Preprint from Proceedings of
Formal Methods in Computer Aided Design, November 2006, San Jose, USA

Advanced Unbounded Model Checking Based on AIGs, BDD Sweeping, And Quantifier Scheduling

Florian Pigorsch, Christoph Scholl, and Stefan Disch
Albert-Ludwigs-Universität Freiburg, Institut für Informatik,
D-79110 Freiburg im Breisgau, Germany
Email: {pigorsch, scholl, disch}@informatik.uni-freiburg.de

***Abstract*— In this paper we present a complete method for verifying properties expressed in the temporal logic CTL. In contrast to the majority of verification methods presented in recent years, we support *unbounded* model checking based on symbolic representations of characteristic functions. Among others, our method is based on an advanced And-Inverter Graph (AIG) implementation, quantifier scheduling, and BDD sweeping. For several examples, our method outperforms BDD based symbolic model checking by orders of magnitude. However, our approach is also able to produce competitive results for cases where BDD are known to perform well.**

I. Introduction

Given a sequential circuit and properties in some temporal logic like CTL or LTL, model checking is a method for verifying these properties [1], [2]. In the early nineties, by introducing *symbolic* model checking, Burch et al. substantially extended the class of systems which can be verified [3], [4]. In symbolic model checking binary decision diagrams (BDDs) [5] are used both for state set representation and for state traversal. Sets of states are represented by characteristic functions which in turn are represented by BDDs.

However, in the last few years SAT based techniques like Bounded Model Checking (BMC) [6], [7] have been attracting much interest, since industrial needs ask for methods avoiding the well known memory explosion problem which may occur during symbolic model checking of large circuits. BMC applied to certain properties (invariants or, more generally, LTL formulas) 'unfolds' the transition relation for k steps in order to find counterexamples. If no counterexample of length k is found, then k is increased and BMC is used again. For *proving* properties using BMC a suitable upper bound on k is needed. In the case of invariants, e.g., the search for counterexamples can be stopped, when k equals the *diameter* of the system, i.e., the maximum length of all shortest paths between states in the system. Then, BMC ends up with a proof of the property. Unfortunately, computing diameters of large systems turns out to be hard. The problem may be reduced to the validity check of a quantified Boolean formula (QBF) with alternating existential and universal quantifiers [6]. Since this check is usually hard for large systems, BMC is mostly used as an incomplete method for finding errors in practice.[1]

In this paper, we present a *complete* method for verifying properties expressed in the temporal logic CTL. Our method is based on a symbolic representation of sets of states. However, our symbolic representation relies on And-Inverter Graphs (AIGs) [9], [10] instead of BDDs. So far, And-Inverter Graphs have been successfully applied in combinational equivalence checking [9], [10] and in BMC for simplifying representations of transition relations [11]. Basically, they are Boolean circuits which consist of AND gates and inverters only. In contrast to BDDs, AIGs do not provide canonical representations of Boolean functions. Since we do not need canonical representations for representing sets of states, we are able to avoid memory blow-ups during the construction of (canonical) BDDs. On the other hand, checks for satisfiability or validity, which are needed during the model checking process, do not come for free as for BDDs, because there are different AIG representations for constants 0 and 1.

In order to obtain as much sharing of subcircuits as possible we make use of a special version of AIGs, the so-called functionally reduced AIGs (FRAIGs) which were introduced by Mishchenko et al. [12] in the context of logic synthesis, technology mapping and combinational equivalence checking. Like general AIGs, FRAIGs still form non-canonical representations of Boolean functions, but they have the additional property that they do not contain any pair of functionally equivalent nodes. This invariant is maintained during construction of FRAIGs by using a SAT solver. In addition, the construction of FRAIGs is assisted by functional simulation in order to avoid unnecessary SAT checks for pairs of nodes for which already simulation is able to prove non-equivalence. Similar ideas for compressing AIG representations using 'SAT sweeping' and functional simulation can be also found in [11].

The most difficult step during model checking using FRAIGs is the elimination of existential quantifiers. As in [13], [14], [15] existential quantifiers $\exists x f$ are eliminated by replacing them by $f|_{x=0} + f|_{x=1}$. Of course, in the worst case the elimination of one quantifier may double the size of the representation. Although it is not very likely that this worst case behavior can be avoided in random examples (since SAT checking is NP hard), we show in our experimental results that we succeed in limiting the increase in size by several measures including a clever choice of the order of quantifications ('quantifier scheduling'). Interestingly, in contrast to a widespread belief [16], [17], [18] our results prove that – for our approach – quantifier elimination by a circuit-based computation of $f|_{x=0} + f|_{x=1}$ is not restricted to models with a small number of inputs (which have to be quantified during symbolic model checking). Our novel method for quantifier scheduling is based on estimations on the AIG sizes of the results after performing quantifier elimination. In Section V we motivate the importance of quantifier scheduling by

[1] Another possibility consists in increasing k up to the length of the longest simple path between two states [8]. Whereas it is easier to determine the length of the longest simple path than to determine the diameter of the system, the longest simple path may be exponentially longer than the diameter. If this is the case, unfolding the transition relation for such a large number of steps will be prohibitive.

giving an example and we describe the approach in more detail. Note that our way of eliminating quantifiers ($\exists x f = f|_{x=0} + f|_{x=1}$) also motivated the use of *functionally reduced* AIGs (FRAIGs) instead of 'standard' AIGs: Since a trivial implementation of quantifying several input variables would lead to an exponential growth of the representation, we need the more aggressive form of enforcing sharing of subcircuits which is provided by FRAIGs.

Other techniques for limiting the sizes of our representations of state sets are node selection heuristics and BDD sweeping:

- Whenever a new node is inserted into our FRAIG representation, we check whether there is already a node in the representation which is functionally equivalent to this new node (using SAT combined with simulation). If there is already a functionally equivalent node, we keep only one representation for the function and replace the representation of one node by the other (this is in contrast to [12] where various representations of the same function are kept for technology mapping purposes). In order to keep the overall size of the representation small we have to select carefully which representation is kept (see Section IV).
- BDD sweeping is known from combinational equivalence checking [9], [10] and builds BDDs for AIG nodes starting at the primary inputs until a certain node limit is reached. BDD sweeping is used there from time to time in order to identify equivalent nodes in the AIG. Since we are using SAT for maintaining the FRAIG invariant we do not need BDD sweeping with this objective. In contrast to the traditional use of BDD sweeping we make use of BDD sweeping in the cone of selected output functions of our FRAIG representation in order to compute smaller AIG representations. After one step of BDD sweeping we check whether our FRAIG representations decrease in size when parts of the FRAIG representation are replaced by subgraphs which are structurally equivalent to the BDDs computed during BDD sweeping.

Interpolation based model checking [17] is related to our approach in the sense that it also provides a method for unbounded model checking. In contrast to our approach [17] does not handle CTL properties, but invariants, and it does not use exact image computations, but overapproximations by so-called Craig interpolants. Due to the overapproximated image computation the method of [17] needs to be applied iteratively on unfoldings of the transition relation for an increasing number of steps (as in Bounded Model Checking). Our method does not need several unfoldings, but it can be used in standard symbolic model checking just replacing BDD representations for state sets by AIG based representations. Other related approaches perform quantifier elimination by using a SAT solver for enumerating all satisfying assignments of a given function [16], [19]. During the enumeration process disjunctions of cubes (or conjunctions of clauses) are collected leading to a two-level representation of the result of the quantification. Characteristic functions for sets of states and transition relations are expressed in conjunctive normal form (CNF) limiting the method to functions having efficient two-level representations. The idea of SAT-based quantifier elimination was refined in [18]. Whereas this method is still based on enumerations of satisfying assignments of a function f, disjunctions of cubes are replaced by disjunctions of cofactors of the function f.

The following novel contributions are introduced by our approach:

- We developed methods for quantifier scheduling which are especially tailored towards our state set representations using FRAIGs. We can show that a proper scheduling of quantifications can lead from exponential representations to representations of linear size.
- The size of the FRAIG representations is limited by heuristics for node selection when functionally equivalent nodes are identified.
- We are using BDD sweeping as a method for non-local logic optimization of our FRAIG representations. BDD sweeping is controlled by heuristics based on the size of the AIG representations and on the success of previous runs of BDD sweeping.

We applied our representations of state sets and of transition functions to CTL model checking. We are using a standard CTL model checking algorithm based on symbolic representations of state sets. However, we make use of degrees of freedom in CTL model checking by preferring operations which are beneficial for our representation (see also Section II).

Our experimental results prove the efficiency of our approach. For several examples, our method outperforms BDD based symbolic model checking by orders of magnitude. However, note that our approach is also able to produce competitive results for cases where BDDs are known to perform well (which was not observed for approaches [13], [14], e.g.). We show in detail how our concepts such as quantifier scheduling, node selection heuristics and BDD sweeping as a non-local optimization step contribute to the success of our experiments.

The paper is structured as follows: We begin with a brief review of CTL model checking in Section II. Then we describe both And-Inverter Graphs (AIGs) in general and the special version of AIGs we use as a data structure for model checking (Section III). In Section IV we describe our heuristics for node selection and in Section V we present our method for quantifier scheduling. AIG compression techniques by BDD sweeping are given in Section VI. After presenting experimental results in Section VII we give some conclusions and future directions in Section VIII.

II. Preliminaries

We use our FRAIG representation in the context of symbolic model checking [3], [4].

Symbolic model checking is applied to Kripke structures (which may be derived from sequential circuits) on the one hand and to a formula of a temporal logic (in our case CTL (Computation Tree Logic)) on the other hand.

An essential step in the recursive evaluation of CTL formulas is the preimage computation which computes for a set of states $Sat(\phi)$ the set of states $Sat(EX\phi)$ with at least one successor in $Sat(\phi)$:

$$\chi_{Sat(EX\phi)}(\vec{q},\vec{x}) := \exists \vec{q}' \exists \vec{x}' \Big(\chi_R(\vec{q},\vec{x},\vec{q}') \cdot \big(\chi_{Sat(\phi)}|_{\substack{\vec{q}\leftarrow\vec{q}' \\ \vec{x}\leftarrow\vec{x}'}}\big)(\vec{q}',\vec{x}') \Big) \tag{1}$$

(As usual χ_M means the characteristic function of set M, $\vec{x}$ represents the current input variables, $\vec{q}$ the current state variables, $\vec{q}\,'$ the next state variables, and $\vec{x}\,'$ the next input variables. χ_R represents the transition relation of the Kripke structure.)

It is well known that the same formula can also be computed based on transition *functions* δ_i of the sequential circuit instead

of the transition *relation* R:

$$\chi_{Sat(EX\phi)}(\vec{q},\vec{x}) := \exists \vec{x}'\Big(\chi_{Sat(\phi)}|\ {\scriptstyle \begin{array}{c} q_1 \leftarrow \delta_1(\vec{q},\vec{x}) \\ \dots \\ q_m \leftarrow \delta_m(\vec{q},\vec{x}) \\ \vec{x} \leftarrow \vec{x}' \end{array}}\ \Big)(\vec{q},\vec{x},\vec{x}')\Big) \quad (2)$$

In our implementation of the model checking procedure we always prefer Equation (2) over Equation (1), since the substitution operation is easy in the AIG context and can be performed in parallel for several substitutions. Although we use sophisticated methods to prevent memory blow-ups due to quantification, in principle quantification needs special attention, since quantifying a single variable has the risk of doubling the size of the representation. If not needed, we do not take this risk and we avoid the additional effort of preventing the representation from increasing.

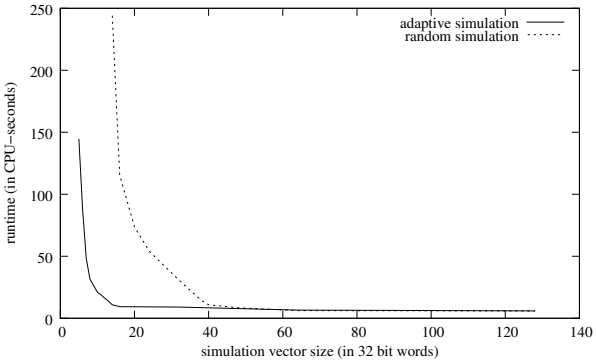

Fig. 1. Impact of adaptive simulation (picojava/icu benchmark)

III. And-Inverter Graphs

Recently, And-Inverter Graphs (AIGs) [9], [10] enjoy a widespread application in combinational equivalence checking and Bounded Model Checking (BMC). They are simply a special kind of directed acyclic graphs representing boolean functions. There are three types of nodes: *and nodes* with two outgoing edges, modeling the Boolean conjunction of the functions represented by the two edges, *variable nodes* with no outgoing edges but labelled with a variable name, representing boolean variables, and a special terminal node with no outgoing edges, forming the constant 0 function.

The edges of an AIG may contain negation marks that denote complementation.

Constructing AIGs using one level structural hashing [10] assures that we do not have two different nodes with the same pair of successors.

A. Functionally Reduced And-Inverter Graphs

AIG representations of Boolean functions are not canonical – for each Boolean function there exist many structurally different AIGs. Actually an AIG may contain functionally redundant nodes, i.e., nodes which are roots of structurally different subgraphs representing the same functions.

Redundant nodes lead to two problems: On the one hand, the graph structure is inefficient. Redundant nodes could be merged to reduce the graph size. On the other, checking the equivalence of two nodes needs additional effort.

To address these problems Mishchenko et al. [12] introduced the notion of functionally reduced AIGs (FRAIGs). The main idea is to check for equivalent nodes using SAT-based equivalence checking techniques while constructing an AIG and to merge them immediately. (In a similar approach Kuehlmann [11] uses 'SAT sweeping' from time to time in order to remove functionally equivalent nodes in AIGs which were not reduced immediately during construction.) This approach establishes the *functional reduction property*: There will not be any two nodes in an FRAIG representing the same Boolean function (and there will not be a pair of nodes where one represents the complement of the Boolean function represented by the other).

B. An AIG Package for Model Checking

Since we use our AIG package for state set representations in CTL model checking, we have different requirements compared to usual packages for combinational equivalence checking. In this section we have a brief look at the key features of our package.

Apart from standard Boolean operations which are translated into AND operations and / or complementations we have to support substitution and existential quantification. Substitution of variables by functions is basically reduced to replacements of inputs of an AIG by subgraphs representing these functions and it can be easily performed for several variables in parallel. Existential quantification $\exists x f$ is reduced to $f|_{x=0}+f|_{x=1}$ (with optimizations described in the following sections).

Whereas in combinational equivalence checking only insertion of nodes has to be supported, we need efficient methods for the deletion of nodes. Nodes have to be deleted when certain state set representations are not needed any longer during the model checking procedure and when functionally equivalent nodes are merged into one representation. The hash table we use for one-level structural hashing (applied as a fast technique for detecting isomorphic AIG nodes) permits lookups in constant time. We have enriched the data structure by adding linked lists to each AIG node chaining all hash table entries affected by this node. This allows for the fast deletion of all occurrences of a node from the hash table without inspecting all table entries.

To maintain the functional reduction property, we use a simulation guided, SAT based equivalence checking method known from the BMC and combinational equivalence checking domains as proposed in [11], [12]. The idea is to avoid powerful methods for easy problems: If for a given pair of nodes simulation is already able to prove non-equivalence, more time consuming SAT checks are not needed. The simulation vectors are initially random, but they are updated using feedback from satisfied SAT instances. Always maintaining a fixed number of simulation vectors we use a simple FIFO replacement method when new vectors are inserted. Figure 1 shows the impact of different simulation vector sizes and the use of learned simulation vectors in a typical model checking run. Depending on the size of the simulation vectors used, the dashed line shows the run times for the complete model checking run, when the learning of distinguishing simulation vectors from satisfied SAT instances is turned off. The solid line shows the same run times for the case that learning is turned on. At least for the smaller sizes of simulation vectors learning leads to considerable improvements of run times. Obviously, by learning we obtain 'good' simulation vectors which are able to prevent time consuming SAT checks during future node insertions.

Additional features of our AIG package which are used during model checking are node selection, quantifier scheduling, and BDD sweeping. They will be described in the following three sections.

IV. Node Selection in Case of Equivalence

When constructing a new node in an AIG, one will often encounter the situation that the current AIG already contains a functionally equivalent node which is the root of a structurally different subgraph. Since the functional reduction invariant must be maintained, only one of the two nodes can be kept and the other one needs to be removed from the AIG. Unlike the approach in [12] where the already existing AIG node is kept and all equivalent nodes are stored in a list of possible structural representations for later technology mapping, our AIG package tries to keep the memory consumption as low as possible and thus destroys redundant nodes. This strategy is vital for a successful employment of AIGs in the model checking domain.

The question is whether to preserve the old, existing node or the newly created one. We use two different heuristics to conquer the problem:

- h_{keep}. We always keep the old node and discard the new node. The drawback of this trivial method is the possible rejection of more efficient structural representations.
- h_{size}. We keep the node that structurally depends on less variables. If the nodes have equal support sizes, we consider the subgraphs (cones) rooted by the two nodes and select the node which has a smaller cone. [2]

In our experiments (see Section VII) we will show that the naive node selection heuristics h_{keep} may result in high and even unmanageable node counts, while the more advanced one is able to reduce the AIGs to reasonable sizes.[3]

V. Quantifier Scheduling

During model checking we eliminate existential quantifiers $\exists x f$ by replacing them by $f|_{x=0}+f|_{x=1}$. In the worst case this elimination may double the size of the representation. Thus, after an existential quantification of a series of variables the size of the representation may potentially show an exponential blow-up. In this section we will present a heuristic method which aims at limiting this (potential) increase in size by a clever choice of the order of quantifications ('quantifier scheduling').

A. A Motivating Example

First of all, we give a motivating example which shows that the order of quantifications may be essential for avoiding memory blow-ups. Consider a simple carry ripple adder which computes the sum $(s_n, \ldots, s_0)$ for two operands $(a_{n-1}, \ldots, a_0)$ and $(b_{n-1}, \ldots, b_0)$. Now we want to compute the set of inputs $(b_{n-1}, \ldots, b_0)$ with the property that there is an input $(a_{n-1}, \ldots, a_0)$ with $2^{n-1} \leq \sum_{i=0}^{n-1} a_i 2^i + \sum_{i=0}^{n-1} b_i 2^i < 2^n$, i.e. with $s_n = 0$ and $s_{n-1} = 1$. The problem may be solved by computing a symbolic (BDD or AIG based) representation of $\overline{s_n} \cdot s_{n-1}$ based on the carry ripple circuit and by computing the existential quantification $\exists a_{n-1} \ldots \exists a_0 \overline{s_n} \cdot s_{n-1}$. The result of the quantification is a representation of a characteristic function for the set of inputs $(b_{n-1}, \ldots, b_0)$ fulfilling the given property. Since it is easy to see that this set of inputs is equal to $\mathbb{B}^n$, the final result has to be equal to 1. Now we consider the two extreme cases for the order of quantification: The first order UP is $(a_0, \ldots, a_{n-1})$ (i.e. we start with the quantification of the least significant bit) and the second order $DOWN$ is $(a_{n-1}, \ldots a_0)$ starting with the quantification of the most significant bit. Remember that we reduce quantification wrt. one variable to the disjunction of positive and negative cofactors. Figure 2.(1) shows the result of the quantification wrt. the first variable a_0 of order UP (for simplicity applied to the given carry ripple circuit, not to the corresponding AIG, which has roughly the same structure). The illustration shows that the propagation of constants 0 and 1 for the negative and the positive cofactor wrt. a_0 already stops at bit position 1 and no subcircuit sharing can be observed in the remaining circuit. Thus, the size of the circuit is almost doubled by quantification.

However, if we quantify variable a_{n-1} first, we obtain the situation shown in Figure 2.(2). In this case most parts of the circuit are shared between the positive and the negative cofactor. Since we use FRAIGs which identify functionally equivalent nodes, the corresponding FRAIG also shows this sharing. The duplication of the number of nodes as observed in the previous case can not be seen here. The effect shown above continues during the following quantifications according to the orders UP and $DOWN$: As shown in Table I for the example of a 14-bit-adder, we observe an exponential blow-up of AIG nodes during quantification according to order UP, whereas the number of AIG nodes is monotonically decreasing for quantification order $DOWN$. (Note that BDD sweeping described in the next section was not used in this experiment.) Altogether our example shows that there may be an exponential gap wrt. AIG sizes between good and bad orders of quantification. So we have a strong need for heuristics computing good quantification orders.

B. A Heuristic Approach to Quantifier Scheduling

Here we present a method for quantifier scheduling which is especially tailored towards our state set representations using FRAIGs. Our greedy method is based on estimations on the sizes of the results after performing quantifier elimination. Before performing a quantification $\exists x f = f|_{x=0} + f|_{x=1}$ for some variable x and some function f represented by a FRAIG, we compute an estimate on the FRAIG size of the final result:

- In a first step we consider the subgraph of the AIG representing f and for each $\epsilon \in \{0, 1\}$ we determine by two traversals of this subgraph the set R_ϵ of nodes which are not removed by propagation of constant ϵ.
- In a second step we compute an estimate for the node sharing between the representations of the positive and the negative cofactor: If a node k which occurs both in R_0 and R_1 is not connected to variable x by a path in the AIG graph, then it does not depend on x and thus the nodes corresponding to k in the representation of $f|_{x=0}$

[2] It is easy to see that h_{size} never replaces a node k by another node k' having k in its cone (which would create a loop in the AIG).

[3] In the case the used heuristics suggest to keep the old node, the only thing to do is to delete the new node. But if the new node is selected, we use a technique similar to implementation techniques known from BDD packages: All edges of the AIG pointing to the old node must be modified to reference the new node. Since the data structure used in our AIG package does not provide an efficient method for finding all predecessors of a node, we need to use a more subtle replacement method: we actually transfer the data of the new node object into the old node object and then delete the new node. By doing this no edge has to be touched.

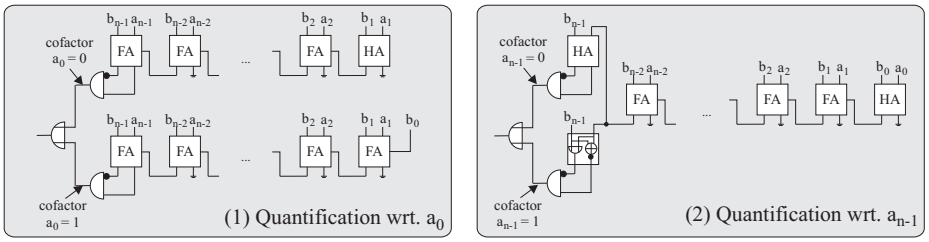

Fig. 2. $\overline{s_n} \cdot s_{n-1}$ after (1) quantification of a_0 and (2) quantification of a_{n-1}.

TABLE I

14-BIT-ADDER, FUNCTION $\overline{s_{14}} \cdot s_{13}$: NUMBER OF AIG NODES AFTER QUANTIFICATIONS OF SINGLE VARIABLES a_i, ORDERS UP AND $DOWN$

Quant. Nr.	orig.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
UP	111	105	106	115	140	197	318	567	1072	2089	4130	8219	16404	32781	1
DOWN	111	105	99	93	87	81	75	69	63	57	51	45	39	33	1

and $f|_{x=1}$ are the same (due to the functional reduction property). So the set $R_{0,1}$ of all such nodes is our estimate for the set of nodes shared between $f|_{x=0}$ and $f|_{x=1}$.[4]

- Finally, our estimate for the size of $f|_{x=0} + f|_{x=1}$ is $1 + |R_0| + |R_1| - |R_{0,1}|$.

For reasons of efficiency our estimate does not consider node sharing between nodes for both cofactors on the one hand and nodes which are not in the subgraph representing f on the other hand. Additionally, possible restructuring of the AIG during node insertion (see Sections IV and VI) is not taken into account. However, as shown by experimental results, our heuristic estimate seems to be reasonable for computing a good order for quantification: Whenever a number of variables $x_1, \ldots, x_n$ has to be quantified for a function f, we compute for each x_i our estimate of the size of $\exists x_i f$ and (greedily) start with quantification of the variable with smallest cost. Then the method is repeated to determine the next variable to be quantified and so on.

We would like to point out that in the case of our motivating example from above our heuristic method leads to quantification order $DOWN$ which produces a monotonically decreasing number of AIG nodes, whereas unfavorable orders like the order UP shown above lead to an exponential peak size in the number of AIG nodes before the final result 1 is computed.

VI. BDD SWEEPING

BDD sweeping [20], [9], [10] is a well-known technique from the domain of combinational equivalence checking (CEC). It builds BDDs for AIG nodes starting at the primary inputs until a certain node limit is reached. Whereas in [20], [9], [10] BDD sweeping is used in order to identify functionally equivalent nodes in the AIG, this is not needed in our case, since we always maintain the functional reduction property using SAT as described in Section III. Here we use BDD sweeping as a means for non-local optimizations of our AIG representations. However, for reasons of efficiency both the number of BDD sweepings and the cost of a single BDD sweeping have to be limited.

From time to time, after certain operations of the AIG package, BDD sweeping is applied to the cone of the corresponding result. BDD sweeping builds a BDD for the cone of the given AIG node starting from the variable nodes and using AND and NOT operations. Variable reordering applied by the BDD package automatically tries to find an optimal variable order in terms of BDD node count. If BDD sweeping is able to compute the BDD for the given AIG node, then we check whether it makes sense to replace the cone of the AIG node by an AIG which is structurally equivalent to the BDD. Here we exploit the fact that any BDD node can be interpreted as a multiplexer, which can be transformed into an AIG with exactly three AIG nodes. Thus, if the size of the generated BDD is smaller than one third of the size of the given cone, we create an AIG from the BDD structure by recursively transforming the BDD nodes to their three-node AIG representation. When inserting this new AIG into the AIG package, node selection heuristics as described in Section IV are used as usual with the additional effect that subgraphs of the new AIG may be replaced by smaller (functional equivalent) representations which are already present in the existing AIG graph.

In order to limit the *cost* of a single BDD sweeping we use a variable *BDD_limit*. Whenever the number of nodes in the BDD package is larger than *BDD_limit*, the BDD construction is aborted.

In order to limit the *number* of BDD sweepings, we decided to confine BDD sweeping to the results of cofactor operations which occur during existential quantification, since existential quantification of a variable is the only operation that has the risk of doubling the size of the representation. Moreover, BDD sweeping is not applied after *all* cofactor operations, but only after a small fraction of cofactor operations controlled by sophisticated heuristics based on the sizes of the results and the success of previous BDD sweepings. To avoid unnecessary BDD sweepings, BDD sweeping is only applied, if the AIG size of the current operation is larger than a some variable *AIG_limit*. *AIG_limit* is initialized to a certain constant (100 in our current implementation) and it evolves as follows:

- If a BDD is successfully built within the node limit *BDD_limit*, but it is not used in the AIG due to its size, *AIG_limit* is multiplied by a certain factor $f_1 > 1$ ($f_1 = 1.2$ in our current implementation).
- If the BDD construction is aborted, since *BDD_limit* is exceeded, *AIG_limit* is multiplied by some larger factor $f_2 > 1$ ($f_2 = 4$ in our current implementation).
- If a BDD is successfully built and a structural equivalent AIG is inserted into the AIG package, then *AIG_limit* is

[4] Situations where a node does not functionally depend on a variable x, but is (structurally) connected to x, may be possible in AIGs due to non-canonicity. However they are neglected for reasons of efficiency.

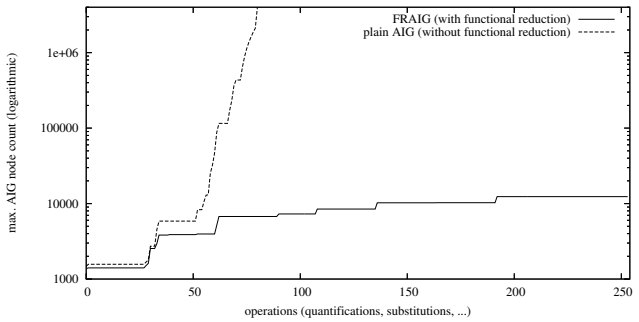

Fig. 3. Impact of functional reduction (picojava/icu).

set to the size of the resulting AIG.

- Whenever BDD sweeping is not applied, since *AIG_limit* is too large, *AIG_limit* is decreased by multiplication with $1 - f_3 \cdot (\frac{1}{2})^{abort}$, where f_3 is a small constant ($0 < f_3 < 1$) and $abort$ is the number of times the BDD construction is aborted due to an exceeded node limit (in our implementation we used $f_3 = 0.01$).

The presented heuristics ensure that unsuccessful BDD sweeping runs result in fewer BDD sweeping runs in the future.

The variable *BDD_limit* for aborting BDD constructions has to be high enough to allow for BDD variable reordering and is set to $\max(100 \cdot \textit{AIG\_limit}, 10^6)$ in our implementation.

Whenever BDD sweeping is aborted, it ends up with a number of BDDs for AIG nodes in the considered cone, since it computes BDDs beginning with the input variables of the cone. The procedure described above can be easily extended by making use of the BDDs computed so far. However, this feature is not yet realized in our prototype implementation.

VII. Experimental Results

We performed a number of experiments for evaluating our approach which we called '*AIG-MC*'. The examples AM2910, Coherence, DAIO, Picojava/ICU, Viper, and Barrelshifter are taken from the *VIS Verification Benchmark* set [21].[5] For each benchmark we checked all CTL formulas provided in the *VIS Verification Benchmark* set. We randomly selected the benchmarks from the set of those benchmarks in the *VIS Verification Benchmark* set which have CTL specifications. Since the prototype implementation of our model checker is currently not yet able to read the benchmark format used in VIS, we had to translate the hierarchical and multi-valued models into flat, binary encoded models. This included a manual adaption of the CTL formulas to the new binary encoding variables.[6] Moreover, we used a pipelined ALU ('PALU') similar to the one presented in [3]. The pipelined ALU includes 16 registers, a combinational adder, a combinational multiplier, and bitwise operations.[7] As in [3] the inputs to the ALU are instruction codes containing a specification of the operation, the source registers and the destination register. For the pipelined ALU we checked the CTL formula $\phi = AG\big(\text{''}\mathtt{R}_2 := \mathtt{R}_0 \oplus \mathtt{R}_1\text{''} \rightarrow \big((AX)^2\mathbf{R}_0 + (AX)^2\mathbf{R}_1 \equiv (AX)^3\mathbf{R}_2\big)\big)$

[5]The number of registers in the barrelshifter was increased from 4 to 10.

[6]The complete set of benchmarks is provided in [22].

[7]The bit width of all operations and registers for palu12,4 is 12, for palu14,4 14 and for palu16,4 16, respectively.

TABLE II
Node selection heuristics in *AIG-MC*

benchmark (ctls)	h_{keep}		h_{size}		
	nodes	time	nodes	repl.	time
am2910 (1-3,5-6)	5354	5.9	2129	232	0.9
barrel10,4 (1)	5918	39.6	5918	0	37.2
coherence (1,7)	1310	0.3	1303	20	0.3
daio (1-4)	50862	10680.9	17069	13078	245.0
decay11	15578	33.6	15578	1	33.7
palu12,4 (xor)	3802	2.1	3802	216	1.9
palu14,4 (xor)	4654	2.7	4654	252	2.6
palu16,4 (xor)	5586	3.6	5586	288	3.3
picojava/icu (1)	23924	21.6	10650	937	8.5
viper (1-3)	15975	7.7	15970	69	11.7

(similar to formula (1) from [3]).[8] The benchmarks named 'decay$n$' contain registers of bit widths n and they compute decaying sums of sequences of inputs according to the formula $register_{new} = \lceil \frac{register_{old}}{2} \rceil + input$.[9]

Note that the barrelshifter example used here is different from the barrelshifter example given in [13], [14]. The examples in [13], [14] do not contain inputs, and thus, quantification is not needed during the fixed point computation of the model checking procedure (see Section II, equation (2)). We did not compare our results to the results from [13], [14], since our goal was to prove that we are able to handle quantification as well. In this sense our experiments show that the objection raised by McMillan ([17], Section 1.1) 'because of the expense of quantifier elimination, this approach is limited to models with a small number of inputs (typically zero or one)' does not apply to our approach.

All experiments were performed on a 2 GHz Dual-Opteron workstation running Debian Linux. We used a timeout of 12 CPU hours.

First of all we demonstrate the effect of functional reduction by means of a typical example in Fig. 3. Fig. 3 shows the number of AIG nodes which are needed during model checking of the picojava/icu benchmark. In this experiment we turned off quantifier scheduling and BDD sweeping in order to concentrate on the effect of functional reduction. The numbers of AIG nodes were recorded after each quantification of a variable and after each substitution, thus the x-axis represents the ongoing progress of the model checking procedure. The numbers of nodes are presented with a logarithmic scale. The dashed line shows the number of nodes which are needed when functional reduction using SAT is turned off, the solid line shows the number of nodes of our FRAIG package using SAT based functional reduction. The example clearly shows that functional reduction is essential for the success in this kind of applications. An AIG package only using structural hashing is not able to provide sufficient compaction. For this reason we always consider results using our FRAIG package with SAT based functional reduction in the following.

In the first experiment we evaluated the effect of our node selection heuristics from Section IV. Table II lists the peak node counts and run times in CPU seconds for the two different proposed node selection heuristics: the naive method h_{keep} (always keeping the already existing node) and h_{size}. For h_{size} Table II also reports the numbers of node replacement

[8]Given an *exor* operation in the instruction register the formula basically checks whether the contents of the destination register in three steps are the same as the *or* operation of the contents of the operand registers in two steps. This would be true for an *or* operation in the instruction register, but is obviously not true for the *exor* operation.

[9]The property asks whether there is a sequence of inputs such that for the binary number R in the register $2^{n-1} \leq R < 2^n$.

TABLE III

AIG-MC W/O, WITH QUANT. SCHEDULING, BDD-MC, VIS

benchmark	ctl	w/o quant. sched. nodes	w/o quant. sched. time	quant. sched. nodes	quant. sched. time	BDD-MC time	VIS time
am2910	1	1132	0.2	1132	0.2	1.5	3.2
	2	1143	0.3	1143	0.3	1.5	3.2
	3	1605	0.8	1650	0.7	1.5	3.2
	4	16039	91.9	13818	51.2	9.0	5.5
	5	1977	1.1	1880	1.6	1.5	3.2
	6	1205	0.3	1181	0.3	1.5	0.8
barrel10,4	1	5918	50.3	5918	51.2	>12h	>12h
coherence	1	1303	0.3	1303	0.3	0.2	0.4
	2	43285	334.3	49010	172.4	1.0	0.4
	3	11744	22.3	10001	13.6	1.4	0.4
	4	25190	72.6	18781	41.7	1.6	0.4
	5	18590	40.5	7895	7.2	2.1	0.5
	6	23814	176.4	25298	88.3	0.8	0.4
	7	1303	0.3	1303	0.3	0.2	0.4
	8	101185	609.8	42042	151.6	37.8	0.4
	9	18590	33.1	7895	5.0	1.6	0.4
daio	1	996	0.6	996	0.6	0.1	0.3
	2	996	0.8	996	0.8	0.2	0.4
	3	1768	1.4	1768	1.4	0.3	0.4
	4	996	0.8	996	0.8	0.2	0.4
decay32	1	1901	6.3	718	1.2	0.5	0.0
decay48	1	2814	30.3	1070	4.0	2.8	0.2
decay64	1	3736	83.0	1422	9.8	21.1	0.3
palu12,4	xor	3802	153.4	3832	76.2	>12h	>12h
palu14,4	xor	4654	59.0	4684	119.4	>12h	>12h
palu16,4	xor	5586	1062.6	5616	91.1	>12h	>12h
picojava/icu	1	8144	6.6	2869	5.0	1.0	2.0
viper	1	15757	3.1	15757	3.0	43.4	75.0
	2	15757	6.7	15757	6.1	43.3	73.0
	3	15757	3.1	15757	3.0	43.3	74.1

TABLE IV

DETAILED RESULTS FOR *AIG-MC*

benchmark	bdd sweeping applic.	bdd sweeping succ.	bdd sweeping limit	sat applic.	sat equiv
am2910	0.06%	59.80%	0%	5.00%	45.58%
barrel10,4	0.01%	0%	100%	1.90%	42.51%
coherence	0.02%	74.06%	0.37%	0.97%	69.47%
decay32	0.39%	11.11%	0%	37.00%	4.65%
decay48	0.26%	9.09%	0%	55.09%	2.40%
decay64	0.20%	7.69%	0%	72.36%	1.50%
daio	0.12%	98.13%	0%	11.11%	93.99%
palu12,4	0.03%	12.5%	62.5%	1.48%	74.07%
palu14,4	0.02%	12.5%	62.5%	1.28%	76.92%
palu16,4	0.02%	12.5%	62.5%	1.12%	78.28%
picojava/icu	0.07%	33.33%	0%	5.06%	31.47%
viper	0.01%	33.33%	0%	4.56%	60.49%

steps. The numbers for each benchmark are averaged over all different CTL formulas. In this experiment we turned off BDD sweeping, because the naive method h_{keep} would never use the results of BDD sweeping. (Thus the results would be biased towards h_{size}, since it can exploit BDD sweeping whilst h_{keep} does not profit from it.) In the comparison we omitted the results for formula 4 of example 'AM2910' and formulas 2-6 and 8-9 of example 'Coherence', since the computation did not finish for h_{keep} within our limit on CPU time. (For formulas 2-6 and 8-9 of 'Coherence' the computation without BDD sweeping did not finish for h_{size} as well, i.e., BDD sweeping is essential for success with this benchmark (see also experiments of Table III).)

The results clearly show that the node selection heuristics are of great importance for obtaining good results: The heuristics h_{size} lead to a considerable decrease in peak node counts. The most impressive examples are AM2910 and DAIO where the peak number of AIG nodes for the naive method are by a factor of 2.5 and 3.0 higher than for h_{size}. Not only the node counts, but also the run times are greatly reduced by h_{size}, in the case of Picojava from 21.6 CPU seconds with h_{keep} to 8.5 CPU seconds with h_{size}, in the case of DAIO even from 3 CPU hours with h_{keep} to about 4 CPU minutes with h_{size}.

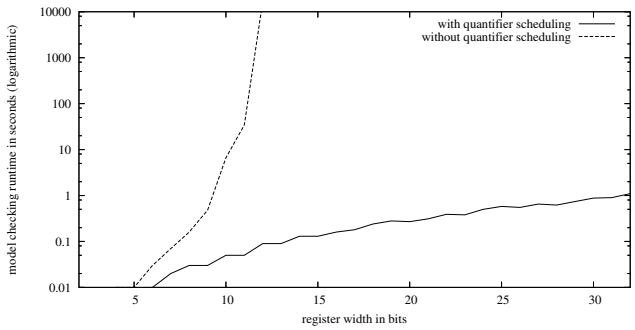

Fig. 4. Example decay$n$ with and without quantifier scheduling, run times.

Since the node selection heuristics h_{size} seems to be the best choice, we always use this method in the following.

In the following experiments we consider our method with BDD sweeping turned on. In the second experiment shown in Table III we evaluated the effect of quantifier scheduling and in the third experiment also shown in this table we compared our results to a BDD based version of our model checker and to the BDD based model checker VIS 2.1 [21]. We ran VIS using dynamic variable ordering with the sifting heuristics, don't care optimization via reachability analysis was turned off.[10]

For the different benchmarks and CTL formulas columns 3 and 4 show the peak node counts and run times for the unmodified quantification order ('w/o quant. sched.') and columns 5 and 6 give the same results for our quantifier scheduling heuristics from Section V. It can be observed that quantifier scheduling improves the peak node counts in most cases; for Coherence, formula 8, e.g., by a factor of 2.4, for decay64 by a factor of 2.6, for Picojava, by a factor of 2.8. The run times are always better or at least in the same range, for example Coherence, formula 8, run times are even reduced from 10.2 CPU minutes to 2.5 CPU minutes.[11]

In Fig. 4 we consider benchmark decay$n$ and have a closer look at the effect of quantifier scheduling: In this case we turned off BDD sweeping in order to perform a separate analysis of the effect of quantifier scheduling: Fig. 4 gives the CPU times for decay$n$ with increasing bit width n both for the version without quantifier scheduling (dashed line) and for the version with quantifier scheduling (solid line) (presented with a logarithmic scale). The experiment shows that without quantifier scheduling the run times grow exponentially with increasing bit width rendering completion of model checking for larger bit widths impossible. (The peak numbers of nodes are not presented here, but they show the same exponential growth.) With quantifier scheduling we observe only a moderate growth of node counts and run times.[12]

[10]For all experiments considered here the default option of performing a reachability analysis before backward model checking gave inferior results. For benchmark am2910 we even observed timeouts (larger than 12 CPU hours) for 5 out of 6 CTL formulas when reachability analysis was turned on.

[11]For examples palu$n$,4 the run times are somewhat misleading (increase of run times for palu14,4 by a factor of 2.0, decrease of run times even by a factor of 11.7 for palu16,4 for the version with quantifier scheduling): A more detailed analysis showed that the run times are almost exclusively due to unsuccessful runs of BDD sweeping (not leading to node replacements in the AIG). Without BDD sweeping the run times remain in the range of a few seconds.

[12]In Table III quantifier scheduling outperforms the version without quantifier scheduling for decay32, decay48, and decay64, but an exponential growth of node counts and run times is not observed. In this experiment BDD sweeping was turned on and it was able to prevent the exponential growth.

The last two columns of Table III give a comparison of our results to our own model checker BDD-MC with FRAIGs replaced by BDDs and to the BDD based model checker VIS, respectively. For barrel10,4, palu12,4, palu14,4, and palu16,4 neither BDD-MC nor VIS were able to provide a result within the CPU limit of 12 CPU hours. However, these examples did not form a problem for our model checker *AIG-MC* and we could solve them within a few seconds (see column 6 of Table III).

In contrast, for the remaining benchmarks taken from the VIS Benchmark set as well as for decay$n$, BDDs are known to perform well and these examples could be solved quickly by VIS. Even for this class of examples our approach finished in shorter time in 10 out of 26 cases and also for the remaining cases we could observe that our approach succeeded in producing competitive results within a few seconds.

For completeness we give some more details for our experiments with BDD sweeping and quantifier scheduling turned on in Table IV. Here again the numbers are averaged over all formulas. Column 2 shows the number of applications of BDD sweeping divided by the total number of attempts to insert a node into the AIG. Column 3 shows the numbers of successful applications of BDD sweeping (i.e. the numbers of BDD sweepings where the results are really used in the AIG package) divided by the total number of BDD sweepings. And finally, column 4 shows the numbers of aborted BDD sweepings (due to exceeded node limits) divided by the total number of BDD sweepings. BDD sweeping is only applied from time to time in all cases and in cases where BDD sweeping is not very successful (especially for examples 'Barrel' and 'PALU') our heuristics from Section VI work in the sense that unsuccessful BDD sweeping runs result in fewer BDD sweeping runs in the future. Column 5 shows the number of SAT checks divided by the total number of attempts to insert a node into the AIG, column 6 shows the fraction of SAT checks which lead to the result that the compared nodes are functionally equivalent. Although we always maintain the functional reduction property of our FRAIGs, the results show that the assistance of SAT by simulation and structural hashing as described in Section III-B assures that SAT is applied only for a small fraction of all node insertions. Moreover, the high percentage of SAT checks proving functional equivalence of two nodes shows the effectiveness of simulation in avoiding unnecessary SAT checks for nodes which are not equivalent.[13]

VIII. Conclusions and Future Work

We presented an approach to unbounded model checking based on And-Inverter Graphs as a representation of characteristic functions. Several methods such as functional reduction using simulation assisted SAT checks, node selection heuristics, quantifier scheduling, and BDD sweeping contribute to the success of our approach. For many examples, our method outperforms BDD based symbolic model checking by orders of magnitude, whereas it is still able to produce competitive results for cases where BDD are known to perform well. Although the experimental results for our current implementation already appear to be impressive, we believe that there remains room for improvement of the heuristics presented in Sections IV, V, and VI. Certainly, our prototype implementation will also profit from the integration of a number of interesting ideas recently developed for optimizing AIG representations such as DAG-aware circuit compression [23], [24] and advanced rewriting methods [15], [24]. In the future we will investigate whether methods for structural SAT solving [9] will be useful in our context and we will explore whether it sometimes makes sense to switch to lazy methods for AIG compression instead of our eager one.

[13] Benchmarks decay$n$ form an exception to this observation: In this case there are many nodes in the representation which are 'almost equivalent', so that simulation with a fixed number of simulation vectors is not very effective in distinguishing between them, leading to a number of SAT checks not proving functional equivalence.

References

[1] A. Sistla and E. Clarke, "The complexity of propositional linear temporal logics," *Journal of the ACM*, vol. 32, no. 3, pp. 733–749, 1985.

[2] E. Clarke, E. Emerson, and A. Sistla, "Automatic Verification of Finite–State Concurrent Systems Using Temporal Logic Specifications," *ACM Trans. on Programming Languages and Systems*, vol. 8, no. 2, pp. 244–263, 1986.

[3] J. Burch, E. Clarke, K. McMillan, D. Dill, and L. Hwang, "Symbolic Model Checking: 10^{20} States and Beyond," *Information and Computation*, vol. 98(2), pp. 142–170, 1992.

[4] K. McMillan, *Symbolic Model Checking*. Kluwer Academic Publisher, 1993.

[5] R. Bryant, "Graph - based algorithms for Boolean function manipulation," *IEEE Trans. on Comp.*, vol. 35, no. 8, pp. 677–691, 1986.

[6] A. Biere, A. Cimatti, E. Clarke, and Y. Zhu, "Symbolic model checking without BDDs," in *Tools and Algorithms for the Constuction and Analysis of Systems*, ser. LNCS, vol. 1579. Springer Verlag, 1999.

[7] A. Biere, A. Cimatti, E. Clarke, M. Fujita, and Y. Zhu, "Symbolic model checking using SAT procedures instead of BDDs," in *Design Automation Conf.*, 1999.

[8] M. Sheeran, S. Singh, and G. Stalmarck, "Checking Safety Properties Using Induction and a SAT-solver," in *FMCAD*, ser. LNCS, W. H. Jr. and S. Johnson, Eds., vol. 1954. Springer, 2000, pp. 407–420.

[9] V. Paruthi and A. Kuehlmann, "Equivalence checking combining a structural SAT-solver, BDDs, and simulation," in *Int'l Conf. on Comp. Design*, 2000, pp. 459–464.

[10] A. Kuehlmann, V. Paruthi, F. Krohm, and M. M.K. Ganai, "Robust Boolean Reasoning for Equivalence Checking and Functional Property Verification," *IEEE Trans. on CAD*, 2002.

[11] A. Kuehlmann, "Dynamic transition relation simplification for bounded property checking," in *Int'l Conf. on Computer-Aided Design*, 2004, pp. 50–57.

[12] A. Mishchenko, S. Chatterjee, R. Jiang, and R. Brayton, "FRAIGs: A unifying representation for logic synthesis and verification," EECS Dept., UC Berkeley, Tech. Rep., 03 2005.

[13] P. Abdullah, P. Bjesse, and N. Een, "Symbolic reachability analysis based on sat-solvers," in *Tools and Algorithms for the Constuction and Analysis of Systems*, ser. LNCS, vol. 1785. Springer-Verlag, 2000.

[14] P. Williams, A. Biere, E. Clarke, and A. Gupta, "Combining decision diagrams and SAT procedures for efficient symbolic model checking," in *Computer Aided Verification*, ser. LNCS, vol. 1855. Springer Verlag, 2000, pp. 124–138.

[15] G. Cabodi, M. Crivellari, S. Nocco, and S. Quer, "Circuit based quantification: Back to state set manipulation with unbounded model checking," in *Design, Automation and Test in Europe*, 2005.

[16] K. McMillan, "Applying SAT methods in unbounded symbolic model checking," in *Computer Aided Verification*, ser. LNCS, vol. 2404. Springer, 2002, pp. 250–264.

[17] ——, "Interpolation and SAT-Based Model Checking," in *Computer Aided Verification*, ser. LNCS. Springer, 2003.

[18] M. Ganai, A. Gupta, and P. Ashar, "Efficient SAT-based unbounded symbolic model checking using circuit cofactoring," in *Int'l Conf. on Computer-Aided Design*, 2004, pp. 510–517.

[19] H.-J. Kang and I.-C. Park, "Sat-based unbounded symbolic model checking," *IEEE Trans. on CAD*, vol. 24, no. 2, pp. 129–140, February 2005.

[20] A. Kuehlmann and F. Krohm, "Equivalence checking using cuts and heaps," in *Design Automation Conf.*, 1997, pp. 263–268.

[21] The VIS Group, "VIS Verification Benchmarks." [Online]. Available: http://vlsi.colorado.edu/∼vis/

[22] F. Pigorsch and C. Scholl, "Collection of benchmarks." [Online]. Available: http://www.informatik.uni-freiburg.de/∼pigorsch/benchmarks.html

[23] P. Bjesse and A. Boralv, "DAG-aware circuit compression for formal verification," in *Int'l Conf. on CAD*, 2004, pp. 42–49.

[24] A. Mishchenko, S. Chatterjee, and R. Brayton, "DAG-aware AIG rewriting," in *Design Automation Conf.*, 2006, pp. 532–535.