FuseIC3: An Algorithm for Checking Large Design Spaces

Rohit Dureja and Kristin Yvonne Rozier
Iowa State University

***Abstract*—The design of safety-critical systems often requires *design space exploration*: comparing several system models that differ in terms of design choices, capabilities, and implementations. Model checking can compare different models in such a set, however, it is continuously challenged by the state space explosion problem. Therefore, learning and reusing information from solving related models becomes very important for future checking efforts. For example, reusing variable ordering in BDD-based model checking leads to substantial performance improvement. In this paper, we present a SAT-based algorithm for checking a set of models. Our algorithm, FuseIC3, extends IC3 to minimize time spent in exploring the common state space between related models. Specifically, FuseIC3 accumulates artifacts from the sequence of over-approximated reachable states, called *frames*, from earlier runs when checking new models, albeit, after careful repair. It uses bidirectional reachability; forward reachability to repair frames, and IC3-type backward reachability to block predecessors to bad states. We extensively evaluate FuseIC3 over a large collection of challenging benchmarks. FuseIC3 is on-average up to 5.48× (median 1.75×) faster than checking each model individually, and up to 3.67× (median 1.72×) faster than the state-of-the-art incremental IC3 algorithm.**

I. Introduction

In the early phases of design, there are several models of the system under development constituting a *design space* [2, 19, 23]. Each model in such a set is a valid design of the system, and the different models differ in terms of core capabilities, assumptions, component implementations, or configurations. We may need to evaluate the different design choices, or to analyze a future version against previous ones in the product line. Model checking can be used to aid system development via a thorough comparison of the set of models. Each model in the set is checked one-by-one against a set of properties representing requirements. However, for large and complex design spaces, such an approach can be inefficient or even fail to scale to handle the combinatorial size of the design space. Nevertheless, model checking remains the most widely used method in industry when dealing with such systems [5, 19, 21, 23, 24].

We assume that different models in the design space have overlapping reachable states, and the models are checked sequentially. In a typical scenario, a model-checking algorithm doesn't take advantage of this information and ends up re-verifying "already explored" state spaces across models. For large models this can be extremely wasteful as every model-checking run re-explores already known reachable states. The problem becomes acute when model differences are small, or when changes in the models are outside the cone-of-influence of the property being checked, i.e., although the reachable states in the models vary, none of them are bad. Therefore, as the number of models grows, learning and reusing information from solving related models becomes very important for future checking efforts.

We present an algorithm that automatically reuses information from earlier model-checking runs to minimize the time spent in exploring the symbolic state space in common between related models. The algorithm, FuseIC3, is an extension to one of the fastest bit-level verification methods, IC3 [6], also known as *property directed reachability* (PDR) [17]. Given a set of models and a safety property, FuseIC3 sequentially checks each model by reusing information: reachable state approximations, counterexamples (cex), and invariants, learned in earlier runs to reduce the set's total checking time. When the difference between two subsequent models is small or beyond the cone-of-influence of the property, the invariant or counterexample from the earlier model may be directly used to verify the current model. Otherwise, FuseIC3 uses reachable state approximations as inputs to IC3 to only explore undiscovered reachable states in the current model. In the former, verification completes almost instantly, while in the latter, significant time is saved. When the stored information cannot be used directly, FuseIC3 repairs and patches it using an efficient SAT-based algorithm. The repair algorithm is the main strength of FuseIC3, and uses features present in modern SAT solvers. It adds "just enough" extra information to the saved reachable states to enable reuse. We demonstrate the industrial scalability of FuseIC3 on a large set of 1,620 real-life models for the NASA NextGen air traffic control system [19, 23], selected benchmarks from HWMCC 2015 [1], and a set of seven models for the Boeing AIR6110 wheel braking system [5]. Our experiments evaluate FuseIC3 along two dimensions; checking all models with the same property, and checking each model with several properties. Lastly, we evaluate the effect of model relatedness on the performance of FuseIC3.

Related Work The idea of reusing model-checking information, like variable orderings, between runs has been extensively used in BDD-based model checking leading to substantial performance improvement [3, 27]. Similarly, intermediate SAT solver clauses and interpolants are reused in bounded model checking [22, 25]. Reusing learned invariants in IC3 speeds up convergence of the algorithm [8]. These techniques enable efficient incremental model checking and are useful in *regression verification* [28] and *coverage computation* [9]. FuseIC3 is an

Artifacts for reproducibility, code, theorem proofs, and detailed experimental results can be found at http://temporallogic.org/research/FMCAD17. Thanks to NSF CAREER Award CNS-1552934 for supporting this work.

incremental algorithm and is applicable in these scenarios.

Product line verification techniques, e.g., with Software Product Lines (SPL), also verify models describing large design spaces [4, 13, 15, 16]. The several *instances* of feature transition systems (FTS) [14] describe a set of models. FuseIC3 relaxes this requirement and can be used to check models that cannot be combined into a FTS. It outputs model-checking results for every model-property pair in the design space without dependence on any *feature*. Nevertheless, SPL instances can be checked using FuseIC3. Large design spaces can also be generated by models that are parametric over a set of inputs. *Parameter synthesis* [10] can generate the many models in a design space that can be checked using FuseIC3. The parameterized model-checking problem [18] deals with infinite homogeneous models. In our case, the models in a set are heterogeneous and finite.

The work most closely related to ours is a state-of-the-art algorithm for incremental verification of hardware [8]. It extends IC3 to reuse the generated proof, or counterexample, in future checker runs. It extracts minimal inductive subclauses from an earlier invariant with respect to the current model. In our analysis, we compare FuseIC3 with this algorithm, and show that with the same amount of information storage, FuseIC3 is faster when checking large design spaces.

Contributions The contributions of our work are manyfold. We present a query-efficient SAT-based algorithm for checking large design spaces, and incremental verification. The algorithm is fully automated, general, and scalable. To the best of our knowledge, FuseIC3 is the first algorithm to reuse reachable state approximations to guide bad-state search in IC3. Our novel procedure to repair state approximations requires little computation effort and is of individual interest. We present an extensive experimental analysis using real-life benchmarks. Lastly, we make all reproducibility artifacts and source code publicly available.

Structure Section II details background information, overviews the typical IC3 algorithm, and defines the notation used throughout the paper. Section III presents the FuseIC3 algorithm. A large-scale experimental evaluation forms Section IV, and Section V concludes by highlighting future work and possible extensions.

II. Preliminaries

A. Definitions

A Boolean transition system, or model M is represented using the tuple (Σ, Q, Q_0, δ) where $Q_0 \subseteq Q$ is the set of *initial states* and δ is the *transition relation* over state variables Σ. A *safety property* is a predicate φ over Σ. A primed variable σ', such that $\sigma \in \Sigma$, represents σ in the next time step. If ψ is a Boolean formula over Σ, ψ' is obtained by replacing each variable in ψ with the corresponding primed variable.

A sequence of states $s_0, s_1, \ldots, s_n$ is a *path* in M if s_0 is an initial state, each $s_i \in Q$ for $0 \leq i \leq n$, and for $0 < i < n$, $(s_i, s_{i+1}) \in \delta$, i.e., there is a valid transition from s_i to s_{i+1}. A state t in a model is *reachable* if there exists an execution path such that $s_n = t$. A model M *satisfies* safety property φ, denoted $M \models \varphi$, when no reachable states of M intersect $\neg\varphi$.

The state variables and their negations are called *literals*. A disjunction of literals is called a *clause*. A Boolean formula containing a conjunction of clauses is said to be in *Conjunctive Normal Form* (CNF).

We assume that a Boolean formula ψ over Σ represents a set of states in M, or $\psi \subseteq Q$. Two Boolean formulas ψ_1 and ψ_2 over Σ *overlap* if $\psi_1 \cap \psi_2 \neq \emptyset$, i.e., they contain common symbolic states. Models M and N are *related* if they contain overlapping reachable states. A *set of models* is a collection of such related models.

B. Overview of IC3

IC3/PDR [6, 17, 26] is a novel verification method based on property directed invariant generation. Given a model $M = (\Sigma, Q, Q_0, \delta)$, and a safety property φ, IC3 incrementally generates an inductive strengthening of φ to prove whether $M \models \varphi$. It maintains a sequence of frames $S_0 = Q_0, S_1, \ldots S_k$ such that each S_i, for $0 < i < k$, satisfies φ and is an over-approximation of states reachable in i-steps or less. If two adjacent frames become equivalent, IC3 has found an inductive invariant and the property holds for the model. If a state violating the property is reachable, a counterexample trace is returned. Throughout IC3's execution, it maintains the following invariants on the sequence of frames:

1) for $i > 0$, S_i is a conjunction of clauses,
2) $S_{i+1} \subseteq S_i$,
3) $S_i \wedge \delta \Rightarrow S'_{i+1}$, and
4) for $i < k$, $S_i \Rightarrow \varphi$.

Each clause added to the frames is an intermediate lemma constructed by IC3 to prove whether $M \models \varphi$. The algorithm proceeds in two phases: a *blocking* phase, and a *propagation* phase. In the blocking phase, S_k is checked for intersection with $\neg\varphi$. If an intersection is found, S_k violates φ. IC3 continues by recursively blocking the intersecting state at S_{k-1}, and so on. If at any point, IC3 finds an intersection with S_0, $M \not\models \varphi$ and a counterexample can be extracted. The propagation phase moves forward the clauses from preceding S_i to S_{i+1}, for $0 < i \leq k$. During propagation, if two consecutive frames become equal, a fix-point has been found and IC3 terminates. The fix-point $\mathcal{I}$ represents the inductive strengthening of φ and has the following properties: $Q_0 \Rightarrow \mathcal{I}$, $\mathcal{I} \wedge \delta \Rightarrow \mathcal{I}'$, and $\mathcal{I} \Rightarrow \varphi$. We refer the reader to [7, 20] for lower-level details of IC3.

C. SAT with Assumptions

In our formulation, we consider SAT queries of the form $\mathsf{sat}(\varphi, \gamma)$, where φ is a CNF formula, and γ is a set of assumption clauses. A query with no assumptions is simply written as $\mathsf{sat}(\varphi)$. Essentially, the query $\mathsf{sat}(\varphi, \gamma)$ is equivalent to $\mathsf{sat}(\varphi \wedge \gamma)$ but the implementation is typically more efficient. If $\varphi \wedge \gamma$ is:

1) SAT, get-sat-model() returns a satisfying assignment.

2) UNSAT, get-unsat-assumptions() returns a unsatisfiable core β of the assumption clauses γ, such that $\beta \subseteq \gamma$, and $\varphi \wedge \beta$ is UNSAT.

We abstract the implementation details of the underlying SAT solver, and assume interaction using the above functions.

D. Notation

We reduce the task of verifying a set of models by restricting the description of our algorithm to two related models $M = (\Sigma, Q_M, Q_{0_M}, \delta_M)$ and $N = (\Sigma, Q_N, Q_{0_N}, \delta_N)$ in the set. Each model has to be checked against a safety property φ. Assume that model M is checked first. The algorithm computes frame sequence R and S for M and N, respectively. $|R|$ denotes number of frames in the sequence R.

III. Algorithm

In this section, we present the main contribution of our paper, FuseIC3. We start with the core idea behind the algorithm by giving the intuition behind recycling IC3-generated intermediate lemmas. We then provide a general overview of different sub-algorithms that help FuseIC3 achieve it's performance. We next describe the two main components: *basic check* and *frame repair* of FuseIC3.

A. Intuition

Recall that frames computed by IC3 represent over-approximated states. When M is checked with IC3, frames $R_0, R_1, \ldots, R_j$, are computed such that $R_i \wedge \delta_M \Rightarrow R'_{i+1}$ for $i < j$ (invariant 3, Section II-B). In the classical case, checking N after M requires resetting and restarting IC3, which then computes frames $S_0, S_1, \ldots, S_k$ for N. Due to to the reset, all intermediate lemmas are lost and verification for N has to start from the beginning. However, since M and N are related, the frames for M and N overlap, and therefore, frames for M can be recycled and potentially reused in the verification for N. The idea is illustrated using Venn diagrams in Fig. 1.

In Fig. 1a, the parallelogram and ellipse represent clauses c_1 and c_2, respectively, in frame R_{i+1} such that $R_{i+1} = c_1 \wedge c_2$, and the triangle represents states reachable from R_i in one step, i.e., $R_i \wedge \delta_M$. So, $R_i \wedge \delta_M \Rightarrow R'_{i+1}$. Now consider a scenario in which we recycle the clauses in R_{i+1} when verifying N. The triangle and the rectangle in Fig. 1b represent the states reachable from S_i in one step. If we were to make $S_{i+1} = R_{i+1}$, we end up with $S_i \wedge \delta_N \not\Rightarrow S'_{i+1}$ since c_1 doesn't contain some states reachable from S_i. Therefore, we have to modify c_1 such that the invariant holds. Fig. 1c and 1d show the two possible modifications of c_1. In the former case, we add states $(S_i \wedge \delta_N) \setminus c_1$ to c_1. In the latter, we over-approximate c_1 to $\hat{c}_1$ such that $S_i \wedge \delta_N \Rightarrow \hat{c}_1$ (a trivial over-approximation is to make c_1 equal to the set of all states). Irrespective of the approach used, we end up with $S_i \wedge \delta_N \Rightarrow \hat{R}'_{i+1} = S'_{i+1}$, where $\hat{R}_{i+1} = \hat{c}_1 \wedge c_2$. Then we check the $(i+1)$-th step over-approximation for intersection with $\neg\varphi$ and IC3 continues. In this way, reusing clauses from model M, saves a lot of effort in rediscovering these clauses for model N. FuseIC3 uses state over-approximations.

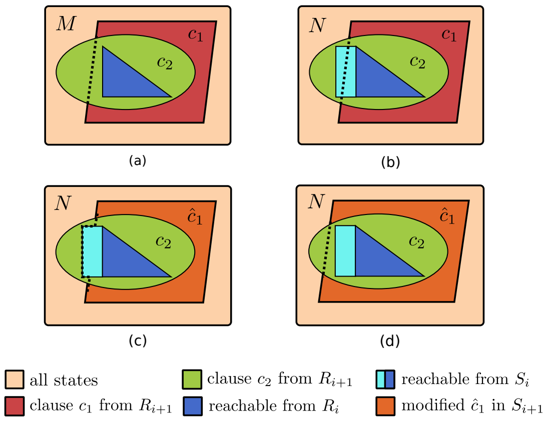

M
c1
c2
(a)
N
c1
c2
(b)
N
ĉ1
c2
(c)
N
ĉ1
c2
(d)
all states
clause c_1 from R_{i+1}
clause c_2 from R_{i+1}
reachable from R_i
reachable from S_i
modified $\hat{c}_1$ in S_{i+1}

Fig. 1. Intuition behind repairing frames computed for one model by IC3, and reusing them for checking another related model.

```
bool FuseIC3 ($Q_0$, $\delta$, $\varphi$)
 1: if CheckInvar($Q_0$, $\delta$, last_invariant, $\varphi$) : return true
 2: if SimulateCex($Q_0$, $\delta$, last_cex, $\varphi$) : return false
 3: $k \leftarrow 0$, $S_k \leftarrow Q_0$   # first frame is initial state
 4: while true :   # main FuseIC3 loop
 5:     while sat($S_k \wedge \neg\varphi$) :   # blocking phase
 6:         $s \leftarrow$ get-sat-model()
 7:         if not recursive_block($s, k$) :
 8:             last_cex ← extract_cex(), return false
 9:     $k \leftarrow k + 1$
10:     $S_k \leftarrow$ FrameRepair($k - 1$)
11:     for $i \leftarrow 1$ to $k - 1$ :   # propagation phase
12:         for each new clause $c \in S_i$ :
13:             if not sat($S_i \wedge c \wedge \delta \wedge \neg c'$) : add $c$ to $S_{i+1}$
14:         if $S_i \equiv S_{i+1}$ :   # found fix-point invariant
15:             last_invariant ← $S_i$, return true
frame FrameRepair (int i)
 1: if not sat($S_i \wedge \delta \wedge \neg R'_{i+1}$) : return $R_{i+1}$
 2: $\mathcal{G} \leftarrow$ FindClauses($S_i$, $\delta$, $R_{i+1}$)
 3: $\hat{R}_{i+1} \leftarrow R_{i+1} \setminus \mathcal{G}$
 4: for each clause $c \in \mathcal{G}$ :
 5:     $\hat{c} \leftarrow$ ExpandClause($S_i$, $\delta$, $c$)
 6:     $\hat{c} \leftarrow$ ShrinkClause($S_i$, $\delta$, $c$, $\hat{c}$)
 7:     $\hat{R}_{i+1} \leftarrow \hat{R}_{i+1} \wedge \hat{c}$
 8: return $\hat{R}_{i+1}$   # repaired frame $R_{i+1}$
```

Fig. 2. High-level description of FuseIC3. Parts of the algorithm for typical IC3 are based on the description in [17, 20].

B. Overview

FuseIC3 is a bidirectional reachability algorithm. It uses forward reachability to reuse frames from a previously-checked related model, and IC3-type backward reachability to recursively block predecessors to bad states. The algorithm description appears in Fig. 2.

FuseIC3 takes as input the initial states Q_0 and the transition relation δ for the current model, and a safety property φ. The internal state maintained by the algorithm is last_invariant, last_cex, and the frames R computed for the last model verified. Initially, the state is empty. Lines 1–2 perform basic checks in an attempt to reuse proofs from an earlier run to verify the current model. Lines 4–15 loop until an invariant or a counterexample is found. FuseIC3 maintains a sequence of frames $S_0, S_1, \ldots, S_k$ for the current model being checked. Whenever a new frame S_k is introduced in line 10, the algorithm reuses a frame from R after repairing it with FRAMEREPAIR. The repaired frame is added to S_k, which after propagation in lines 11–15, is checked for intersection with a bad state. A typical execution of IC3 follows until a new frame is introduced. Upon termination, R is replaced with the current set of frames S, and last_invariant and last_cex are updated accordingly.

FRAMEREPAIR takes as input an integer i. It checks if R_{i+1} can be used as is in line 1. If yes, R_{i+1} is returned. Otherwise, the frame is repaired in lines 2–7. FINDCLAUSES finds violating clauses in R_{i+1}. Each of these clauses is repaired in lines 4–7 using EXPANDCLAUSE and SHRINKCLAUSE. After repair, the updated frame $\hat{R}_{i+1}$ is returned.

The models in a set are checked sequentially. When FuseIC3 is run on the first model in the set, it reduces to running typical IC3. During propagation and when $k < |R|$, only repaired clauses (from FRAMEREPAIR) and discovered clauses for the current model are propagated. When $k \geq |R|$, FRAMEREPAIR returns an empty frame and all clauses from earlier frames take part in propagation.

C. Basic Checks

It is possible that the changes in design between two models are very small, and are outside the cone-of-influence of the verification procedure. Therefore, although the models are different, they might have the same over-approximated inductive invariant with respect to the property being checked. A similar argument applies for two models that fail a property. In this case, a counterexample for the first model might be a valid counterexample for the second model. Both these checks can be carried out in very little time as explained below. For the case when M and N have different state variables, cone-of-influence with respect to variables in N is applied on the invariant/counterexample before performing the checks.

a) Inductive Invariant: If $\mathcal{I}_M$ is an inductive invariant for M with respect to φ, it satisfies the following three conditions:

1) $Q_{0_M} \Rightarrow \mathcal{I}_M$,
2) $\mathcal{I}_M \wedge \delta_M \Rightarrow \mathcal{I}'_M$, and
3) $\mathcal{I}_M \Rightarrow \varphi$.

If changes in N are outside the cone-of-influence of $\mathcal{I}_M$, then $N \models \varphi$ if the above conditions hold for N (checked using three SAT calls).

b) Counterexample Trace: If $M \not\models \varphi$, IC3 generates a counterexample trace $s_0, s_1, \ldots s_k$ such that

1) $s_0 \in Q_{0_M}$,
2) $(s_i, s_{i+1}) \in \delta_M$ for $i < k$, and
3) $s_k \in \neg\varphi$.

Simulate the counterexample trace for M on N and check if it satisfies the above three conditions (using $k+1$ SAT calls). If the conditions are satisfied, conclude that $N \not\models \varphi$.

To summarize, if changes in two subsequent models are outside the cone-of-influence of the proofs generated by IC3,

```
bool CHECKINVARIANT ($Q_0$, $\delta$, invariant $\mathcal{I}$, $\varphi$)
1: if not sat($Q_0 \wedge \neg\mathcal{I}$) and not sat($\mathcal{I} \wedge \delta \wedge \neg\mathcal{I}'$) and not
   sat($\mathcal{I} \wedge \neg\varphi$) : return true
2: else return false

bool SIMULATECEX ($Q_0$, $\delta$, trace $s$, $\varphi$)
1: if not sat($s_0 \wedge Q_0$) : return false
2: if not sat($s_k \wedge \neg\varphi$) : return false
3: for $i \leftarrow 0$ to len($s$) :
4:     if not sat($s_i \wedge \delta \wedge s'_{i+1}$) : return false
5: return true  # valid counterexample
```

Fig. 3. CHECKINVARIANT evaluates the last known invariant against the current model, and returns *true* if invariant holds, otherwise *false*. SIMULATECEX simulates the last known counterexample on the current model, and returns *true* if successful, otherwise, *false*.

```
FINDCLAUSES (frame $S$, $\delta$, frame $R$)
1: for each clause $c_i \in R$ :  # configure solver assertions
2:     introduce auxiliary variable $y_i$
3:     for each literal $l \in c'_i$ :
4:         add assertion $\neg l \vee y_i$ to solver
5: $\mathcal{G} \leftarrow \emptyset$  # set is initially empty
6: while sat($S \wedge \delta, (\neg y_1 \vee \neg y_2 \vee \ldots \vee \neg y_k)$) :
7:     $\alpha \leftarrow$ get-sat-model()
8:     for each $y_1, y_2, \ldots y_k$ :
9:         if $\alpha(y_i)$ == $\perp$ :
10:            add $c_i$ to $\mathcal{G}$ and remove $y_i$ from sat query
11: return $\mathcal{G}$  # set of violating clauses
```

Fig. 4. FINDCLAUSES algorithm to find all clauses in R that lead to violation of $S \wedge \delta \not\Rightarrow R'$. Upon termination, $\mathcal{G}$ contains violating clauses.

verification completes almost instantly. The pseudo-code for these two basic checks is given in Fig. 3.

D. Frame Repair

We want to expand clauses in frame R_{i+1} that are responsible for the violation of $S_i \wedge \delta_N \Rightarrow R'_{i+1}$. The satisfiability model is a pair of states (a, b) such that $a \in S_i$, $b \notin R_{i+1}$, and $(a, b) \in \delta_M$. In other words, b is missing from some, or all clauses in R_{i+1}. If all such missing states are added to clauses in R_{i+1}, resulting in $\hat{R}_{i+1}$, the condition $S_i \wedge \delta_N \Rightarrow \hat{R}'_{i+1}$ becomes valid and $\hat{R}_{i+1}$ can be reused in checking N. Adding these states one-by-one requires several calls to the underlying SAT solver and is infeasible in practice (reduces to all-SAT). Instead, the violating clauses in R_{i+1} are over-approximated. The over-approximation ends up adding several states to R_{i+1} that are in the post-image of multiple states in S_i. As the first step in repairing the frame, we want to find all such violating clauses.

Find Violating Clauses: Let's assume frame R_{i+1} is composed of clauses $C = \{c_1, c_2, \ldots c_n\}$. There are clauses $\mathcal{G} \subseteq C$ such that the assertion $S_i \wedge \delta_N \Rightarrow c'$ is violated for all $c \in \mathcal{G}$. Set $\mathcal{G}$ can be found by brute-forcing the assertion check for all clauses in C. However, such an approach doesn't scale since IC3 frames can have thousands of clauses. Algorithm FINDCLAUSES, which is inspired by the *Invariant Finder* algorithm in [8], efficiently finds such violating clauses. The pseudo-code for the algorithm is given in Fig. 4.

```
EXPANDCLAUSE (frame $S$, $\delta$, clause $\hat{c}$)
 1: $v \leftarrow$ all primed variables in $\delta$
 2: $l \leftarrow$ all variables in clause $\hat{c}'$
 3: $B \leftarrow v \setminus l$  # variables not in clause $\hat{c}$
 4: while $|B| > 0$ and sat($S \wedge \delta \wedge \neg\hat{c}'$) :
 5:     $\alpha \leftarrow$ get-sat-model()
 6:     randomly pick any $b' \in B$
 7:     if $\alpha(b') == \top$ : add $b$ to clause $\hat{c}$
 8:     else if $\alpha(b') == \bot$ : add $\neg b$ to clause $\hat{c}$
 9:     remove $b'$ from $B$
10: if sat($S \wedge \delta \wedge \neg\hat{c}'$) : return $\emptyset$
11: return $\hat{c}$  # expanded clause; $S \wedge \delta \Rightarrow \hat{c}'$
```

Fig. 5. EXPANDCLAUSE algorithm to add literals to clause c such that $S \wedge \delta \Rightarrow \hat{c}'$. Upon termination, an empty set is returned if expansion fails.

FINDCLAUSES takes as input frame $S = S_i$, transition relation $\delta = \delta_N$, and frame $R = R_{i+1}$. Upon termination, it returns all violating clauses. An auxiliary variable y_i is introduced for each clause c_i in R in line 2. Lines 3–4 are equivalent to adding the assertion $c_i \Rightarrow y_i$ to the solver. Lines 6–10 loop until the query in line 6 is SAT. On every iteration of the loop, there is at least one y_i that is assigned *false*. Clauses c_i corresponding to all such y_i are added to $\mathcal{G}$ and y_i is removed from the query. When the query becomes UNSAT, $\mathcal{G}$ contains all violating clauses in R, and is returned. In practice, multiple y_i are assigned *false* which helps terminate the loop faster.

Lemma 1. FINDCLAUSES *returns all clauses in R_{i+1} that are responsible for $S_i \wedge \delta \not\Rightarrow R'_{i+1}$.*

After discovering all violating clauses, FuseIC3 attempts to expand them before reusing R_{i+1} to check model N. In the trivial case, each violating clause can be removed from R_{i+1} altogether. However, doing this is quite wasteful. For example, consider a frame in which all clauses are violating. Reusing this frame entails restarting IC3 from an empty frame, a scenario we want to avoid. Instead, we rely on efficient use of the SAT solver to over-approximate the violating clauses.

Expand Violating Clauses: A clause c is violating if none of its literals match the literals in state b (recall the model (a, b) to the SAT query $S_i \wedge \delta_N \Rightarrow R'_{i+1}$). If any literal from b is added to c, resulting in $\hat{c}$, then $b \in \hat{c}$. Fundamentally, we want to add literals to clause c without actually enumerating all such b such that the assertion $S_i \wedge \delta_N \Rightarrow \hat{c}'$ holds. A literal can be added as is, or in its negated form. Adding both makes the assertion trivially valid. For example, consider a system with variables x, y, z, and a violating clause $c = (x \vee y)$. Our aim is to add states to c. Either z or $\neg z$ can be added to c, but not both. However, deciding what to add to make the assertion valid is beyond the scope of a SAT solver. Instead, we use an efficient randomized algorithm, EXPANDCLAUSE, to add literals to clause c. The pseudo-code for the algorithm is given in Fig. 5.

EXPANDCLAUSE takes as input frame $S = S_i$, transition relation $\delta = \delta_N$, and the violating clause $c \in R_{i+1}$. Initially, $\hat{c} = c$. Lines 1–3 find all variables that are missing from c and store them in set B. The loop in lines 4–9 is repeated until set B becomes empty, or the query $S \wedge \delta \Rightarrow \hat{c}'$ becomes valid. In the latter case, enough literals have been added to expand c and the algorithm can terminate. From the SAT model α, randomly pick an assignment to a variable in B. If the assignment is *true*, add the variable as is to $\hat{c}$, otherwise, negate variable and add to $\hat{c}$. The added variable is removed from B and the loop continues. When all possible variables have been added to $\hat{c}$ and the assertion is still SAT, return $\hat{c}$ to be the empty clause ($c =$ *true*, or set of all states) in line 10.

```
SHRINKCLAUSE (frame $S$, $\delta$, clause $c$, clause $\hat{c}$)
    assert(not sat($S \wedge \delta \wedge \neg\hat{c}'$))
 1: $v \leftarrow$ {literals in $\hat{c}$} $\setminus$ {literals in $c$}
 2: for each $l \in v$ :
 3:     $g \leftarrow v \setminus l$  # drop literal $l$
 4:     if not sat($S \wedge \delta \wedge \neg c', \neg g'$) :
 5:         $v \leftarrow$ {literals $j$ | $j' \in$ get-unsat-assumptions()}
 6: return $c \vee \bigvee${literals in $v$}
```

Fig. 6. SHRINKCLAUSE algorithm to remove excess literals from clause c while maintaining $S \wedge \delta \Rightarrow c'$.

Lemma 2. EXPANDCLAUSE *expands clause c to $\hat{c}$ such that the assertion $S_i \wedge \delta \Rightarrow \hat{c}'$ is valid.*

Shrink Expanded Clauses: Due to the nature of the randomized algorithm, we may end up adding more states than required to the expanded clauses. As a last step in repairing the frame, we remove the excess states added from all such clauses, albeit, maintaining the over-approximation. FuseIC3 uses UNSAT assumptions generated in the proof for $S_i \wedge \delta \Rightarrow \hat{c}'$ to shrink clause $\hat{c}$. The SHRINKCLAUSE algorithm tries dropping a subset of the newly added literals from $\hat{c}$. The pseudo-code for the algorithm is given in Fig. 6.

SHRINKCLAUSE takes as input frame $S = S_i$, transition relation $\delta = \delta_N$, non-expanded violating clause c, and the expanded non-violating clause $\hat{c}$. Set v contains all literals that were added to c by EXPANDCLAUSE. Lines 2–5 loop until enough literals have been dropped from $\hat{c}$ such that the $S_i \wedge \delta_N \wedge \neg c' \wedge \neg v'$ is valid. On each iteration of the loop, a literal l to drop from v is chosen. If the assertion is UNSAT, we can successfully drop l from v, and replace v with the UNSAT assumptions in the query. However, if the assertion is SAT, l is a required literal in v and we try dropping another literal.

Lemma 3. SHRINKCLAUSE *removes all possible literals from $\hat{c}$ such that the assertion $S_i \wedge \delta \Rightarrow \hat{c}'$ is valid.*

The violating clause may appear in future frames in R (due to the propagation phase when checking M). The modification is reflected in all occurrences of the clause. All such violating clauses in R_{i+1} are repaired.

Theorem 1. FRAMEREPAIR *returns repaired frame $\hat{R}_{i+1}$ such that $S_i \wedge \delta \Rightarrow \hat{R}'_{i+1}$ is valid.*

The repaired frame $\hat{R}_{i+1}$ is added to the set of frames for N at step $i+1$. Therefore, $S_{i+1} = \hat{R}_{i+1}$, and we continue with the normal execution of IC3. Clauses are propagated from frames

S_j, for $j \leq i$, to S_{i+1}, which is then checked for intersection with bad states and, if any are found, IC3 tries recursively blocking them at earlier steps.

IV. Experimental Analysis

We extensively experimentally analyze FuseIC3. We summarize the setup used for the experiments, briefly detail our benchmarks, and end with experimental results.

A. Setup

FuseIC3 is coded in C++ and uses MathSAT5 [11] as the underlying SAT solver. It takes SMV models or AIGER files as input. The IC3 part of FuseIC3 is based on the description in [17] and `ic3ia`.[1] We compare the performance of FuseIC3 with typical IC3 (typ), and incremental IC3 (inc). The algorithm for incremental IC3 is part of IBM's RuleBase model checker [3]. We coded inc based on the description in [8] to the best of our understanding. All experiments were performed on Iowa State University's Condo Cluster comprising of nodes having two 2.6GHz 8-core Intel E5-2640 processors, 128 GB memory, and running Enterprise Linux 7.3. Each model-checking run had exclusive access to a node.

B. Benchmarks

We evaluated FuseIC3 over a large collection of challenging benchmarks. The benchmarks are derived from real-world case studies and modified benchmarks from HWMCC 2015.

1) Air Traffic Controller (ATC) Models: are a large set of 1,620 real-world models representing different possible designs for NASA's NextGen air traffic control (ATC) system [19]. The set of models are generated from a contract-based, parameterized NUXMV model. Each model is checked against 34 safety properties. The entire evaluation consists of 34 model-sets (one for each property) containing 1,620 models.

2) Selected Benchmarks from HWMCC 2015: We considered a total of 548 benchmark models from the single safety property track [1]. Of the 548, 110 models were solved using our implementation of IC3 within a timeout of 5 minutes. To create a model-set, we generated 200 mutations of each of the 110 benchmarks. The original model was mutated to only modify the transition system, and not the safety property implicit in the AIGER file; 1% of the assignments were randomly modified. An assignment of the form $g = g_1 \wedge g_2$ was selected with probability 0.01 and changed to $g = 0$, $g = 1$, $g = \neg g_1 \wedge g_2$, $g = g_1 \wedge \neg g_2$, $g = \neg g_1 \wedge \neg g_2$, $g = g_1 \wedge g_2$, $g = g_1$, $g = \neg g_1$, $g = g_2$, or $g = \neg g_2$, with equal probability. Therefore, the full evaluation consists of 110 model-sets, each consisting of one property and 200 models.

3) Wheel Braking System (WBS) Models: are a set of seven real-world models representing possible designs for the Boeing AIR6110 wheel braking system [5]. Each model is checked against ∼300 safety properties. However, the properties checked for each model are not the same. We evaluate FuseIC3 using this benchmark to measure performance when a model is checked against several related or similar properties. Each model was checked using a timeout of 120 minutes.

[1]https://es-static.fbk.eu/people/griggio/ic3ia/

TABLE I
Summary of results for 34 sets of 1,620 models each for NASA Air Traffic Control System.

Algorithm	Cumulative Time in minutes	Median Speedup	
		v/s typ (avg)	v/s inc (avg)
Typical IC3 (typ)	2502.70	-	-
Incremental IC3 (inc)	2180.57	1.29 (1.3)	-
FuseIC3	**1683.53**	**1.75** (5.48)	**1.34** (3.67)

C. Results

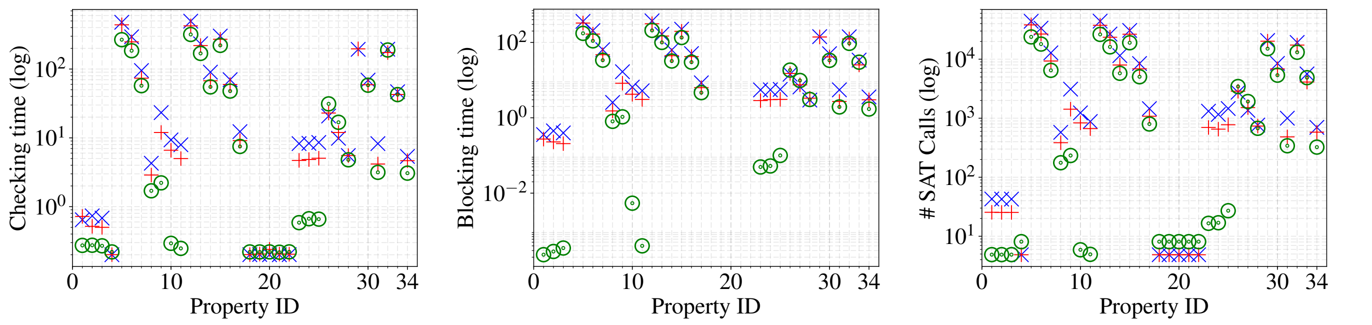
1) Air Traffic Controller (ATC) Models: Each of the 34 model-sets were checked using a timeout of 720 minutes per algorithm. The models in a set were checked in random order. Table I gives a summary of the results. FuseIC3 is median 1.75× (average 5.48×) faster compared to typical IC3, and median 1.34× (average 3.67×) faster compared to incremental IC3. On the other hand, incremental IC3 is median 1.29× (average 1.3×) faster than typical IC3.

Fig. 7a shows time taken by the algorithms on each model-set. FuseIC3 is almost always faster than typical IC3, and incremental IC3. However, for some very small instances, typical IC3 is faster; both incremental IC3 and FuseIC3 require a certain overhead in extracting information from the last checker run. FuseIC3 tries minimizing the time spent in exploring the common state space between models. In terms of the IC3 algorithm, this relates to time spent in finding bad states and blocking them at earlier steps (blocking phase). Fig. 7b shows time taken by each algorithm in blocking discovered bad states. FuseIC3 spends considerably less time in the blocking phase compared to typical IC3 and incremental IC3. Therefore, FuseIC3 is successful in reusing a major part of the already-discovered state space between different checker runs, a major requirement when checking large design spaces. Fig. 7c shows the total number of calls made to the underlying SAT solver by each algorithm. FuseIC3 makes fewer SAT calls and takes less time to check each model-set. For small models, FuseIC3 makes more SAT calls compared to typical IC3.

2) Benchmarks from HWMCC 2015: Each of the 110 model-sets were checked using a timeout of 120 minutes per algorithm. The models in a set were checked in random order. 91 of 110 model-sets were solved by all algorithms within the timeout. Incremental IC3 solved two more model-sets compared to typical IC3, while FuseIC3 solved five more compared to typical IC3. Table II gives a summary of results.

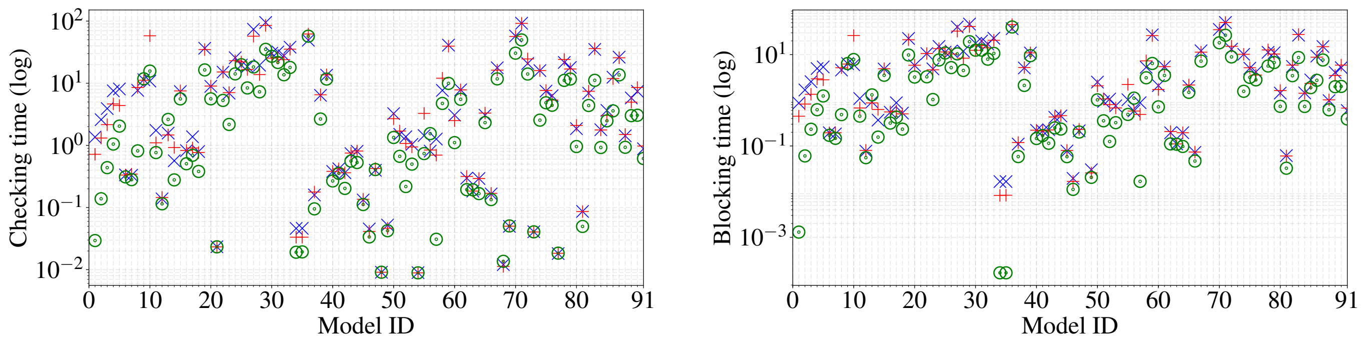
Fig. 8a shows time taken by the algorithms in checking each benchmark model-set. FuseIC3 is median 1.75× (average 3.18×) faster than typical IC3, and median 1.72× (average 2.56×) faster than incremental IC3. Significant speedup is achieved when checking model-sets containing large models with FuseIC3. Performance for model-sets containing small models is similar for all algorithms. Fig. 8b shows time spent by each algorithm in blocking predecessors to bad states.

To estimate performance of FuseIC3 on model-sets with varying degree of overlap among models, we picked the `bobtuint18neg` benchmark from HWMCC 2015. 100



(a) Cumulative checking time per set (minutes) (b) Cumulative blocking time per set (minutes) (c) Number of SAT calls per set

Fig. 7. Comparison between IC3 (×), incremental IC3 (+), and FuseIC3 (⊙) on NASA Air Traffic Control System models. There are a total of 34 properties. 1,620 models are checked per property. A point represents cumulative time taken to check all models for a property by an algorithm.



(a) Cumulative checking time per design space (minutes) (b) Cumulative blocking time per design space (minutes)

Fig. 8. Comparison between IC3 (×), incremental IC3 (+), and FuseIC3 (⊙) on 91 benchmarks from HWMCC 2015. Each model is converted to a model-set containing 200 models, generated by 1% mutation of the original. A point represents cumulative time for checking all mutated versions of a model.

TABLE II
SUMMARY OF RESULTS FOR 91 OF 110 SETS OF 200 MODELS EACH FOR SELECTED HWMCC 2015 BENCHMARKS.

Algorithm	Cumulative Time in minutes	Median Speedup	
		v/s typ (avg)	v/s inc (avg)
Typical IC3 (typ)	1024.60	-	-
Incremental IC3 (inc)	1026.30	1.04 (1.07)	-
FuseIC3	**545.31**	**1.75** (3.18)	**1.72** (2.56)

model-sets with varying degrees of mutation, between 0.5% to 50%, of the original model were generated. Each model-set consists of 100 models each. Each set was checked using a timeout of 60 minutes per algorithm. Model-sets corresponding to higher mutation values time out (SAT solvers are tuned for practical designs and random mutations create SAT instances that don't always correspond to real designs). However, FuseIC3 is able to verify more models in a set for almost all mutation percentages. Fig. 9 gives a summary of the adjusted relative speedup between checking using FuseIC3 versus typical IC3. Even at higher mutation percentages, FuseIC3 is significantly faster.

3) Wheel Braking System Models: A model in the design space was checked against several properties, differently from the other benchmarks that checked all models in a set with the same property. Each model was checked using a timeout of 120 minutes. The properties for each model were checked

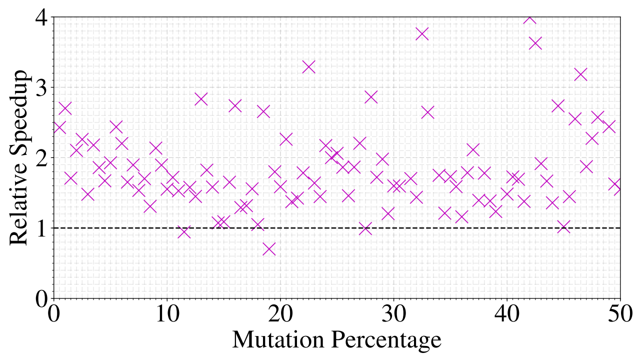


Fig. 9. Adjusted relative speedup between checking using FuseIC3 versus typical IC3. 100 models are generated for every mutation percentage between 0.5% to 50% in steps of 0.5%, and are checked against the same property.

in random order. Table III gives a summary of the results.

Compared to other benchmarks, FuseIC3 achieves a smaller speedup when checking the WBS models. Although some properties being checked for the models are similar, i.e., the bad states representing the negation of the property overlap, the order in which they are checked greatly influences the performance of FuseIC3. In the random ordering used for the experiment, FuseIC3 is able to reuse frames without any repair (the same model is being checked), however, it spends a lot of time in blocking predecessors to bad states. Nevertheless, it is faster than checking all properties on a model using typical

TABLE III
COMPARISON BETWEEN TYPICAL IC3, INCREMENTAL IC3, AND FUSEIC3 FOR AIR6110 WHEEL BRAKING SYSTEM (TIME IS IN MINUTES).

	Typical IC3	Incremental IC3		FuseIC3		
Model	Time	Time	v/s typ	Time	v/s typ	v/s inc
M_1	4.36	5.02	0.87	**3.72**	1.17	1.35
M_2	15.78	16.65	0.95	**14.80**	1.07	1.13
M_3	12.43	13.48	0.92	**11.24**	1.11	1.20
M_4	12.45	13.66	0.91	**11.09**	1.12	1.23
M_5	15.92	17.04	0.93	**14.71**	1.08	1.16
M_6	**16.85**	17.79	0.95	17.04	0.99	1.04
M_7	12.95	13.67	0.95	**12.12**	1.07	1.13
	90.73 (total)	97.31 (total)	0.95 (median)	84.72 (total)	1.11 (median)	1.20 (median)

IC3. On the other hand, incremental IC3 is slower compared to typical IC3. It is able to extract the minimal inductive invariant (invariant finder) instantly, however, suffers from the same problem as FuseIC3. Incremental IC3, and FuseIC3 will benefit if similar properties are checked in order.

V. CONCLUSIONS AND FUTURE WORK

FuseIC3, a SAT-query efficient algorithm, significantly speeds up model checking of large design spaces. It extends IC3 to minimize time spent in exploring the state space in common between related models. FuseIC3 spends less time during the blocking phase (Fig. 7b and Fig. 8b) due to success in reusing several clauses, has to learn fewer new clauses, and makes fewer SAT queries. The smallest salvageable unit in FuseIC3 is a clause; due to this granularity, FuseIC3 is able to selectively reuse stored information and is faster than the state-of-the-art algorithms that rely on reusing a coarser CNF invariant [8]. FuseIC3 is industrially applicable and scalable as witnessed by its superior performance on a real-life set of 1,620 NASA air traffic control system models (achieving an average 5.48× speedup), and benchmarks from HWMCC 2015 (achieving an average 3.18× speedup). Despite spending significant time in learning new clauses for the Boeing wheel braking system models, FuseIC3 is still faster than the previous best algorithm, typical IC3, when checking properties in random order; FuseIC3's performance will improve if similar properties are checked in order. We contribute to the available benchmarks by releasing all artifacts for reproducibility.

Ordering of models and properties in the design space improves the performance of FuseIC3, much like variable ordering in BDDs. Heuristics for optimizing model ordering are a promising topic for future work. Preprocessing the models and properties, based on knowledge about the design space, before checking them with FuseIC3 may remove redundancies in the design space. We plan to extend FuseIC3 to checking liveness properties by using it is a safety checker [12]. We also to plan to investigate extending FuseIC3 to reuse intermediate results of SAT queries, generalized clauses, and IC3 proof obligations across models. Finally, since checking large design spaces is becoming commonplace, we plan to develop more model-set benchmarks and make them publicly available.

REFERENCES

[1] "HWMCC 2015," http://fmv.jku.at/hwmcc15/.

[2] C. Bauer, K. Lagadec, C. Bès, and M. Mongeau, "Flight control system architecture optimization for fly-by-wire airliners," *J. Guidance, Control, and Dynamics*, vol. 30, no. 4, 2007.

[3] I. Beer, S. Ben-David, C. Eisner, and A. Landver, "RuleBase: An industry-oriented formal verification tool," in *DAC*, 1996.

[4] S. Ben-David, B. Sterin, J. M. Atlee, and S. Beidu, "Symbolic model checking of product-line requirements using SAT-based methods," in *ICSE*, vol. 1, 2015, pp. 189–199.

[5] M. Bozzano, A. Cimatti, A. Fernandes Pires, D. Jones, G. Kimberly, T. Petri, R. Robinson, and S. Tonetta, "Formal design and safety analysis of AIR6110 wheel brake system," in *CAV*, 2015.

[6] A. R. Bradley, "SAT-Based Model Checking without Unrolling," in *VMCAI*, 2011, pp. 70–87.

[7] A. R. Bradley, "Understanding IC3," in *SAT*, 2012, pp. 1–14.

[8] H. Chockler, A. Ivrii, A. Matsliah, S. Moran, and Z. Nevo, "Incremental Formal Verification of Hardware," in *FMCAD*, 2011, pp. 135–143.

[9] H. Chockler, O. Kupferman, and M. Y. Vardi, "Coverage metrics for temporal logic model checking," *FMSD*, vol. 28, no. 3, pp. 189–212, 2006.

[10] A. Cimatti, A. Griggio, S. Mover, and S. Tonetta, "Parameter synthesis with IC3," in *FMCAD*, 2013, pp. 165–168.

[11] A. Cimatti, A. Griggio, B. J. Schaafsma, and R. Sebastiani, "The MathSAT5 SMT solver," in *TACAS*, 2013, pp. 93–107.

[12] K. Claessen and N. Sörensson, "A liveness checking algorithm that counts," in *FMCAD*, 2012, pp. 52–59.

[13] A. Classen, M. Cordy, P. Heymans, A. Legay, and P.-Y. Schobbens, "Model checking software product lines with snip," *(STTT)*, pp. 1–24, 2012.

[14] A. Classen, M. Cordy, P.-Y. Schobbens, P. Heymans, A. Legay, and J.-F. Raskin, "Featured transition systems: Foundations for verifying variability-intensive systems and their application to ltl model checking," *IEEE Trans. Softw. Eng.*, vol. 39, no. 8, pp. 1069–1089, 2013.

[15] A. Classen, P. Heymans, P.-Y. Schobbens, and A. Legay, "Symbolic model checking of software product lines," in *ICSE*, 2011, pp. 321–330.

[16] A. Classen, P. Heymans, P.-Y. Schobbens, A. Legay, and J.-F. Raskin, "Model checking lots of systems: efficient verification of temporal properties in software product lines," in *ICSE*, 2010, pp. 335–344.

[17] N. Een, A. Mishchenko, and R. Brayton, "Efficient Implementation of Property Directed Reachability," in *FMCAD*, 2011, pp. 125–134.

[18] E. A. Emerson and V. Kahlon, "Reducing model checking of the many to the few," in *CADE*, 2000, pp. 236–254.

[19] M. Gario, A. Cimatti, C. Mattarei, S. Tonetta, and K. Y. Rozier, "Model checking at scale: Automated air traffic control design space exploration," in *CAV*, 2016.

[20] A. Griggio and M. Roveri, "Comparing Different Variants of the IC3 Algorithm for Hardware Model Checking," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 35, no. 6, pp. 1026–1039, Jun 2016.

[21] P. James, F. Moller, H. N. Nguyen, M. Roggenbach, S. Schneider, and H. Treharne, "On modelling and verifying railway interlockings: Tracking train lengths," *Science of Computer Programming*, vol. 96, no. 3, 2014.

[22] J. Marques-Silva, "Interpolant learning and reuse in sat-based model checking," *Theoretical Computer Science*, vol. 174, no. 3, pp. 31 – 43, 2007.

[23] C. Mattarei, A. Cimatti, M. Gario, S. Tonetta, and K. Y. Rozier, "Comparing different functional allocations in automated air traffic control design," in *FMCAD*, 2015.

[24] F. Moller, H. N. Nguyen, M. Roggenbach, S. Schneider, and H. Treharne, "Defining and model checking abstractions of complex railway models using CSP—B," in *HVC*, 2013.

[25] P. Schrammel, D. Kroening, M. Brain, R. Martins, T. Teige, and T. Bienmüller, "Incremental bounded model checking for embedded software," *Formal Aspects of Computing*, 2016.

[26] F. Somenzi and A. R. Bradley, "IC3: Where Monolithic and Incremental Meet," in *FMCAD*, 2011, pp. 3–8.

[27] B. Yang, R. E. Bryant, D. R. O'Hallaron, A. Biere, O. Coudert, G. Janssen, R. K. Ranjan, and F. Somenzi, "A performance study of bdd-based model checking," in *FMCAD*, 1998, pp. 255–289.

[28] G. Yang, M. B. Dwyer, and G. Rothermel, "Regression model checking," in *ICSM*, 2009, pp. 115–124.