turing lecture

## Turing Lecture from the winners of the 2007 ACM A.M. Turing Award.

In 1981, Edmund M. Clarke and E. Allen Emerson, working in the USA, and Joseph Sifakis working independently in France, authored seminal papers that founded what has become the highly successful field of model checking. This verification technology provides an algorithmic means of determining whether an abstract model—representing, for example, a hardware or software design—satisfies a formal specification expressed as a temporal logic (TL) formula. Moreover, if the property does not hold, the method identifies a counterexample execution that shows the source of the problem.

The progression of model checking to the point where it can be successfully used for complex systems has required the development of sophisticated means of coping with what is known as the state explosion problem. Great strides have been made on this problem over the past 28 years by what is now a very large international research community. As a result many major hardware and software companies are beginning to use model checking in practice. Examples of its use include the verification of VLSI circuits, communication protocols, software device drivers, real-time embedded systems, and security algorithms.

The work of Clarke, Emerson, and Sifakis continues to be central to the success of this research area. Their work over the years has led to the creation of new logics for specification, new verification algorithms, and surprising theoretical results. Model checking tools, created by both academic and industrial teams, have resulted in an entirely novel approach to verification and test case generation. This approach, for example, often enables engineers in the electronics industry to design complex systems with considerable assurance regarding the correctness of their initial designs. Model checking promises to have an even greater impact on the hardware and software industries in the future.
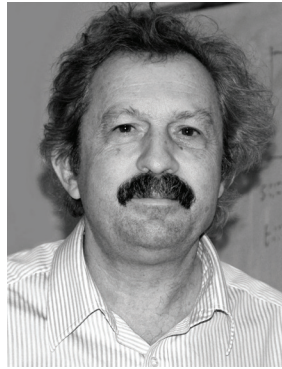
—*Moshe Y. Vardi, Editor-in-Chief*
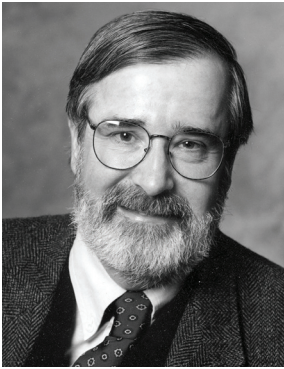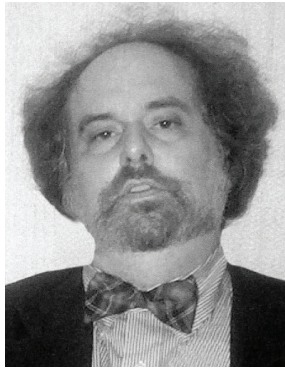
Edmund M. Clarke   **E. Allen Emerson**   **Joseph Sifakis**

Model Checking: Algorithmic Verification and Debugging

By Edmund M. Clarke, E. Allen Emerson, and Joseph Sifakis

1. E. ALLEN EMERSON
MODEL CHECKING: A BIRD'S-EYE VIEW

1.1. Formal Verification

Formal verification of program correctness hinges on the use of mathematical logic. A program is a mathematical object with well-defined, although possibly complex and intuitively unfathomable, behavior. Mathematical logic can be used to describe precisely what constitutes correct behavior. This makes it possible to contemplate mathematically establishing that the program behavior conforms to the correctness specification. In most early work this involved constructing a formal proof of correctness. In contradistinction, model checking avoids proofs.

Floyd-Hoare-style deductive verification was the prevailing mode of formal verification going back to the 1960s. This classic and elegant approach entailed manual proof construction, typically using axioms and inference rules in a formal deductive system, often oriented toward sequential programs. Such proof construction was tedious, difficult, and required human ingenuity. This field was a great intellectual success, spawning work on compositional or modular proof systems, soundness of program proof systems, and their completeness; see Section 3. Case studies confirmed that this approach really worked for small programs, although a short program might require a long proof. However, manual verification did not scale up well to large programs. The proofs were just too hard to construct.
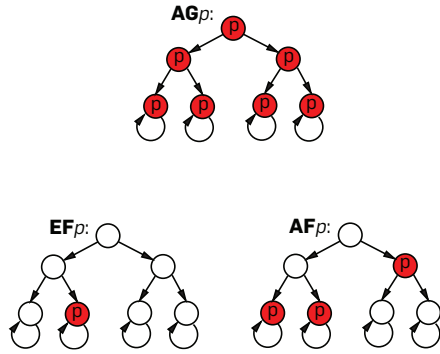
1.2. Temporal Logics

In view of the difficulties in trying to construct program proofs it seemed like there ought to be a better way. The way was inspired by the use of *Temporal Logic* (TL), a formalism for describing change over time. If a program can be specified in TL, it can be realized as a finite state system. This suggested the idea of model checking—to check if a finite state graph is a model of a TL specification.

The critical suggestion of using TL for reasoning about ongoing concurrent programs was made in Pnueli's landmark paper.[39] Such systems ideally exhibit nonterminating behavior so that they do not conform to the Hoare-style paradigm. They are also typically nondeterministic. Examples include hardware circuits, microprocessors, operating systems, banking networks, communication protocols, automotive electronics, and many modern medical devices. Pnueli used a TL with basic temporal operators **F** (*sometime*) and **G** (*always*). Augmented with **X** (*next-time*) and **U** (*until*), this is today known as LTL (Linear Time Logic).

Another widely used logic is CTL (Computation Tree Logic)[10] (cf. Emerson and Clarke, and Ben-Ari et al.[20, 4]). Its basic temporal modalities are **A** (for *all* futures) or **E** (for *some* future) followed by one of **F** (sometime), **G** (always), **X** (next-time), and **U** (until); compound formulae are built up from nestings and propositional combinations of CTL subformulae. CTL is a branching time logic as it can distinguish between **AF**p (along all futures, P eventually holds and

Figure 1. Basic temporal operators.

is thus inevitable) and **EF**p (along some future, P eventually holds and is thus possible). The branching time logic CTL* subsumes both CTL and LTL. (See Figure 1.)

Temporal logic formulae are interpreted over a given finite *state graph*, also called a *(Kripke) structure*, M comprised of a set S of *states*, a total binary *transition relation* $R \subseteq S \times S$, and a *labelling* L of states with atomic facts (propositions such as P) true there. There may also be a distinguished (start) state s_0. As usual in mathematical logic, to be precise in defining a logic we use the meta-notation $M, s_0 \models f$ as shorthand for "in structure M at state s_0 formula f is true," for f a CTL (or CTL*) formula. When s_0 is understood, we may write $M \models f$. For example, $M, s_0 \models \mathbf{AF}P$ iff for all paths $x = s_0, s_1, s_2, \ldots$ in M we have $\exists i \geq 0, P \in L(s_i)$.

When doing specification in practice we may write just **AF**p to assert that formula p is inevitable. An LTL formula h is interpreted over a path and then over a structure by implicit universal path quantification: in practical specification we write h but mean $\mathbf{A}h$.

The LTL formula $\mathbf{G}\neg(C_1 \wedge C_2)$ captures mutual exclusion for the critical sections, corresponding to assertions C_1 and C_2, of processes 1 and 2, respectively. In CTL, we would write $\mathbf{AG}\neg(C_1 \wedge C_2)$ for mutual exclusion, and $\mathbf{AG}(T_1 \Rightarrow \mathbf{AF}C_1)$ for "whenever process 1 enters its trying region (T_1) it inevitably enters its critical section (C_1)." The CTL formula **AGEF**$start$ asserts the system can always be restarted; this is not expressible in LTL. The CTL* formula **EGF**$send$ asserts the existence of a *fair* behavior along which the *send* condition occurs repeatedly. Such fairness conditions are important in ensuring that goals are fulfilled in concurrent systems.

The logics LTL, CTL, and CTL* have turned out to be very influential, spawning industrial extensions and uses, plus many academic applications as well as theoretical results. There are prominent industrial logics, tailored for hardware verification using special "macros," i.e., compact high-level operators that expand into longer combinations of basic operators. These include IBM Sugar based on CTL, Intel ForSpec based on LTL, and PSL (IEEE-1850 standard), incorporating features from CTL*.

Finally, there is also the (propositional) mu-calculus[31] (cf. Emerson and Clarke[20]), a particular but very general TL. It permits temporal correctness properties to be characterized as fixed points or *fixpoints* of recursive definitions. For example $\mathbf{EF}p = p \vee \mathbf{EX}(\mathbf{EF}p)$. The mu-calculus plays a vital role in model checking. It is very expressive: CTL, CTL*, as well as LTL, can be encoded in the mu-calculus. The fixed point characterizations of temporal correctness properties underlie many conventional and symbolic model checking algorithms, as well as tools used in practice.

1.3. Model Checking

In the early 1980s Clarke and Emerson proposed *model checking*, a method for automatic (and algorithmic) verification of finite state concurrent systems[10]; independently Quielle and Sifakis proposed essentially the same method.[41] In model checking, TL is used to specify correct system behavior. An efficient, flexible search procedure is used to find correct temporal patterns in the finite state graph of the concurrent system. The orientation of the method is to provide a practical verification method. The technical formulation of the model checking problem is simply: Given a finite structure M, state s, and a TL formula f, does $M, s \models f$? An alternative formulation is, given M and f, calculate $\{s : M, s \models f\}$. The main result of Clarke and Emerson[10] is that CTL model checking can be done in time $O(|f| \cdot |M|^2)$; that is, in time polynomial in the formula and the structure sizes. (The result in Queille and Sifakis[41] was with respect to a slighly weaker TL.)

The algorithm was based on fixpoint characterizations of basic temporal modalities. For example, let $f(Z)$ denote $p \vee \mathbf{AX}Z$. We see that $\mathbf{AF}p = f(\mathbf{AF}p)$ is a fixpoint of $f(Z)$, since $\mathbf{AF}p$ holds iff p holds or $\mathbf{AXAF}p$ holds. In general, there may be multiple fixpoints. It can be shown that $\mathbf{AF}p$ is the *least fixpoint*, which we shall write $\mu Z = f(Z)$, with $f(Z)$ as above. Intuitively, least fixpoints capture only well-founded or finite behaviors. The fixpoint characterization $\mu Z = f(Z)$ of a property makes it possible to calculate *iteratively* the set of states where $\mathbf{AF}p$ is true. This utilizes the fact that every formula corresponds to the set of states in which it is true. We compute the maximum of the ascending chain of increasingly larger under-approximations to the desired set of states: $false \subseteq f(false) \subseteq f^2(false) \subseteq \ldots \subseteq f^k(false) = f^{k+1}(false)$, where k is at most the size of the (finite) state space. More generally, the *Tarski–Knaster Theorem* (cf. Tarski[44]) permits the ascending iterative calculation $\cup f^i(false)$ of any temporal property r characterized as a least fixpoint $\mu Z = f(Z)$, provided that $f(Z)$ is *monotone*, which is ensured by Z only appearing un-negated. For greatest fixpoints, one starts the calculation at *true*. Essentially the same algorithm was given in Queille and Sifakis.[41]

The following are noteworthy extensions. CTL model checking can be done in time $O(|M| \cdot |f|)$,[11] i.e., linear in the size of the state graph and linear in the size of the formula. LTL model checking can be done in time $O(|M| \cdot \exp(|f|))$; since M is usually very large while f is small, the exponential factor may be tolerable.[33] The automata-theoretic approach to LTL model checking is described in Vardi and Wolper.[46] A succinct fixpoint characterization of fairness from Emerson and Lei[23] is used to make LTL model checking more efficient in practice. Branching time CTL* model

checking can be efficiently reduced to linear time LTL model checking for the same overall bound.[24]

1.4. Expressiveness

An important criterion for a logic is expressiveness, reflecting what correctness properties can and cannot be captured by the logic. Interesting properties include safety properties ("nothing bad happens," e.g., $\mathbf{G}\neg bad$), liveness properties ("something good happens," e.g., $\mathbf{F}goal$), and fairness properties ("something is recurrent", e.g., $\mathbf{GF}try$). It is arguable that expressiveness in model checking is the most fundamental characteristic, perhaps even more critical than efficiency. It is imperative that one be able to express all the correctness properties that are needed. If this basic requirement is not met, there is no point in using the verification method in the first place. In actual usage, a particular formalism, commonly a system of TL, provides the needed expressive power. It includes a few basic temporal operators, which can be combined to yield virtually limitless assertions. Another benefit of TL is that it is related to natural language, which can facilitate its use.

The ability to describe complex patterns of system behavior is basic. LTL is naturally suited to the task. Along paths, it is in a sense expressively complete, equivalent to the First Order Language of Linear Order,[19] e.g. $\mathbf{G}P = \forall t\,(t \geq 0 \Rightarrow P(t))$. A property such as $\mathbf{G}_2 P$, meaning that P holds at all even moments $0, 2, 4, \ldots$ is not expressible in LTL. It can be useful in hardware verification applications where it is needed to count clock cycles. The (linear time) mu-calculus as well as PSL can express this property (cf. Wolper[47]).

CTL is well suited to capture correctness over computation trees. The branching time capability of distinguishing between necessary and possible behaviors using explicit path quantifiers ($\mathbf{A}$, $\mathbf{E}$) provides significant expressive power. The existence of a bad path, $\mathbf{EF}bad$, is not expressible by any formula $\mathbf{A}h$ where h is in LTL, nor even any universal CTL* formula where all path quantifiers are $\mathbf{A}$ (and only atomic propositions appear negated). Thus, LTL is not closed under semantic negation: writing the invariant $\mathbf{G}\neg bad$ means $\mathbf{AG}\neg bad$ whose semantic negation is $\mathbf{EF}bad$ which, as above, is not expressible by any $\mathbf{A}h$ formula.[21] There has been an ongoing debate as to whether LTL or branching time logic is better for program reasoning. Linear time offers the advantage of simplicity, but at the cost of significantly less expressiveness. Branching time's potentially greater expressiveness may incur greater conceptual (and computational) complexity.

A related criterion is succinctness, reflecting how compactly a property can be expressed. The CTL* formula $\mathbf{E}(\mathbf{F}P_1 \wedge \mathbf{F}P_2)$ is not a CTL formula, but is semantically equivalent to the longer CTL formula $\mathbf{EF}(P_1 \wedge \mathbf{EF}P_2) \vee \mathbf{EF}(P_2 \wedge \mathbf{EF}P_1)$. For n conjuncts, the translation is exponential in n. In practice, the most important is the criterion of convenience, reflecting how easily and naturally properties can be expressed. Expressiveness and succinctness may be partially amenable to mathematical definition and investigation. Succinctness and convenience often correlate but not always. Convenience, however, is inherently informal.

Yet it is extremely important in actual use. That is why, e.g., many person-years were devoted to formulating industrial-strength logics such as PSL.

1.5. Efficiency

Another important criterion, efficiency, is related to questions of the complexity of the model checking problem for a logic and the performance of model checking algorithms for the logic. An algorithm that has potentially high complexity in theory but is repeatedly observed to exhibit significantly lower complexity in actual use is likely to be preferred to one with better theoretical complexity but inferior observed performance. Moreover, there are trade-offs. For instance, a more expressive logic is likely to be less efficient. A more succinct logic is likely to be more convenient yet even less efficient. Some experience is required to reach a good trade-off. For many model checking applications, M is sufficiently small that it can be explicitly represented in computer memory. Such basic enumerative model checking may be adequate for systems with 10^6 states.

However, many more systems M have an astronomically or even infinitely large state space. There are some fundamental strategies to cope with large state spaces. Foremost, is the use of abstraction where the original, large, complex system M is simplified, by suppressing inessential detail (cf. Clarke and Emerson[10]), to get a (representation of a) smaller and simpler system $\overline{M}$.

Compact representations of the state graph yield another important strategy. The advent of symbolic model checking, combining CTL, fixpoint computation, and data structures for compact representation of large state sets, made it possible to check many systems with an astronomical number of states (cf. Burch et al.[8]).

If there are many replicated or similar subcomponents, it is often possible to factor out the inherent symmetry in the original M resulting in an exponentially reduced abstract $\overline{M}$[43] (cf. Sistla et al. and Clarke et al.[43, 9]). Most work on symmetry has required the use of explicit representation of M. Natural attempts to combine symmetry and symbolic representation were shown inherently infeasible.[14] However, a very advantageous combination based on dynamically reorganizing the symbolic representation overcomes these limitations.[25] Finally, one may have an infinite state system comprised of, e.g., a (candidate) dining philosophers solution M_n for all sizes $n > 1$. In many situations, this parameterized correctness problem is reducible to model checking a fixed finite-size system M_c (cf. Clarke et al., and Emerson and Kahlon[9, 22]).

1.6. Evolution of Model Checking

The early reception of model checking was restrained. Model checking originated in the theoretical atmosphere of the early 1980s. There was a field of study known as Logics of Programs, which dealt with the theory and sometime use of logic for reasoning about programs. Various modal and temporal logics played a prominent role. The key technical issue under investigation for such a logic was satisfiability: Given any formula f, determine whether there exists some structure M such that $M \models f$. Analyzing the decidability and

complexity of satisfiability for these logics was the major focus. However, model checking refers to the truth under *one* given interpretation *M* of a given formula *f*. This notion was implicit in the Tarskian definition of truth but, classically, was not viewed as an interesting problem. The idea that model checking should provide for verification of finite state systems was not appreciated. The early reaction to model checking then was mostly one of confusion and disinterest. It seemed a disconcerting novelty. It was not satisfiability. It was not validity. What was it? It was even dubbed "disorienting." Many felt it could not possibly work well in practice. In more recent times, some more favorable comments have been made. Model checking is "an acceptable crutch"—Edsger W. Dijkstra; it is "a first step toward engineerization of the field"—A. Pnueli.[40]

What factors contributed to model checking's successful deployment? First, the initial framework was feasible and comprehensible. It built on a helpful combination of TL and algorithms. It provided a "push-button," i.e., automated, method for verification. It permitted bug detection as well as verification of correctness. Since most programs are wrong, this is enormously important in practice. Incidently, the limited support for bug detection in proof-theoretic verification approaches contributes to their slower adoption rate. Moreover, while a methodology of constructing a program hand-in-hand with its proof certainly has its merits, it is not readily automatable. This hampers its deployment. With model checking, the separation of system development from verification and debugging (see Section 2) has undoubtedly facilitated model checking's industrial acceptance. The development team can go ahead and produce various aspects of the system under design. The team of verifiers or verification engineers can conduct verification independently. Hopefully, many subtle bugs will be detected and fixed. As a practical matter, the system can go into production at whatever level of "acceptable correctness" prevails at deadline time. Lastly, Moore's Law has engendered larger computer main memory, which enabled the development of ever more powerful model checking tools.

1.7. Discussion and Summary
What are the key accomplishments of model checking? The key contribution is that verification using model checking is now done *routinely* on a widespread basis for many large systems, including industrial-strength systems. Large organizations from hardware vendors to government agencies depend on model checking to facilitate achieving their goals. In contrast to 28 years ago, we no longer just talk about verification; we do it. The somewhat surprising conceptual finding is that verification can be done extremely well by automated search rather than manual proofs.

Model checking realizes in small part the *Dream of Leibniz [1646–1716]* (cf. Davis[18]). This was a proposal for a universal reasoning system. It was comprised of a *lingua characteristica universalis*, a language in which all knowledge could be formally expressed. TL plays a limited formulation of this role. There was also a *calculus ratiocinator*, a method of calculating the truth value of such a formalized assertion. Model checking algorithms provide the means of calculating truth. We hope that, over time, model checking will realize an increasingly large portion of Leibniz' Dream.

2. EDMUND M. CLARKE
MY 28-YEAR QUEST TO CONQUER THE STATE EXPLOSION PROBLEM

2.1. Model Checkers and Debugging
Model checkers typically have three main components: (1) a *specification language*, based on propositional TL,[39] (2) a way of encoding a state machine representing the system to be verified, and (3) a *verification procedure*, that uses an *intelligent* exhaustive search of the state space to determine if the specification is true or not. If the specification is not satisfied, then most model checkers will produce a counterexample execution trace that shows why the specification does not hold. It is impossible to overestimate the importance of this feature. The counterexamples are invaluable in debugging complex systems. Some people use model checking just for this feature. The EMC model checker[11] did not give counterexamples for universal CTL properties that were false or witnesses for existential properties that were true. Michael C. Browne added this feature to the MCB model checker in 1984. It has been an important feature of model checkers ever since. (See Figure 2.)

Figure 2. A model checker with counterexamples.
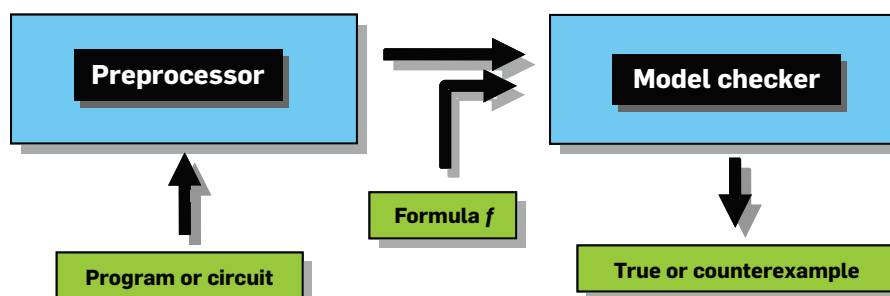
2.2. State Explosion Problem

State explosion is the major problem in model checking. The number of global states of a concurrent system with many processes can be enormous. It is easy to see why this is true. The asynchronous composition of n processes, each having m states, may have m^n states. A similar problem occurs with data. The state-transition system for an n-bit counter will have 2^n states. All model checkers suffer from this problem. Complexity-theoretic arguments can be used to show that the problem is unavoidable in the worst case (assuming P is different than PSPACE). Fortunately, steady progress has been made over the past 28 years for special types of systems that occur frequently in practice. In fact, the state explosion problem has been the driving force behind much of the research in model checking and the development of new model checkers. We discuss below key breakthroughs that have been made and some of the important cases where additional research is needed.

2.3. Major Breakthroughs

2.3.1. Symbolic Model Checking with OBDDs. In the original implementation of the model checking algorithm, transition relations were represented explicitly by adjacency lists.[11] For concurrent systems with small numbers of processes, the number of states was usually fairly small, and the approach was often quite practical. In systems with many concurrent parts the number of states in the global state-transition system was too large to handle. In the fall of 1987, McMillan, then a graduate student at Carnegie Mellon, realized that by using a symbolic representation for the state-transition systems, much larger systems could be verified. The new symbolic representation was based on Bryant's *ordered binary decision diagrams* (OBDDs). OBDDs provide a canonical form for Boolean formulas that is often substantially more compact than conjunctive or disjunctive normal form, and very efficient algorithms have been developed for manipulating them. Because the symbolic representation captures some of the regularity in the state space determined by circuits and protocols, it is possible to verify systems with an extremely large number of states—many orders of magnitude larger than could be handled by the explicit-state algorithms. With the new representation for state-transition systems, we could verify some examples that had more than 10^{20} states.[8, 35] Since then, various refinements of the OBDD-based techniques have pushed the state count up to more than 10^{120}.

2.3.2. Partial Order Reduction. Verifying software poses significant problems for model checking. Software tends to be less structured than hardware. In addition, concurrent software is usually *asynchronous*, i.e., most of the activities taken by different processes are performed independently, without a global synchronizing clock. For these reasons, the state explosion problem is particularly serious for software. Consequently, model checking has been used less frequently for software verification than for hardware verification. One of the most successful techniques for dealing with asynchronous systems is the *partial order reduction*. This technique exploits the independence of concurrently executed events. Intuitively, two events are *independent* of each other when executing them in either order results in the same global state. In this case, it is possible to avoid exploring certain paths in the state-transition system. Model checking algorithms that incorporate the partial order reduction are described in several different papers. The *stubborn sets* of Valmari,[45] the *persistent sets* of Godefroid[27] and the *ample sets* of Peled,[38] differ on the actual details, but contain many similar ideas. The SPIN model checker developed by Holzmann uses the ample-set reduction to great advantage.

2.3.3. Bounded Model Checking with SAT. Although symbolic model checking with OBDDs was the first big breakthrough on the state explosion problem and is still widely used, OBDDs have a number of problems that limit the size of the models that can be checked with this technique. The ordering of variables on each path from the root of the OBDD to a leaf has to be the same. Finding an ordering that results in a small OBDD is quite difficult. In fact, for some Boolean formulas no space-efficient ordering is possible. A simple example is the formula for the middle output bit of a combinational multiplier for two n-bit numbers. It is possible to prove that the OBDD for this formula has size that is exponential in n for all variable orderings.

Propositional satisfiability (SAT) is the problem of determining whether a propositional formula in conjunctive normal form ("product of sums form" for Boolean formulas) has a truth assignment that makes the formula true. The problem is NP-complete (in fact, it is usually the first example of this class that students see). Nevertheless, the increase in power of modern SAT solvers over the past 15 years on problems that occur in practice has been phenomenal. It has become the key enabling technology in applications of model checking to both computer hardware and software. Bounded Model Checking (BMC) of computer hardware using a fast SAT solver is now probably the most widely used model checking technique. The counterexamples that it finds are just the satisfying instances of the propositional formula obtained by unwinding to some fixed depth the state-transition system for the circuit and the negation of its specification in linear TL.

The basic idea for BMC is quite simple (cf. Biere et al.[7]). The extension to full LTL obscures the simplicity so we will just describe how to check properties of the form **F**P where the property P is an atomic proposition (e.g., "Message_ Received"). BMC determines whether there is a counterexample of length k (we assume $k \geq 1$). In other words, it checks if there is a path of length k ending in a cycle in which each state is labeled with $\neg P$ (see Figure 3). Assume that the state-transition system M has n states. Each state can be encoded by a vector $\vec{v}$ of $\lceil \log(n) \rceil$ Boolean variables.

Figure 3. Counterexample of length at most k.

The set of initial states can be specified by a propositional formula $I(\vec{v})$ which holds for exactly those assignments to $\vec{v}$ that correspond to initial states. Likewise, the transition relation can be given by a propositional formula $R(\vec{v}, \vec{v}')$. A path of length k starting in an initial state can be encoded by means of the following formula:

$$\text{path}(k) = I(\vec{v}_0) \wedge R(\vec{v}_0, \vec{v}_1) \wedge \cdots \wedge R(\vec{v}_{k-1}, \vec{v}_k). \quad (1)$$

The path ends in a cycle if and only if

$$\text{cycle}(k) = R(\vec{v}_k, \vec{v}_0) \vee \cdots \vee R(\vec{v}_k, \vec{v}_{k-1}) \vee R(\vec{v}_k, \vec{v}_k). \quad (2)$$

The property P is false in each of the k steps if and only if

$$\text{property}(k) = \neg P(\vec{v}_0) \wedge \neg P(\vec{v}_1) \wedge \cdots \wedge \neg P(\vec{v}_k). \quad (3)$$

Thus, the liveness property $\mathbf{F}P$ has a counterexample of length k if and only if the conjunction $\Omega(k)$ of Formulas 1, 2, and 3 is satisfiable.

$$\Omega(k) = \text{path}(k) \wedge \text{cycle}(k) \wedge \text{property}(k). \quad (4)$$

We start with $k = 1$. If the formula $\Omega(k)$ is satisfiable, we know that $\mathbf{F}P$ has a counterexample of length k. A counterexample execution trace can be extracted from the satisfying assignment to $\Omega(k)$. If the formula $\Omega(k)$ is not satisfiable, then it could be the case that either the temporal formula $\mathbf{F}P$ holds on all paths starting from an initial state (and our specification is true) or there is a counterexample that is longer than k. When $\Omega(k)$ is unsatisfiable, we can do one of two things: Either increase the value of k and look for longer counterexamples or stop if time or memory constraints are exceeded.

We note that the idea of checking safety properties such as $\mathbf{G}P$ by reduction to propositional satisfiability is implicit in the work of Kautz and Selman.[30] However, they do not consider more general temporal properties such as the liveness property that we consider above.

In practice, BMC can often find counterexamples in circuits with thousands of latches and inputs. Armin Biere recently reported an example in which the circuit had 9510 latches and 9499 inputs. This resulted in a propositional formula with 4×10^6 variables and 1.2×10^7 clauses. The shortest bug of length 37 was found in 69 seconds! Many others have reported similar results.

Can BMC ever be used to prove correctness if no counterexamples are found? It is easy to see that for safety and liveness properties of the form $\mathbf{F}P$ and $\mathbf{G}P$ where P is a propositional formula, if there is a counterexample, then there is one that is less than the diameter (i.e., the longest shortest path between any two states) of the state-transition system. So, the diameter could be used to place an upper bound on how much the transition relation would need to be unwound. Unfortunately, it appears to be computationally difficult to compute the diameter when the state-transition system is given implicitly as a circuit or in terms of propositional formulas for the set of initial states, the transition relation, and the set of bad states. Other ways for making BMC complete are based on cube

enlargement,[36] circuit co-factoring,[26] induction,[42] and Craig interpolants.[37] But, the problem remains a topic of active research. Meanwhile, an efficient way of finding subtle counterexamples is still quite useful in debugging circuit designs.

2.3.4. The Abstraction Refinement Loop. This technique uses counterexamples to refine an initial abstraction. We begin by defining what it means for one state-transition system to be an abstraction of another. We write $M_\alpha = \langle S_\alpha, s_0^\alpha, R_\alpha, L_\alpha \rangle$ to denote the *abstraction* of state-transition system $M = \langle S, s_0, R, L \rangle$ with respect to an *abstraction mapping* α. (Here we include the start states s_0 and s_0^α as parts of the state-transition systems.) We assume that the states of M are labeled with atomic propositions from a set A of atomic propositions, and that M_α is labeled with atomic propositions from a set A_α that is a subset of A. We call M the *concrete* system and M_α the *abstract* system.

DEFINITION 1. *A function* $\alpha: S \to S_\alpha$ *is an* abstraction mapping *from the concrete system M to the abstract system M_α with respect to the propositions in A_α if and only if*

- $\alpha(s_0) = s_0^\alpha$.
- *If there is a transition from state s to state t in M, then there is a transition from $\alpha(s)$ to $\alpha(t)$ in M_α.*
- *For all states s, $L(s) \cap A_\alpha = L_\alpha(\alpha(s))$.*

The three conditions ensure that M_α *simulates* M. Note that only identically labeled states of the concrete model (modulo propositions absent from A_α) will be mapped into the same state of the abstract model (see Figure 4). The key theorem relating concrete and abstract systems is the *Property Preservation Theorem*:

THEOREM 1 (CLARKE, GRUMBERG, and LONG[13]). *If a universal CTL* property holds on the abstract model, then it holds on the concrete model.*

Here, a universal CTL* property is one that contains no existential path quantifiers when written in negation-normal form. For example, $\mathbf{AF}P$ is a universal property but $\mathbf{EF}P$ is not.

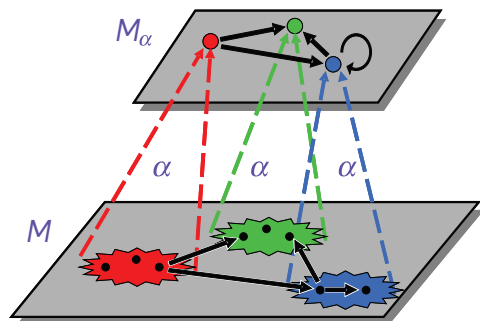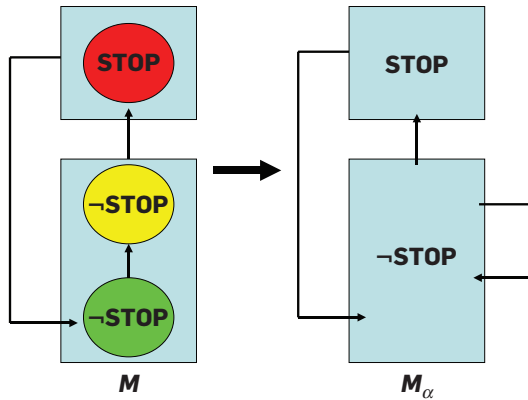**Figure 4. A concrete system and its abstraction.**

Figure 5. Spurious counterexample.

The converse of the theorem is not true as Figure 5 illustrates. A universal property that holds in the concrete system may fail to hold in the abstract system. For example, the property **AGF** *STOP* (*infinitely often* STOP) holds in *M*, but not in M_α. Thus, a counterexample to the property in the abstract system may fail to be a counterexample in the concrete system. Such counterexamples are said to be *spurious* counterexamples. This leads to a verification technique called *Counterexample Guided Abstraction Refinement* (CEGAR).[12] Universal properties are checked on a series of increasingly precise abstractions of the original system. If the property holds, then by the Property Preservation Theorem, it must hold on the concrete system and we can stop. If it does not hold and we get a counterexample, then we must check the counterexample on the concrete system in order to make sure that it is not spurious. If the counterexample checks on the concrete system, then we have found an error and can also stop. If the counterexample is

spurious, then we use information in the counterexample to refine the abstraction mapping and repeat the loop. The CEGAR Loop in Figure 6 generalizes an earlier abstraction technique for sequential circuits called the *localization reduction*, which was developed by R. Kurshan.[32] CEGAR is used in many software model checkers including the SLAM Project at Microsoft.[1]

2.4. State Explosion Challenges for the Future

The state explosion problem is likely to remain the major challenge in model checking. There are many directions for future research on this problem, some of which are listed below.

- Software model checking, in particular, combining model checking and static analysis
- Effective model checking algorithms for real-time and hybrid systems
- Compositional model checking of complex systems
- Symmetry reduction and parameterized model checking
- Probabilistic and statistical model checking
- Combining model checking and theorem proving
- Interpreting long counterexamples
- Scaling up even more

3. JOSEPH SIFAKIS
THE QUEST FOR CORRECTNESS:
CHALLENGES AND PERSPECTIVES

3.1. Where Are We Today?

Verification techniques have definitely found important applications. After the first two decades of intensive research and development, recent years have been characterized by a shift in focus and intensity.

Figure 6. The CEGAR loop.
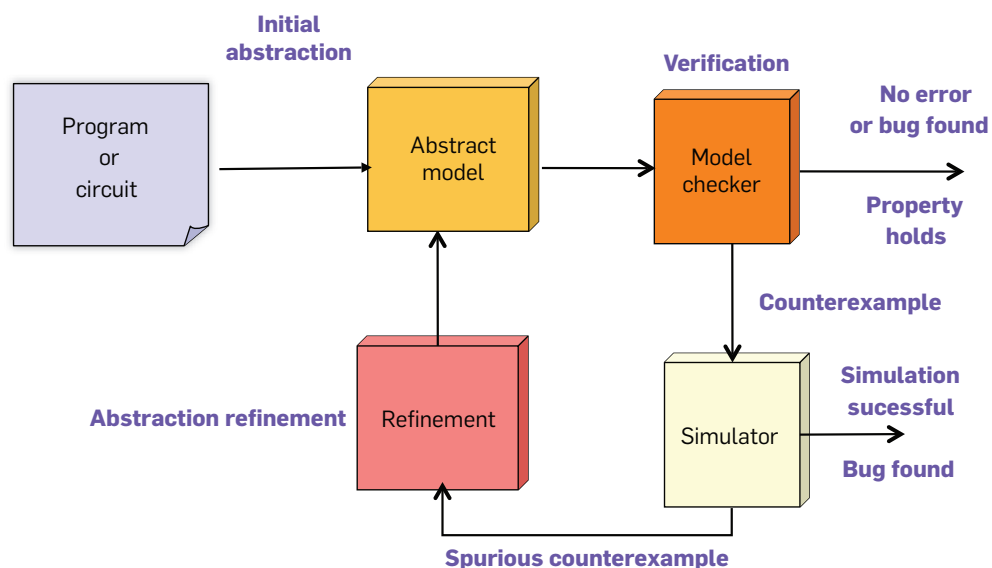
Algorithmic verification involves three different tasks: (1) requirements specification, (2) building executable system models, and (3) developing scalable algorithms both for checking requirements and for providing diagnostics when requirements are not met. The status for each of these tasks is discussed below.

3.1.1. Requirements Specification.
Requirements characterize the expected behavior of a system. They can be expressed following two paradigms. *State-based* requirements specify a system's observable behavior by using transition systems. *Property-based* requirements use a declarative style. These requirements are expressed as sets of formulas in a formalism such as a TL. A combination of the two paradigms is necessary for enhanced expressiveness, such as in the PSL language. The state-based paradigm is adequate for characterizing causal dependencies between events, e.g., sequences of actions. In contrast, the property-based paradigm is more appropriate for global properties, e.g., liveness and mutual exclusion. For concurrent systems, an important trend is toward semantic variations of state-based formalisms such as Live Sequence Charts.[17]

Using TLs has certainly been a breakthrough in understanding and formalizing requirements for concurrent systems. Nonetheless, subtle differences in the formulation of common concepts such as liveness and fairness, which depend on the underlying time model (e.g., branching or linear time), show that writing rigorous logic specifications is not trivial.

Furthermore, the declarative and dense style in the expression of property-based requirements is not always easy to master and understand. Requirements must be *sound*. That is, they must be satisfiable by some model. In addition, they must be *complete*. That is, no important information is omitted about the specified system. In contrast to soundness, which is a well-understood property and can be checked automatically by using decision procedures, there is no consensus as to what precisely constitutes completeness in requirements specifications, nor how to go about achieving it. Absolute completeness, which means that specifications describe the system exactly, has only a theoretical interest and is probably unattainable for non-trivial systems.

Existing requirements specification formalisms are mainly appropriate for expressing functional requirements. We lack rigorous formalisms for extra-functional requirements for security properties (e.g., privacy), reconfigurability properties (e.g., noninterference of configurable features), and quality of service (e.g., degree of jitter).

3.1.2. Building Executable Models.
Successful application of verification methods requires techniques for building executable models that *faithfully* represent a system or an abstraction of it. Faithfulness means that the system to be verified and its model are related through a checkable semantics–preserving relation. This will ensure soundness of the model. In other words, any property that we can verify for the model will hold for the real system. Furthermore, to avoid errors in building models and to cope with their complexity, models should be generated automatically from system descriptions.

For hardware verification, it is relatively straightforward to generate exact logical finite-state models, expressed as systems of boolean equations, e.g., from RTL descriptions. This probably explains the strong and immediate success of model checking in the area. For software, the problem is more difficult. In contrast to logical hardware models, we need to define formally the semantics of the programming language. This may not be an easy task for languages such as C or Java, as it requires some clarification of concepts and additional assumptions about their semantics. Once the semantics is fixed, tractable models can be extracted from real software through abstraction. This allows us to cope with complexity of data and dynamic features. Currently, we do not know how to build faithful models for systems consisting of hardware and software, at the same level of detail as for pure hardware or software. Ideally, for a system consisting of application software running on a platform, the corresponding model could be obtained as the composition of models for the software and the platform. The main difficulty is in understanding and formalizing the interaction between these two types of models, in particular by taking into account timing aspects and resources such as memory and energy. In addition, this should be done at some adequate level of abstraction, allowing tractable models.

Today, we can specify and verify only high-level timed models with tools such as Uppaal[3] for schedulability analysis. These models take into account hardware timing aspects and some abstraction of the application software. The validation of even relatively simple systems such as a node in a wireless sensor network is carried out by testing physical prototypes or by ad-hoc simulation. We need theory, methods, and tools for modeling complex heterogeneous systems.[2] Weaknesses in the state of the art are also seen in standards and languages for system modeling. Efforts for extending UML to cover scheduling and resource management issues have failed to provide a rigorous basis for this. At the same time, extensions of hardware description languages to encompass more asynchronous execution models such as SystemC and TLM can be used only for simulation, due to a lack of formal semantic foundations.

3.1.3. Scalable Verification Methods.
Today we have fairly efficient verification algorithms. However, all suffer from well-known inherent complexity limitations when applied to large systems. To cope with this complexity, I see two main avenues.

The first avenue is to develop new abstraction techniques, in particular for specific semantic domains depending on the data handled by the system and on the properties to be verified. The convergence between model checking and abstract interpretation[16] could lead to significant breakthroughs. These two main algorithmic approaches, which have developed rather independently for almost three decades, have a common foundation: solving fixpoint equations in specific semantic domains.

Initially, model checking focused on the verification of finite state systems such as hardware or complex

control-intensive reactive systems such as communication protocols. Later, research on model checking addressed verification of infinite state systems by using abstractions.[13, 34] The evolution of abstract interpretation is driven by the concern for finding adequate abstract domains for efficient verification of program properties by computing approximations of reachability sets. Model checking has had a broader application scope, including hardware, software, and systems. Furthermore, depending on the type of properties to be checked, model checking algorithms may involve computation of multiple fixed points. I believe that the combination of the two algorithmic approaches can still lead to significant progress in the state of the art, e.g., by using libraries of abstract domains in model checking algorithms.

The second avenue addresses significant long-term progress in defeating complexity. It involves moving from monolithic verification to compositional techniques. We need divide-and-conquer approaches for inferring global properties of a system from the properties of its components. The current state of the art does not meet our initial expectations. The main approach is by "assume-guarantee," where properties are decomposed into two parts. One is an assumption about the global behavior of the system within which the component resides; the other is a property guaranteed by the component when the assumption about its environment holds. As discussed in a recent paper,[15] many issues make it difficult to apply assume-guarantee rules, in particular because synthesis of assumptions (when feasible) may cost as much as monolithic verification.

In my opinion, any *general* compositional verification theory will be highly intractable and will be of theoretical interest only. We need to study compositionality results for particular classes of properties and/or particular classes of systems as explained below.

3.2. From a posteriori Verification to Constructivity

A big difference between Computer Engineering and more mature disciplines based on Physics, e.g., Electrical Engineering, is the importance of verification for achieving correctness. These disciplines have developed theory guaranteeing by construction the correctness and predictability of artifacts. For instance, the application of Kirchoff's laws allows building circuits that meet given properties.

My vision is to investigate links between compositional verification for specific properties and results allowing constructivity. Currently, there exists in Computer Science an important body of constructivity results about architectures and distributed algorithms.

1. We need theory and methods for building faithful models of complex systems as the composition of heterogeneous components, e.g., mixed software/hardware systems. This is a central problem for ensuring correct interoperation, and meaningful refinement and integration of heterogeneous viewpoints. Heterogeneity has three fundamental sources which appear when composing components with different (a) execution models, e.g., synchronous and asynchronous execution, (b) interaction mechanisms such as locks, monitors, function calls, and message passing, and (c) granularity of execution, e.g., hardware and software.[29]

We need to move from composition frameworks based on the use of a single low-level parallel composition operator, e.g., automata-based composition, to a unified composition paradigm encompassing architectural features such as protocols, schedulers, and buses.

2. In contrast to existing approaches, we should investigate compositionality techniques for high-level composition operators and specific classes of properties. I propose to investigate two independent directions:

   • One direction is studying techniques for specific classes of properties. For instance, finding compositional verification rules guaranteeing deadlock-freedom or mutual exclusion instead of investigating rules for safety properties in general. Potential deadlocks can be found by analysis of dependencies induced by interactions between components.[28] For proving mutual exclusion, a different type of analysis is needed.
   • The other direction is studying techniques for particular architectures. Architectures characterize the way interaction among a system's components is organized. For instance, we might profitably study compositional verification rules for ring or star architectures, for real-time systems with preemptable tasks and fixed priorities, for time-triggered architectures, etc. Compositional verification rules should be applied to high-level coordination mechanisms used at the architecture level, without translating them into a low-level automata-based composition.

The results thus obtained should allow us to identify "verifiability" conditions (i.e., conditions under which verification of a particular property and/or class of systems becomes scalable). This is similar to finding conditions for making systems testable, adaptable, etc. In this manner, compositionality rules can be turned into correct-by-construction techniques.

Recent results implemented in the D-Finder tool[5, 6] provide some illustration of these ideas. D-Finder uses heuristics for proving compositionally global deadlock-freedom of a component-based system, from the deadlock-freedom of its components. The method is compositional and proceeds in two steps.

   • First, it checks that individual components are deadlock-free. That is, they may block only at states where they are waiting for synchronization with other components.
   • Second, it checks if the components' interaction graph is acyclic. This is a sufficient condition for establishing global deadlock-freedom at low cost. It depends only on the system architecture. Otherwise, D-Finder symbolically

computes increasingly strong global deadlock-free invariants of the system, based on results from the first step. Deadlock-freedom is established if there exists some invariant that is satisfied by the system's initial state.

Benchmarks published in Bensalem et al.[6] show that such a specialization for deadlock-freedom, combined with compositionality techniques, leads to significantly better performance than is possible with general-purpose monolithic verification tools.

A posteriori verification is not the only way to guarantee correctness. System designers develop complex systems, by carefully applying architectural principles that are operationally relevant and technically successful. Verification should advantageously take into account architectures and their features. There is a large space to be explored, between full constructivity and a posteriori verification. This vision can contribute to bridging the gap between Formal Methods and the body of constructivity results in Computer Science. ⓒ

References

1. Ball, T., Rajamani, S.K. The SLAM toolkit. In *Computer-Aided Verification (CAV'01)*. Volume 2102 of *Lecture Notes in Computer Science* (2001), 260–264.
2. Basu, A., Bozga, M., Sifakis, J. Modeling heterogeneous real-time components in BIP. In *SEFM* (2006), 3–12.
3. Behrmann, G., Cougnard, A., David, A., Fleury, E., Larsen, K.G., Lime, D. Uppaal-tiga: Time for playing games! In *CAV*. W. Damm and H. Hermanns, eds. Volume 4590 of *Lecture Notes in Computer Science* (Springer, 2007), 121–125.
4. Ben-Ari, M., Pnueli, A., Manna, Z. The temporal logic of branching time. *Acta Inf. 20* (1983), 207–226.
5. Bensalem, S., Bozga, M., Nguyen, T.-H., Sifakis, J. D-finder: A tool for compositional deadlock detection and verification. In *CAV*. A. Bouajjani and O. Maler, eds. Volume 5643 of *Lecture Notes in Computer Science* (Springer, 2009), 614–619.
6. Bensalem, S., Bozga, M., Sifakis, J., Nguyen, T.-H. Compositional verification for component-based systems and application. In *ATVA*. S.-D. Cha, J.-Y. Choi, M. Kim, I. Lee, and M. Viswanathan, eds. Volume 5311 of *Lecture Notes in Computer Science* (Springer, 2008), 64–79.
7. Biere, A., Cimatti, A., Clarke, E.M., Zhu, Y. Symbolic model checking without BDDs. In *Proceedings of the 5th International Conference on Tools and Algorithms for Construction and Analysis of Systems (TACAS'99)*. R. Cleaveland, ed. Volume 1579 of *Lecture Notes in Computer Science* (Springer-Verlag, Mar. 1999), 193–207.
8. Burch, J.R., Clarke, E.M., McMillan, K.L., Dill, D.L., Hwang, L.J. Symbolic Model Checking: 10^{20} states and beyond. *Inf. Comput. 98*, 2 (June 1992), 142–170. Originally presented at the 1990 Symposium on Logic in Computer Science (LICS'90).
9. Clarke, E., Grumberg, O., Peled, D. *Model Checking*. MIT Press, 1999.
10. Clarke, E.M., Emerson, E.A. Design and synthesis of synchronization skeletons using branching time temporal logic. In *Logics of Programs: Workshop, Yorktown Heights, NY, May 1981*. Volume 131 of *Lecture Notes in Computer Science* (Springer, 1981). 52–71.
11. Clarke, E.M., Emerson, E.A., Sistla, A.P. Automatic verification of finite-state concurrent systems using temporal logic specifications. *ACM Trans. Prog. Lang. Syst. 8*, 2 (1986), 244–263. Originally presented at the 1983 Symposium on Principles of Programming Languages (POPL'83).
12. Clarke, E.M., Grumberg, O., Jha, S., Lu, Y., Veith, H. Counterexample-guided abstraction refinement for symbolic Model Checking. *J. ACM 50*, 5 (2003), 752–794. Originally presented at the 2000 Conference on Computer-Aided Verification (CAV'00).
13. Clarke, E.M., Grumberg, O., Long, D.E. Model checking and abstraction. *ACM Trans. Program. Lang. Syst. 16*, 5 (1994), 1512–1542.
14. Clarke, E.M., Jha, S., Enders, R., Filkorn, T. Exploiting symmetry in temporal logic model checking. *Formal Methods Sys Design 9*, 1/2 (1996), 77–104.
15. Cobleigh, J.M., Avrunin, G.S., Clarke, L.A. Breaking up is hard to do: An evaluation of automated assume-guarantee reasoning. *ACM Trans. Softw. Eng. Methodol. 17*, 2 (2008), 1–52.
16. Cousot, P., Cousot, R. Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fixpoints. In *POPL* (1977), 238–252.
17. Damm, W., Harel, D. LSCs: Breathing life into message sequence charts. *Formal Methods Sys. Design 19*, 1 (2001), 45–80.
18. Davis, M. *The Universal Computer: The Road from Leibniz to Turing*. W. W. Norton & Co., 2000.
19. Emerson, E.A. Temporal and modal logic. In *Handbook of Theoretical Computer Science*. J. van Leeuwen, ed. Volume B, chapter 16, Elsevier Science (1990), 995–1072.
20. Emerson, E.A., Clarke, E.M. Characterizing correctness properties of parallel programs using fixpoints. In *Lecture Notes in Computer Science 85*, Automata, Languages and Programming (July 1980), 169–181.
21. Emerson, E.A. Halpern, J.Y. "Sometimes" and "Not Never" revisited: On branching time versus linear time. *J. ACM 33* (1986), 151–178.
22. Emerson, E.A., Kahlon, V. Reducing model checking of the many to the few. In *CADE*. D.A. McAllester, ed. Volume 1831 of *Lecture Notes in Computer Science* (Springer, 2000), 236–254.
23. Emerson, E.A., Lei, C.-L. Efficient model checking in fragments of the propositional mu-calculus (extended abstract). In *Proceedings, Symposium on Logic in Computer Science, 16–18 June 1986*, Cambridge, MA, USA, 1986, 267–278.
24. Emerson, E.A., Lei, C.-L. Modalities for model checking: Branching time logic strikes back. *Sci. Comput. Progr. 8*, 3 (1987), 275–306.
25. Emerson, E.A., Wahl, T. Dynamic symmetry reduction. In *TACAS*. N. Halbwachs and L.D. Zuck, eds. Volume 3440 of *Lecture Notes in Computer Science* (Springer, 2005), 382–396.
26. Ganai, M.K., Gupta, A., Ashar, P. Efficient SAT-based unbounded symbolic model checking using circuit cofactoring. In *International Conference on Computer-Aided Design (ICCAD'04)* (2004), 510–517.
27. Godefroid, P. Using partial orders to improve automatic verification methods. In *Computer-Aided Verification (CAV'90)*. Volume 531 of *Lecture Notes in Computer Science* (1990), 176–185.
28. Gößler, G., Sifakis, J. Composition for component-based modeling. *Sci. Comput. Progr. 55*, 1–3 (2005), 161–183.
29. Henzinger, T.A., Sifakis, J. The discipline of embedded systems design. *IEEE Comp. 40*, 10 (2007), 32–40.
30. Kautz, H.A., Selman, B. Planning as satisfiability. In *10th European Conference on Artificial Intelligence* (1992), 359–363.
31. Kozen, D. Results on the propositional mu-calculus. *Theor. Comput. Sci. 27* (Dec. 1983), 333–354.
32. Kurshan, R.P. *Computer-Aided Verification of Coordinating Processes*. Princeton University Press, 1994.
33. Lichtenstein, O., Pnueli, A. Checking that finite state concurrent programs satisfy their linear specification. In *POPL* (1985), 97–107.
34. Loiseaux, C., Graf, S., Sifakis, J., Bouajjani, A., Bensalem, S. Property preserving abstractions for the verification of concurrent systems.
35. McMillan, K.L. *Symbolic Model Checking: An Approach to the State Explosion Problem*. Kluwer Academic Publishers, 1993.
36. McMillan, K.L. Applying SAT methods in unbounded symbolic model checking. In *Computer-Aided Verification (CAV'02)*. Volume 2404 of *Lecture Notes in Computer Science* (2002), 250–264.
37. McMillan, K.L. Interpolation and SAT-based model checking. In *Computer-Aided Verification (CAV'03)*. Volume 2725 of *Lecture Notes in Computer Science* (2003), 1–13.
38. Peled, D. Combining partial order reductions with on-the-fly Model-Checking. In *Computer Aided Verification (CAV'94)*. Volume 818 of *Lecture Notes in Computer Science* (1994), 377–390.
39. Pnueli, A. The temporal logic of programs. Presented at FOCS, Oct. 1977.
40. Pnueli, A. Verification engineering: A future profession (A. M. Turing Award Lecture). Presented at *PODC* (Aug. 1997).
41. Queille, J.P., Sifakis, J. Specification and verification of concurrent systems in CESAR. In *Proceedings of the 5th International Symposium on Programming* (1982), 337–350.
42. Sheeran, M., Singh, S., Stålmarck, G. Checking safety properties using induction and a SAT-solver. In *Formal Methods in Computer-Aided Design (FMCAD'02)*. Volume 1954 of *Lecture Notes in Computer Science* (2000), 108–125.
43. Sistla, A.P., Gyuris, V., Emerson, E.A. SMC: a symmetry-based model checker for verification of safety and liveness properties. *ACM Trans. Softw. Eng. Methodol. 9*, 2 (2000), 133–166.
44. Tarski, A. A lattice-theoretical fixpoint theorem and its applications. *Pacific J. Math. 5* (1955), 285–309.
45. Valmari, A. A stubborn attack on the state explosion problem. In *Computer-Aided Verification (CAV'90)*. Volume 531 of *Lecture Notes in Computer Science* (1990).
46. Vardi, M.Y., Wolper, P. An automata-theoretic approach to automatic program verification (preliminary report). In *Proceedings, Symposium on Logic in Computer Science, 16–18 June 1986*, Cambridge, MA, USA, 1986, 332–344.
47. Wolper, P. Temporal logic can be more expressive. *Inform. Control 56* (1983), 72–99.

Edmund M. Clarke
FORE Systems University Professor
Computer Science Department
Carnegie Mellon University, Pittsburgh, PA.

E. Allen Emerson
Regents Chair and Professor
Department of Computer Science
University of Texas, Austin, TX.

Joseph Sifakis
Schneider-INRIA Chair
Verimag Laboratory Gieres, France.