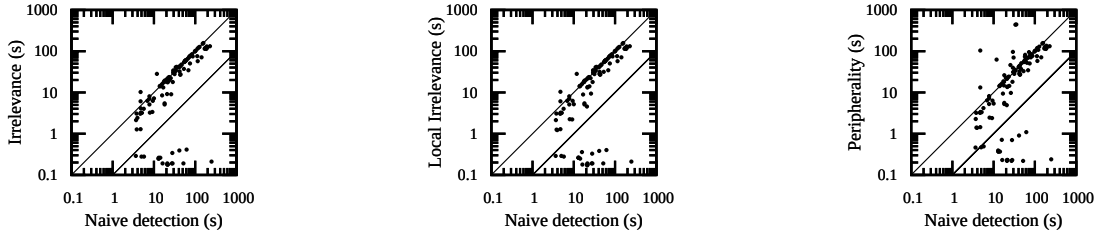

Fig. 4. $\mathcal{S}_{\mathcal{A}}$: Comparison of execution times. Where applicable, all times include times for both the partial and model-checking passes.

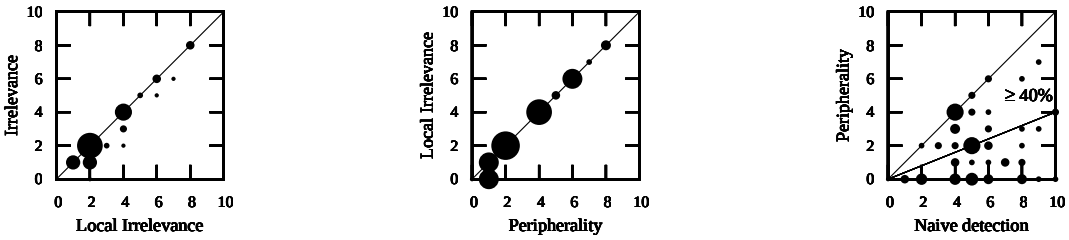

Fig. 5. $\mathcal{S}_{\mathcal{A}}$: Comparison of the number of vacuous variables detected by partial pass. Larger points represent more test cases than the smaller points.

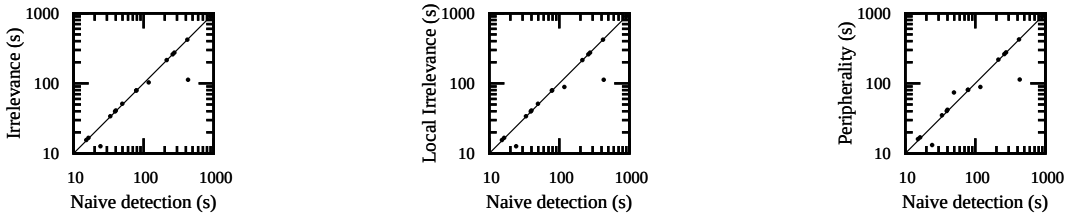

Fig. 6. $\mathcal{S}_{\mathcal{B}}$: Comparison of execution times. Where applicable, all times include times for both the partial and model-checking passes.

ingly, these are in the same range as proof sizes for $\mathcal{S}_{\mathcal{A}}$. This could be explained by the fact that even though these models are more complex, properties are simpler.

Scatter plots in Fig. 6 compare the execution times of ***VaqTree*** parametrized with local irrelevance and peripherality, with naive detection for this benchmark. Execution times are measured as described in Section V-A, and the graphs are interpreted in the same way as those in Fig. 4. Since $\mathcal{S}_{\mathcal{B}}$ had low vacuity, we did not expect our techniques to find it without the help of naive detection. However, graphs in Fig. 6 show that our techniques do in fact detect vacuity, as indicated by the points that appear below the diagonal. Both local irrelevance and peripherality detect the same amount of vacuity in $\mathcal{S}_{\mathcal{B}}$, but local irrelevance is slightly faster than peripherality.

Surprisingly, peripherality introduces a low overhead in this benchmark – points over the diagonal are near it, unlike what we see in Fig. 4. To explain this behavior, we hypothesized that in non-vacuous cases with low clause/variable ratios and large proofs, peripherality is much slower than naive detection. In $\mathcal{S}_{\mathcal{B}}$, we found that 15 of the test cases have a clause/variable ratio between 2.62-3.66, much higher than the ratios encountered in $\mathcal{S}_{\mathcal{A}}$. The remaining three cases had ratios in the same range as the **abp04** model. However, two of these produce trivial proofs, and the last one exhibits vacuity. These results empirically support our hypothesis.

C. Conclusions

In summary, we observed that local irrelevance performs best out of our proposed partial methods, finding most vacuity in the least amount of time. In 95% of both benchmark suites, we found ***VaqTree*** with local irrelevance to be at most 3% slower, and usually much faster, than the naive detection. In several tests of the $\mathcal{S}_{\mathcal{A}}$ benchmark, peripherality was noticeably slower than naive detection. On the industrial benchmark $\mathcal{S}_{\mathcal{B}}$, the overhead produced by peripherality was negligible. Interestingly, this suggests that peripherality may be a viable alternative to local irrelevance on industrial models. We plan to investigate this further in the future. Thus, we believe that both local irrelevance and peripherality can be used to replace naive detection. We plan to enhance our methods by developing a heuristic based on the clause/variable ratio and proof size that indicates when naive detection should be applied instead. Finally, ***VaqTree*** outputs the vacuity results for each timed variable p_i as a byproduct of its partial pass. This information gives an *explanation of non-vacuity*, indicating which time steps have been important for deciding whether a given variable was vacuous, thus facilitating debugging.

VI. Summary and Related Work

In this paper, we showed how to exploit the UNSAT core and resolution proof produced by a successful run of BMC for vacuity detection. We introduced three vacuity detection methods that can be applied with little overhead after one model-checking run in order to quickly identify vacuous variables and reduce the number of additional model-checking runs required. Two of these methods, irrelevance and local irrelevance, exploit the UNSAT core, and the third, peripherality, is based on analyzing the resolution proof. We built a tool ***VaqTree***, which

is based on these methods, and showed that it is effective for speeding up vacuity detection.

Related work on vacuity detection has been described in Section III. Additionally, our work is related to research in declarative modeling. In particular, our use of the UNSAT core to detect vacuity was inspired by [23], which addresses the problem of identifying overconstraint in declarative models. While similar in spirit to vacuity detection in model checking, declarative models have no explicit transition relation; instead, transitions are expressed with constraints [24], [25]. An overconstraint occurs when the model satisfies a safety property because all violations of the formula have been accidentally ruled out by the declared constraints. In order to detect such overconstraints, [23] introduces the idea of *core extraction*: declarative models are reduced to SAT instances, from which an UNSAT core can be extracted if the property holds. If a constraint's clauses do not appear in the UNSAT core, the constraint is called *irrelevant*, and is a source of overconstraint (similar to Def. 3). The cone-of-influence technique [1] is also similar to Def. 3. However, as both of these techniques are model-based, neither can be used to detect vacuity.

Our experiments show that local irrelevance and peripherality can detect more vacuous variables than irrelevance. Therefore, detecting overconstraint in declarative models may also benefit from methods that analyze the structure of the resolution proof. In the future, we propose to investigate how a notion equivalent to peripherality can be defined in the declarative setting. Another goal of future work is to increase the power of resolution proof-based vacuity detection methods. In this paper, we restricted ourselves to using results of only one BMC run, and to methods with linear time complexity in the size of the resolution proof or better. However, it is possible that the most optimal trade-off between speed and effectiveness of vacuity detection algorithms lies in the domain of multiple resolution proofs, where we can find the minimal UNSAT core [26] or reduce the resolution proof using interpolation [27].

McMillan [6] uses interpolation to prove that a particular bound is sufficient to imply the unbounded satisfaction of a BMC problem. We intend to combine our techniques with this algorithm in order to prove that bounded vacuity for the correct k implies that the property also holds vacuously in the unbounded case.

Interpolation can also be used to detect vacuity. Given two sets of clauses, A and B, such that $A \cup B$ is unsatisfiable, an interpolant C is a set of clauses whose variables appear in both A and B, such that $B \cup C$ is unsatisfiable and $A \Rightarrow C$ [28]. Intuitively, if C is minimal, then C is the reason why $A \cup B$ is unsatisfiable. This intuition suggests that if an interpolant of CL_K and CL_e could be found, then all variables not appearing in it could be considered vacuous. However, we did not include this technique in our empirical evaluation, as our interpolant generator was comparatively slower.

Another means of speeding vacuity detection for BMC is to iteratively check the k-step vacuity of each variable starting with $k = 0$. Since $K \not\models_{k_1} \varphi[p \leftarrow x]$ implies $K \not\models_{k_2} \varphi[p \leftarrow x]$ for all $k_2 > k_1$, if a variable is proven non-vacuous at some step k, then it can be omitted from subsequent checks of higher k. This method is orthogonal to our techniques, and the vacuity detection at each step could be carried out by ***VaqTree***.

REFERENCES

[1] E. Clarke, O. Grumberg, and D. Peled, *Model Checking*.MIT Press, 1999.

[2] I. Beer, S. Ben-David, C. Eisner, and Y. Rodeh, "Efficient Detection of Vacuity in ACTL Formulas," in *Proc. of CAV'97*, ser. LNCS, vol. 1254, 1997, pp. 279–290.

[3] O. Kupferman and M. Vardi, "Vacuity Detection in Temporal Model Checking," in *Proc. of CHARME'99*, ser. LNCS, vol. 1703, 1999.

[4] M. Purandare and F. Somenzi, "Vacuum Cleaning CTL Formulae," in *Proc. of CAV'02*, ser. LNCS, vol. 2404, 2002, pp. 485–499.

[5] R. Armoni, L. Fix, A. Flaisher, O. Grumberg, N. Piterman, A. Tiemeyer, and M. Vardi, "Enhanced Vacuity Detection in Linear Temporal Logic," in *Proc. of CAV'03*, vol. 2725, July 2003, pp. 368–380.

[6] K. McMillan, "Interpolation and SAT-Based Model Checking," in *Proc. of CAV'03*, ser. LNCS, vol. 2725, July 2003, pp. 1–13.

[7] A. Biere, A. Cimatti, E. M. Clarke, and Y. Zhu, "Symbolic Model Checking without BDDs," in *Proc. of TACAS'99*, ser. LNCS, vol. 1579, 1999.

[8] J. Simmonds, J. Davies, A. Gurfinkel, and M. Chechik, "Exploiting Resolution Proofs for LTL Vacuity Detection," ftp://ftp.cs.toronto.edu/pub/reports/csrg/547/TR-547.ps, DCS, University of Toronto, CSRG TR 547, 2007.

[9] A. Cimatti, M. Pistore, M. Roveri, and R. Sebastiani, "Improving the Encoding of LTL Model Checking into SAT," in *Proc. of VMCAI'02*, ser. LNCS, vol. 2294, 2002, pp. 196–207.

[10] A. Cimatti, E. M. Clarke, E. Giunchiglia, F. Giunchiglia, M. Pistore, M. Roveri, R. Sebastiani, and A. Tacchella, "NuSMV 2: An OpenSource Tool for Symbolic Model Checking." in *Proc. of CAV'02*, ser. LNCS, vol. 2404, 2002, pp. 359–364.

[11] N. Een and N. Sörensson, "The MiniSat Page," http://www.cs.chalmers.se/Cs/Research/FormalMethods/MiniSat/Main.html, April 2006.

[12] L. Zhang and Z. Fu, "Boolean Satisfiability Research Group at Princeton," http://www.princeton.edu/~chaff/, September 2006.

[13] L. Zhang and S. Malik, "Validating SAT Solvers Using an Independent Resolution-Based Checker: Practical Implementations and Other Applications," in *Proc. of DATE'03*, 2003, pp. 10 880–10 885.

[14] A. Gurfinkel and M. Chechik, "How Vacuous Is Vacuous?" in *Proc. of TACAS'04*, ser. LNCS, vol. 2988, March 2004, pp. 451–466.

[15] ——, "Extending Extended Vacuity," in *Proc. of FMCAD'04*, ser. LNCS, vol. 3312, 2004, pp. 306–321.

[16] K. Namjoshi, "An Efficiently Checkable, Proof-Based Formulation of Vacuity in Model Checking," in *Proc. of CAV'04*, ser. LNCS, vol. 3114, 2004, pp. 57–69.

[17] I. Beer, S. Ben-David, C. Eisner, and Y. Rodeh, "Efficient Detection of Vacuity in Temporal Model Checking," *FMSD*, vol. 18, no. 2, 2001.

[18] I. Haifa, "CNF Benchmarks from IBM Formal Verification Benchmarks Library," 2007. [Online]. Available: http://www.haifa.il.ibm.com/projects/verification/RB_Homepage/benchmark%s.html

[19] J. Simmonds and J. Davies, "VaqTree," http://www.cs.toronto.edu/~jsimmond/VaqTree, 2007.

[20] M. Heimdahl, S. Rayadurgam, W. Visser, G. Devaraj, and J. Gao, "Auto-generating Test Sequences Using Model Checkers: A Case Study," in *Proc. of FATES'03*, ser. LNCS, vol. 2931, 2003, pp. 42–59.

[21] B. Selman, D. Mitchell, and H. Levesque, "Generating Hard Satisfiability Problems," *Artificial Intelligence*, vol. 81, no. 1-2, pp. 17–29, 1996.

[22] E. Zarpas, "Benchmarking SAT Solvers for Bounded Model Checking," in *SAT*, 2005, pp. 340–354.

[23] I. Shlyakhter, R. Seater, D. Jackson, M. Sridharan, and M. Taghdiri, "Debugging Overconstrained Declarative Models Using Unsatisfiable Cores," in *Proc. of ASE'03*, October 2003, pp. 94–105.

[24] D. Jackson, "Alloy: a Lightweight Object Modelling Notation," *ACM TOSEM*, vol. 11, no. 2, pp. 256–290, 2002.

[25] J. M. Spivey, *The Z Notation: a Reference Manual*. Prentice Hall, 1992.

[26] R. Gershman, M. Koifman, and O. Strichman, "Deriving Small Unsatisfiable Cores with Dominators," in *Proc. of CAV'06*, ser. LNCS, vol. 4144, 2006, pp. 109–122.

[27] W. Craig, "Linear Reasoning. A New Form of the Herbrand-Gentzen Theorem," *JSL*, vol. 22, pp. 250–268, 1957.

[28] T. Henzinger, R. Jhala, R. Majumdar, and K. McMillan, "Abstractions from Proofs," in *Proc. of POPL'04*. ACM, January 2004, pp. 232–244.

APPENDIX

A. Proofs of Theorems

Proofs of selected theorems are given. Additional proofs can be found in [8, Appendix A2].

Theorem 2 *If p is locally irrelevant with respect to $BMC_k(K, \varphi)$ and Π, then φ is k-step p-vacuous.*

Proof: Let $BMC_k(K, \varphi) = CL_K \cup CL_e$ and U be the UNSAT core of Π. Assume that p is locally irrelevant in $BMC_k(K, \varphi)$. So for all p_i, either p_i does not appear in U, or p_i is local to $CL_e \cap U = U_e$ or to $CL_K \cap U = U_K$ by Def. 4. Let $U_{e'}$ be U_e with each occurence of p_i replaced by x_i. Since each p_i that has been replaced is local to U_e, and $U_K \cup U_e = U$ is unsatisfiable, then $U_K \cup U_{e'}$ is also unsatisfiable. Since $U_{e'} \subseteq CL_e[p \leftarrow x]$, the set of clauses $CL_K \cup CL_e[p \leftarrow x]$ is unsatisfiable as well. Therefore, $K \models_k \varphi[p \leftarrow x]$ holds, so φ is p-vacuous. □

Theorem 3 *Let Π be a proof of unsatisfiability of $BMC_k(K, \varphi)$. If a variable p of φ is peripheral in Π, then φ is k-step p-vacuous.*

Proof: Let $BMC_k(K, \varphi) = CL_K \cup CL_e$ and U be the UNSAT core of Π. Assume that p is peripheral in $BMC_k(K, \varphi)$. Let $U_{e'}$ be the result of replacing each p_i with x_i in $CL_e \cap U$. Then $(CL_K \cap U) \cup U_{e'}$ is still unsatisfiable, since every resolution on x_i must be local to $CL_e \cap U$, and every resolution on p_i must be local to $CL_K \cap U$ by the peripherality of p_i. Since $U_{e'} \subseteq CL_e[p \leftarrow x]$, $CL_K \cup CL_e[p \leftarrow x]$ is unsatisfiable as well. Therefore, $K \models_k \varphi[p \leftarrow x]$, and φ is p-vacuous. □

Theorem 4 *Assume φ is k-step p-vacuous in K. Then, there exists a resolution proof Π of unsatisfiability of $BMC_k(K, \varphi)$ such that p is peripheral in Π.*

Proof: Assume that φ is p-vacuous. Then, the BMC problem $BMC_k(K, \varphi[p \leftarrow x]) = CL_K \cup CL_e[p \leftarrow x]$ is unsatisfiable, and there exists a resolution proof Π establishing this. We must show that this proof can be transformed to a proof of unsatisfiability of $BMC_k(K, \varphi) = CL_K \cup CL_e$ in which each p_i is peripheral with respect to CL_K and CL_e.

If Π does not contain a clause that has both p_i and x_i for some i, then for Π' obtained from Π by replacing each occurrence of x_i by p_i, (a) Π' is a well-formed resolution proof, and (b) $Roots(\Pi') \subseteq CL_K \cup CL_e$. That is, Π' is a resolution proof establishing that $BMC_k(K, \varphi)$ is unsatisfiable.

We now show how such a proof can be constructed from an arbitrary proof Π of unsatisfiability of $BMC_k(K, \varphi[p \leftarrow x])$. Let $U_K = Roots(\Pi) \cap CL_K$, and $CL_{e'} = CL_e[p \leftarrow x]$. Then, if p_i occurs in any clause of U_K, it is local to U_K. Let L be the set of all local variables of U_K, $C = \exists L \cdot U_K$ be a formula resulting from existentially eliminating these local variables, and $CNF(C)$ be the CNF encoding of C. For any i, p_i does not appear in C. Furthermore, the set of clauses $CNF(C) \cup CL_{e'}$ is unsatisfiable. Thus, there exists a resolution proof Π' establishing this such that $Roots(\Pi') \subseteq CL_{e'} \cup CNF(C)$. Finally, since $U_K \Rightarrow C$, for each clause $c \in C$ there exists a resolution proof Π_c such that $Leaf(\Pi_c) = c$ and $Roots(\Pi_c) \subseteq U_K$. By combining the proofs $\{\Pi_c \mid c \in CNF(C)\}$ and Π', we obtain a proof of unsatisfiability of $U_K \cup CL_{e'}$ that does not contain a clause with variables x_i and p_i. □

B. Experiments

Table I shows detailed results of our experiments. In this table, column "Benchmark" indicates the benchmark the test case belongs to; "Test case" is the case's unique identifier inside the benchmark, "Model" is the SMV model tested; "# var. in M" is the number of variables in the model; "k" is the number of steps used to run BMC; "op. in φ" shows the property operators (e.g., *2G* means that two G operators appear in the property); "# var. in φ" is the number of atomic variables present in the property; "# vac. vars." is the number of vacuous variables; and "# resol. in Π" is the number of resolutions in the resolution proof. The next three columns report the time needed for model-checking: "Gen. CNF" is the time NuSMV took to generate the corresponding CNF theory; "Test SAT" and "Gen. Π" are the time MiniSat took to test satisfiability and generate the corresponding resolution proof respectively; and "Total" is the sum of the previous three columns.

For the naive method, we report the total times for the CNF theory generation ("Gen. CNF") and for satisfiability testing ("Test SAT"). One CNF theory is produced per each atomic variable. For irrelevance, local irrelevance and peripherality, we report how many vacuous variables were found by the partial pass ("# vac. vars. found"), how long ***VaqTree*** took to do the corresponding analysis ("Anal.") and how much time was needed to do the completeing pass ("Extra runs").

For example, test case **8** analyzes a five-variable, two temporal operator (G,U) property of the **pci** model (which has 40 variables). Only three of these variables are vacuous. The resolution proof generated when $k = 13$ has 4,283 resolutions. This property was checked in 5.59 seconds. Naive vacuity detection required five model-checking runs, taking 25.85 seconds to generate the corresponding CNF theories and 2.89 seconds to test their satisfiability, requiring a total of 28.74 seconds. Irrelevance took 0.27 seconds to find two of the vacuous variables during the partial pass. It then took 17.60 seconds to carry out the completing pass, so the total time required by irrelevance to find all three vacuous variables is 17.87 seconds. Local irrelevance took 0.28 seconds to analyze the resolution proof, finding the same two vacuous variables as irrelevance. Thus, it also takes 17.60 seconds to run the completing pass, so the total time required by local irrelevance is 17.88 seconds. Finally, peripherality took 0.47 seconds to execute the partial pass and found the same two vacuous variables; it also required 17.60 seconds to run the completing pass, taking a total of 18.07 seconds to produce complete results for test case **8**.

VaqTree, the complete experimental results and some test cases are available at [19].

TABLE I: Statistics for vacuity detection experiments on NuSMV distribution and other examples.

Bench-mark	Test case	Model (M)	# var. in M	k	op. in φ	# var. in φ	# vac. vars.	# resol. in Π	Model Checking				Naive			Irrelevance				Local Irrelevance (LI)				Peripherality (P)			
									Gen. CNF (s)	Test SAT (s)	Gen. Π (s)	Total (s)	Gen. CNF (s)	Test SAT (s)	Total (s)	# vac. vars. found	Anal. (s)	Extra runs (s)	Total (s)	# vac. vars. found	Anal. (s)	Extra runs (s)	Total (s)	# vac. vars. found	Anal. (s)	Extra runs (s)	Total (s)
$\mathcal{S}_\mathcal{A}$	1	pci	40	13	G,U	4	1	19792	4.69	0.23	5.9	10.82	20.66	2.77	23.43	0	0.34	23.43	23.77	0	0.34	23.43	23.77	0	0.81	23.43	24.24
$\mathcal{S}_\mathcal{A}$	2	pci	40	13	G,U	4	3	1649	5.13	0.14	5.64	10.91	11.75	1.30	13.05	3	0.26	0	0.26	3	0.26	0	0.26	3	0.37	0	0.37
$\mathcal{S}_\mathcal{A}$	3	pci	40	13	G,U	4	3	1649	5.09	0.13	5.32	10.54	12.03	2.14	14.17	3	0.26	0	0.26	3	0.25	0	0.25	3	0.37	0	0.37
$\mathcal{S}_\mathcal{A}$	4	pci	40	13	G,U	3	1	7725	4.80	0.18	5.65	10.63	12.68	1.73	14.41	0	0.29	14.41	14.7	0	0.29	14.41	14.70	0	0.50	14.41	14.91
$\mathcal{S}_\mathcal{A}$	5	pci	40	13	G,U	3	1	7555	4.76	0.18	5.55	10.49	12.36	1.56	13.92	0	0.28	13.92	14.20	0	0.28	13.92	14.20	0	0.50	13.92	14.42
$\mathcal{S}_\mathcal{A}$	6	pci	40	13	G,U	4	3	1705	4.66	0.12	5.68	10.46	11.66	1.19	12.85	3	0.25	0	0.25	3	0.26	0	0.26	3	0.39	0	0.39
$\mathcal{S}_\mathcal{A}$	7	pci	40	13	G,U	4	3	1705	4.67	0.14	5.42	10.23	11.68	1.40	13.08	3	0.25	0	0.25	3	0.26	0	0.26	3	0.37	0	0.37
$\mathcal{S}_\mathcal{A}$	8	pci	40	13	G,U	5	3	4283	4.95	0.22	5.59	10.76	25.85	2.89	28.74	2	0.27	17.60	17.87	2	0.28	17.60	17.88	2	0.47	17.60	18.07
$\mathcal{S}_\mathcal{A}$	66	prod-cell	39	10	G,F	6	6	5275	0.88	0.04	1.25	2.17	5.55	0.18	5.73	6	0.28	0	0.28	6	0.28	0	0.28	6	0.49	0	0.49
$\mathcal{S}_\mathcal{A}$	67	prod-cell	39	10	G,F	5	5	5320	1.02	0.04	1.41	2.47	4.81	0.16	4.97	5	0.28	0	0.28	5	0.29	0	0.29	5	0.47	0	0.47
$\mathcal{S}_\mathcal{A}$	68	prod-cell	39	10	G,F	4	4	3798	0.91	0.03	1.27	2.21	3.57	0.12	3.69	2	0.27	1.86	2.13	2	0.27	1.86	2.13	2	0.43	1.86	2.29
$\mathcal{S}_\mathcal{A}$	69	prod-cell	39	10	G,F	8	8	2764	0.99	0.03	1.26	2.28	7.52	0.23	7.75	1	0.26	6.78	7.04	1	0.26	6.78	7.04	1	0.42	6.78	7.2
$\mathcal{S}_\mathcal{A}$	70	prod-cell	39	10	G,F	5	5	5232	1.20	0.04	1.33	2.57	4.63	0.15	4.78	1	0.28	3.82	4.10	2	0.28	2.86	3.14	2	0.48	2.86	3.34
$\mathcal{S}_\mathcal{A}$	71	prod-cell	39	10	G,F	4	4	4068	1.35	0.03	1.27	2.65	3.87	0.10	3.97	2	0.27	2.16	2.43	3	0.27	0.95	1.22	3	0.44	0.95	1.39
$\mathcal{S}_\mathcal{A}$	72	prod-cell	39	10	G,F	4	4	2756	0.96	0.03	1.27	2.26	3.64	0.13	3.77	1	0.26	2.82	3.08	1	0.26	2.82	3.08	1	0.40	2.82	3.22
$\mathcal{S}_\mathcal{A}$	73	prod-cell	39	10	G,F	6	6	4425	0.84	0.04	1.30	2.18	5.47	0.19	5.66	2	0.28	3.74	4.02	2	0.28	3.74	4.02	2	0.46	3.74	4.2
$\mathcal{S}_\mathcal{A}$	74	prod-cell	39	10	G,F	5	5	3802	0.92	0.04	1.28	2.24	4.55	0.17	4.72	4	0.27	1.01	1.28	4	0.28	1.01	1.29	4	0.43	1.01	1.44
$\mathcal{S}_\mathcal{A}$	75	prod-cell	39	10	G,F	5	5	2802	0.91	0.03	1.44	2.38	4.53	0.14	4.67	2	0.26	2.80	3.06	2	0.26	2.80	3.06	2	0.41	2.80	3.21
$\mathcal{S}_\mathcal{A}$	76	prod-cell	39	10	G,F	8	8	3732	1.16	0.03	1.36	2.55	7.72	0.21	7.93	5	0.28	2.96	3.24	6	0.27	1.98	2.25	6	0.46	1.98	2.44
$\mathcal{S}_\mathcal{A}$	77	prod-cell	39	10	G,F	9	9	3010	1.50	0.03	1.28	2.81	8.93	0.22	9.15	6	0.27	3.12	3.39	7	0.27	1.94	2.21	7	0.45	1.94	2.39
$\mathcal{S}_\mathcal{A}$	78	prod-cell	39	10	G,F	5	5	2585	0.86	0.03	1.25	2.14	4.98	0.14	5.12	2	0.26	2.93	3.19	2	0.26	2.93	3.19	2	0.40	2.93	3.33
$\mathcal{S}_\mathcal{A}$	79	prod-cell	39	10	G,F	5	5	2556	1.06	0.03	1.30	2.39	4.70	0.12	4.82	2	0.26	2.98	3.24	2	0.26	2.98	3.24	2	0.40	2.98	3.38
$\mathcal{S}_\mathcal{A}$	80	prod-cell	39	10	G,F	4	4	5317	1.26	0.04	1.27	2.57	3.53	0.12	3.65	4	0.29	0	0.29	4	0.29	0	0.29	4	0.46	0	0.46
$\mathcal{S}_\mathcal{A}$	81	prod-cell	39	10	2G,2F	10	10	2497	3.15	0.06	1.29	4.5	9.68	0.27	9.95	3	0.26	6.97	7.23	4	0.26	4.94	5.20	4	0.42	4.94	5.36
$\mathcal{S}_\mathcal{A}$	82	prod-cell	39	10	G,F	8	8	2348	0.88	0.033	1.25	2.16	7.52	0.22	7.74	3	0.27	4.84	5.11	3	0.26	4.84	5.10	3	0.41	4.84	5.25
$\mathcal{S}_\mathcal{A}$	83	abp4	13	19	G,F	1	0	1289374	2.79	10.73	34.14	47.66	2.93	1.79	4.72	0	5.51	4.72	10.23	0	5.72	4.72	10.44	0	98.62	4.72	103.34
$\mathcal{S}_\mathcal{A}$	84	abp4	13	19	G,F	3	2	1050234	3.14	6.45	29.43	39.02	8.43	20.76	29.19	0	5.07	29.19	34.26	0	5.22	29.19	34.41	0	67.54	29.19	96.73
$\mathcal{S}_\mathcal{A}$	85	abp4	13	19	G,F	3	2	2246095	2.99	19.03	49.63	71.65	8.81	26.43	35.24	0	8.23	33.78	42.01	0	8.22	33.78	42	0	412.30	33.78	446.08
$\mathcal{S}_\mathcal{A}$	86	abp4	13	19	G,2F	2	0	795705	3.07	5.04	21.28	29.39	5.54	6.29	11.83	0	2.69	25.64	28.33	0	2.71	25.64	28.35	0	37.21	25.64	62.85
$\mathcal{S}_\mathcal{A}$	93	toyFGS04	151	18	F	6	6	297	18.88	0.26	5.27	24.41	114.78	0.76	115.54	3	0.23	57.39	57.62	3	0.22	57.39	57.61	3	0.29	57.39	57.68
$\mathcal{S}_\mathcal{A}$	94	toyFGS04	151	18	F	12	12	308	19.13	0.16	5.28	24.57	224.79	1.40	226.19	6	0.26	132.43	132.69	6	0.26	132.43	132.69	6	0.33	132.43	132.76
$\mathcal{S}_\mathcal{A}$	95	toyFGS04	151	18	F	6	0	318	18.35	0.15	5.17	23.67	126.28	32.03	158.31	0	0.22	158.31	158.53	0	0.22	158.31	158.53	0	0.29	158.31	158.60
$\mathcal{S}_\mathcal{A}$	96	toyFGS04	151	18	F	4	0	308	18.57	0.14	5.45	24.16	75.18	22.26	97.44	0	0.22	97.44	97.66	0	0.22	97.44	97.66	0	0.27	97.44	97.71
$\mathcal{S}_\mathcal{A}$	97	toyFGS04	151	18	G	4	0	8072	14.14	0.21	3.3	17.65	57.91	10.60	68.51	0	0.33	68.51	68.84	0	0.33	68.51	68.84	0	0.60	68.51	69.11
$\mathcal{S}_\mathcal{A}$	98	toyFGS04	151	18	G	6	0	7985	14.47	0.21	3.63	18.31	88.94	11.48	100.42	0	0.34	100.42	100.76	0	0.34	100.42	100.76	0	0.68	100.42	101.10
$\mathcal{S}_\mathcal{A}$	99	toyFGS04	151	18	F	6	6	293	19.80	0.15	5.61	25.56	111.91	0.66	112.57	2	0.21	75.08	75.29	2	0.22	75.08	75.30	2	0.27	75.08	75.35
$\mathcal{S}_\mathcal{A}$	107	msi_wtrans	30	40	G	5	3	66	21.85	0.20	8.39	30.44	120.15	65.70	185.85	3	0.21	112.59	112.80	3	0.20	112.59	112.79	3	0.24	112.59	112.83
$\mathcal{S}_\mathcal{A}$	108	msi_wtrans	30	40	F	5	4	66	23.53	0.20	9.15	32.88	120.16	73.28	193.44	3	0.2	120.30	120.50	3	0.21	120.30	120.51	3	0.25	120.30	120.55
$\mathcal{S}_\mathcal{A}$	109	msi_wtrans	30	40	F	6	4	66	21.56	0.21	8.46	30.23	156.61	93.23	249.84	4	0.21	0	0.21	4	0.21	0	0.21	4	0.24	0	0.24
$\mathcal{S}_\mathcal{B}$	1	IBM_FV_2002_03	111	20	G	8	8	7480	4.54	0.09	3.8	8.43	36.21	0.67	36.88	7	0.35	4.67	5.02	7	0.35	4.67	5.02	7	0.74	4.67	5.41
$\mathcal{S}_\mathcal{B}$	2	IBM_FV_2002_04	223	20	G	4	3	45065	7.62	0.92	5.71	14.25	29.66	3.83	33.49	0	0.59	33.49	34.08	0	0.59	33.49	34.08	0	1.67	33.49	35.16
$\mathcal{S}_\mathcal{B}$	3	IBM_FV_2002_05	310	20	G	2	1	32776	11.82	0.62	10.02	22.46	22.97	1.31	24.28	1	0.44	12.21	12.65	1	0.44	12.21	12.65	1	1.02	12.21	13.23
$\mathcal{S}_\mathcal{B}$	4	IBM_FV_2002_09	233	20	F	9	9	2	8.96	0.17	0	9.13	81.02	1.22	82.24	9	0.17	0	0.17	9	0.17	0	0.17	9	0.17	0	0.17
$\mathcal{S}_\mathcal{B}$	5	IBM_FV_2002_10	224	20	G	3	2	78523	54.23	8.45	46.09	108.77	165.88	93.22	259.1	0	0.7	259.1	259.8	0	0.7	259.1	259.8	0	2.33	259.1	261.43
$\mathcal{S}_\mathcal{B}$	6	IBM_FV_2002_10	224	20	G	4	3	177536	53.3	30.21	56.61	140.12	219.74	199.25	418.99	0	1.12	418.99	420.11	0	1.12	418.99	420.11	0	5.8	418.99	424.79
$\mathcal{S}_\mathcal{B}$	7	IBM_FV_2002_10	224	20	G	4	4	9119	53.97	0.97	40.84	95.78	218.45	211.21	429.66	3	0.32	112.58	112.9	3	0.32	112.58	112.9	3	0.61	112.58	113.19
$\mathcal{S}_\mathcal{B}$	8	IBM_FV_2002_10	224	20	G	2	0	155775	54.99	9.22	46.75	110.96	108.76	165.1	273.86	0	0.99	273.86	274.85	0	0.99	273.86	274.85	0	3.7	273.86	277.56
$\mathcal{S}_\mathcal{B}$	9	IBM_FV_2002_10	224	20	G	2	1	197053	54.96	65.43	79.32	199.71	110.2	103.82	214.02	0	1.09	214.02	215.11	0	1.12	214.02	215.14	0	4.5	214.02	218.52
$\mathcal{S}_\mathcal{B}$	10	IBM_FV_2002_17_1	1584	20	G	2	0	96085	38.58	1.15	14.23	53.96	75.78	2.35	78.13	0	1.07	78.13	79.2	0	1.08	78.13	79.21	0	2.47	78.13	80.6
$\mathcal{S}_\mathcal{B}$	11	IBM_FV_2002_17_2	1583	20	G	1	0	77553	38.5	0.86	13.68	53.04	38.82	1.35	40.17	0	0.88	40.17	41.05	0	0.88	40.17	41.05	0	1.74	40.17	41.91
$\mathcal{S}_\mathcal{B}$	12	IBM_FV_2002_17_2	1583	20	G	2	1	73790	38.47	0.82	23.98	63.25	77.01	1.77	78.78	0	0.89	78.78	79.67	0	0.89	78.78	79.76	0	1.98	78.78	80.76
$\mathcal{S}_\mathcal{B}$	13	IBM_FV_2002_19	121	20	G	1	0	35769	9.56	6.49	13.31	29.36	9.65	5.52	15.17	0	0.39	15.17	15.56	0	0.4	15.17	15.57	0	0.83	15.17	16
$\mathcal{S}_\mathcal{B}$	14	IBM_FV_2002_21	79	20	G	1	0	25508	8.8	5.61	11.98	26.39	8.68	7.61	16.29	0	0.37	16.29	16.66	0	0.37	16.29	16.66	0	0.71	16.29	17
$\mathcal{S}_\mathcal{B}$	15	IBM_FV_2002_22	104	20	G	1	0	53300	14.58	7.74	17.84	40.16	14.78	24.42	39.2	0	0.53	39.2	39.73	0	0.53	39.2	39.73	0	1.2	39.2	40.4
$\mathcal{S}_\mathcal{B}$	16	IBM_FV_2002_23	103	20	G	8	8	7618	14.52	0.69	11.29	26.5	115.68	2.36	118.04	1	0.36	103.01	103.37	2	0.35	88.41	88.76	2	0.74	88.41	89.15
$\mathcal{S}_\mathcal{B}$	17	IBM_FV_2002_27	43	20	G	8	6	431223	3.09	4.54	15.98	23.61	24.43	25.01	49.44	0	1.77	49.44	51.21	0	1.78	49.44	51.22	0	24.5	49.44	73.94
$\mathcal{S}_\mathcal{B}$	18	IBM_FV_2002_31_2	227	20	G	17	17	2	9.81	0.19	0	10	168.2	2.57	170.77	17	0.19	0	0.19	17	0.19	0	0.19	17	0.19	0	0.19