Proceedings of SAT Competition 2025: Solver and Benchmark Descriptions, Proceedings of SAT Competitions, Vienna, 2025.

Challenging Certificates from Model Checking

Nils Froleyks
Johannes Kepler University
Linz, Austria

Emily Yu
Institute of Science and Technology Austria
Klosterneuburg, Austria

Armin Biere
University of Freiburg
Freiburg, Germany

The Hardware Model Checking Competition 2024 [1] introduced certificates to the bit-level track of the competition in the form of *witness circuits* [2]. Checking the correctness of a witness circuit entails solving a set of five SAT formulas. If all of them are unsatisfiable the witness circuit is valid, and the safety of the original model is proven.

The model checkers participating in the bit-level track of the competition have to prove that all states a sequential Boolean circuit can reach meet a given property. A circuit $M = (I, L, R, F, P, C)$ is modeled as a set of non-deterministic inputs I and a set of latches L which are initialized using their *reset function* in R and change value according to their transition function in F. Finally, C is a constraint that can be used to restrict the states to be considered, and P is the property to proof.

We use the following notation to denote that the latches in L are in a reset state $R\{L\} = \bigwedge_{\ell \in L} \ell{=}r_\ell(I, L)$, or follow the transition function $F_{0,1}\{L\} = \bigwedge_{\ell \in L} \ell_1{=}f_\ell(I_0, L_0)$. In the latter we use a lower index to refer to multiple temporal copies, i.e., multiple states along a trace.

If a circuits violates the property P, the model checkers have to return a sequence of inputs satisfying:

$$R_0\{L\} \wedge \bigwedge_{i \in [0,n)} F_{i,i+1}\{L\} \wedge \bigwedge_{i \in [0,n]} C_i \wedge \neg P_n.$$

If the property holds they have to return a witness circuit $W = (I', L', R', F', P', C')$ which satisfies:

Reset: $R\{K\} \wedge C \Rightarrow R'\{K\} \wedge C'$
Transition: $F_{0,1}\{K\} \wedge C_0 \wedge C_1 \wedge C'_0 \Rightarrow F'_{0,1}\{K\} \wedge C'_1$
Property: $(C \wedge C') \Rightarrow (P' \Rightarrow P)$
Base: $R'\{L'\} \wedge C' \Rightarrow P'$
Step: $P'_0 \wedge F'_{0,1}\{L'\} \wedge C'_0 \wedge C'_1 \Rightarrow P'_1$

Where $K = L \cap L'$ is the intersection between the circuits. Additionally, R' has to be stratified [3].

These checks are generated using CERTIFAIGER [4] and translated to CNF, thus yielding five formulas per model checking instance. Usually, all checks except for *Transition* and *Step* are trivial, as they do not encode the transition behavior, which usually is the most complex part.

Table list the 20 benchmarks submitted to the SAT competition. Since all of the witness circuits produced by RIC3 are correct, the submitted benchmarks only contain unsatisfiable benchmarks. Many of them could not be solved by MINISAT within a time limit of 10 hours. On the other hand, all 1492 valid certificates produced during the model checking competition where successfully checked within the same time limit using KISSAT 4.0. In fact, the overall certification overhead for the winning model checker RIC3 was only 34%.

Model	**rIC3**	**Check**	**miniSAT**
x-epic_a19-p15	0.64 s	Transition	$> 10h$
x-epic_a10-p53	1.28 s	Transition	438 s
cal182_cal182	3.22 s	Transition	$> 10h$
yosyshq_cv32e40x-p500	6.64 s	Transition	$> 10h$
yosyshq_cv32e40x-p749	6.74 s	Transition	$> 10h$
yosyshq_veer-p15	8.86 s	Transition	$> 10h$
yosyshq_axi-p23	9.88 s	Transition	34989 s
bv_ILA_Piccolo_JALR_sanity	13.14 s	Transition	26415 s
x-epic_a19-p16	13.89 s	Step	$> 10h$
x-epic_a19-p16	13.89 s	Transition	$> 10h$
bv_ILA_Piccolo_BEQ_sanity	17.20 s	Transition	27000 s
nla_freire1_valuebound1	17.81 s	Transition	$> 10h$
yosyshq_veer-p28	23.90 s	Transition	33821 s
bv_rocket_1951	32.62 s	Transition	766 s
yosyshq_axi-p06	35.16 s	Transition	33453 s
float-benchs_zonotope_2	37.98 s	Transition	394 s
2018D_VexRiscv-regch0-20-p1	150.54 s	Step	2788 s
nla_hard-ll_valuebound20	188.18 s	Transition	$> 10h$
nla_dijkstra-u_valuebound1	822.94 s	Step	$> 10h$
nla_dijkstra-u_valuebound1	822.94 s	Transition	$> 10h$

TABLE I
THE TABLE LIST BENCHMARKS SELECTED FROM THE HARDWARE MODEL CHECKING COMPETITION 2024 WITH THE TIME RIC3 TOOK TO PERFORM MODEL CHECKING, AND THE TIME MINISAT TOOK TO VALIDATE THE STATED CERTIFICATE VALIDATION CHECK.

The benchmark generator is available on Zenodo [5].

REFERENCES

[1] A. Biere, N. Froleyks, and M. Preiner, "Hardware model checking competition (hwmcc'20)," 2020. [Online]. Available: https://hwmcc.github.io/2020

[2] N. Froleyks, E. Yu, M. Preiner, A. Biere, and K. Heljanko, "Introducing certificates to the hardware model checking competition," in *Computer Aided Verification (CAV)*, 2025.

[3] E. Yu, N. Froleyks, A. Biere, and K. Heljanko, "Stratified certification for k-Induction," in *22nd Formal Methods in Computer-Aided Design, FMCAD 2022, Trento, Italy, October 17-21, 2022*, A. Griggio and N. Rungta, Eds. IEEE, 2022, pp. 59–64.

[4] N. Froleyks, E. Yu, and A. Biere, "Certifaiger on github," 2025. [Online]. Available: https://github.com/Froleyks/certifaiger

[5] N. Froleyks, E. Yu, and A. Biere, "Challenging certificates from model checking," Apr. 2025. [Online]. Available: https://doi.org/10.5281/zenodo.15267003