Hardware Model Checking Competition 2013

HWMCC'13

Armin Biere, Keijo Heljanko, Martina Seidl, Siert Wieringa

presented at

Formal Methods in Computer Aided Design 2013

FMCAD'13

Portland, Oregon, USA

Tuesday

October 22, 2013

including minor updates from October 27, 2013

solver	sat+uns		sat		uns		to	mo	uk	real sec	time sec	sum mb	max mb
iimc	G	140	G	78	S	62	45	0	0	4954	11634	58232	5136
simplive	S	122	S	68	B	54	43	20	0	7933	23483	30941	2990
tiprbmc12		122		58		64	63	0	0	8818	8751	3295	186
tiprbmc	B	120		58		62	65	0	0	7153	7083	3170	187
v3		108		63		45	75	2	0	10335	28843	26110	6716
tip		79		15	G	64	106	0	0	4307	4261	910	110
tipbmc		64	B	64		0	95	21	5	1681	1633	1119	220
aigbmclgl*		62		62		0	97	26	0	3458	3425	4925	3223
aigbmc*		56		56		0	96	33	0	2744	2715	1105	300

G = gold (winner), S = silver (2nd), B = bronze (3rd)

to = time out, mo = memory out, real = wall clock time
time = process time, sum = sum of max-mem, max = max mem
(restricted to solved benchmarks, e.g. for real, time, sum, max)
uk = unknown (tipmbmc reached 100k bound limit)
* hors concours

	properties solved			score based		
solver	sat+uns	sat	uns	sat+uns	sat	uns
mulprove	97397	30379	67018	G 60.67%	S 21.42%	G 39.26%
mpmc12+	90520	31039	59481	55.54%	21.17%	34.37%
tiprbmc	76661	31913	44748	S 50.68%	20.71%	29.96%
tip	75166	30302	44864	48.67%	18.17%	S 30.50%
v3	75849	31553	44296	B 44.66%	B 20.67%	B 23.99%
tipbmc	32273	32273	0	21.90%	G 21.90%	0.00%
aigbmclgl*	29693	29693	0	18.07%	18.07%	0.00%
aigbmc*	4063	4063	0	17.60%	17.60%	0.00%

+ mpmc12 winner of last year but has now discrepancies:
6s249-26 6s386-11 6s386-1 6s386-12 bob12m15m-0

* hors concours

$$\text{score} = \frac{1}{178} \cdot \sum_{i=1}^{178} \frac{\#\text{solved}_i}{\#\text{properties}_i}$$

solver	sat+uns		sat		uns		to	mo	s11	uk	real	time	space	max
suprove	G	138		48	G	90	100	10	0	0	14466	32981	145583	6636
suprove12		133		48		85	101	14	0	0	13981	29279	113454	4740
simple		132	G	51		81	107	9	0	0	11009	34802	138233	6636
iimc	S	113	S	38	S	75	134	1	0	0	15376	52115	55186	5564
v3	B	105	B	36	B	69	137	6	0	0	10633	41378	100085	5450
nuxmvexp		94		26		68	144	10	0	0	14136	55388	103733	6477
nuxmv		94		27		67	142	12	0	0	15212	59526	113280	6666
simbip		83		26		57	106	14	0	45	1104	2914	74992	4986
tip		78		23		55	170	0	0	0	8540	8499	3749	674
pdtrav		74		20		54	143	1	0	30	8962	27637	82892	3490
minireachic3tp		67		30		37	181	0	0	0	8031	15970	26565	5859
minireachic3		63		29		34	185	0	0	0	7945	15804	22887	5945
simpsat		43		22		21	105	11	0	89	12492	40406	39769	6613
smspdr		42		12		30	164	31	0	11	9903	9864	26731	4936
blimcn*		39		36		3	201	8	0	0	4965	4944	10767	2520
blimc*		37		34		3	202	9	0	0	3354	3333	9424	1327
tipbmc		35		35		0	171	20	0	22	3420	3400	8581	2374
aigbmclgl*		31		31		0	192	25	0	0	3408	3371	14296	2406
shiftbmcn		30		29		1	197	21	0	0	3073	3055	12378	2721
shiftbmc		29		28		1	196	23	0	0	2868	2851	12001	2530
fussiasto		29		12		17	200	1	11	7	2953	2936	7303	1396
aigbmc*		26		26		0	182	40	0	0	3170	3153	12015	2323
shiftbmcna		25		25		0	183	12	0	28	4518	4496	3976	752

* hors concours

solver	bounds	insts	capped	score
tipbmc	3101	64	14	94.53%
blimc*	3022	64	12	94.22%
blimcn*	3010	64	12	94.14%
aigbmclgl*	2757	64	9	93.94%
shiftbmc	2750	64	10	93.71%
shiftbmcn	2745	64	10	93.70%
aigbmc*	2636	64	8	93.51%
nuxmvexp	2412	64	6	93.48%
nuxmv	2411	64	6	93.45%
v3	2534	64	6	92.76%
shiftbmcna	2807	59	11	86.60%
pdtrav	2662	58	10	85.23%
tip	525	64	0	84.43%
smspdr	1748	56	5	79.61%
suprove12	1714	49	8	69.56%
suprove	1615	42	5	61.71%

bounds = sum of bounds reached (capped at bound 100)
* hors concours

Motivation

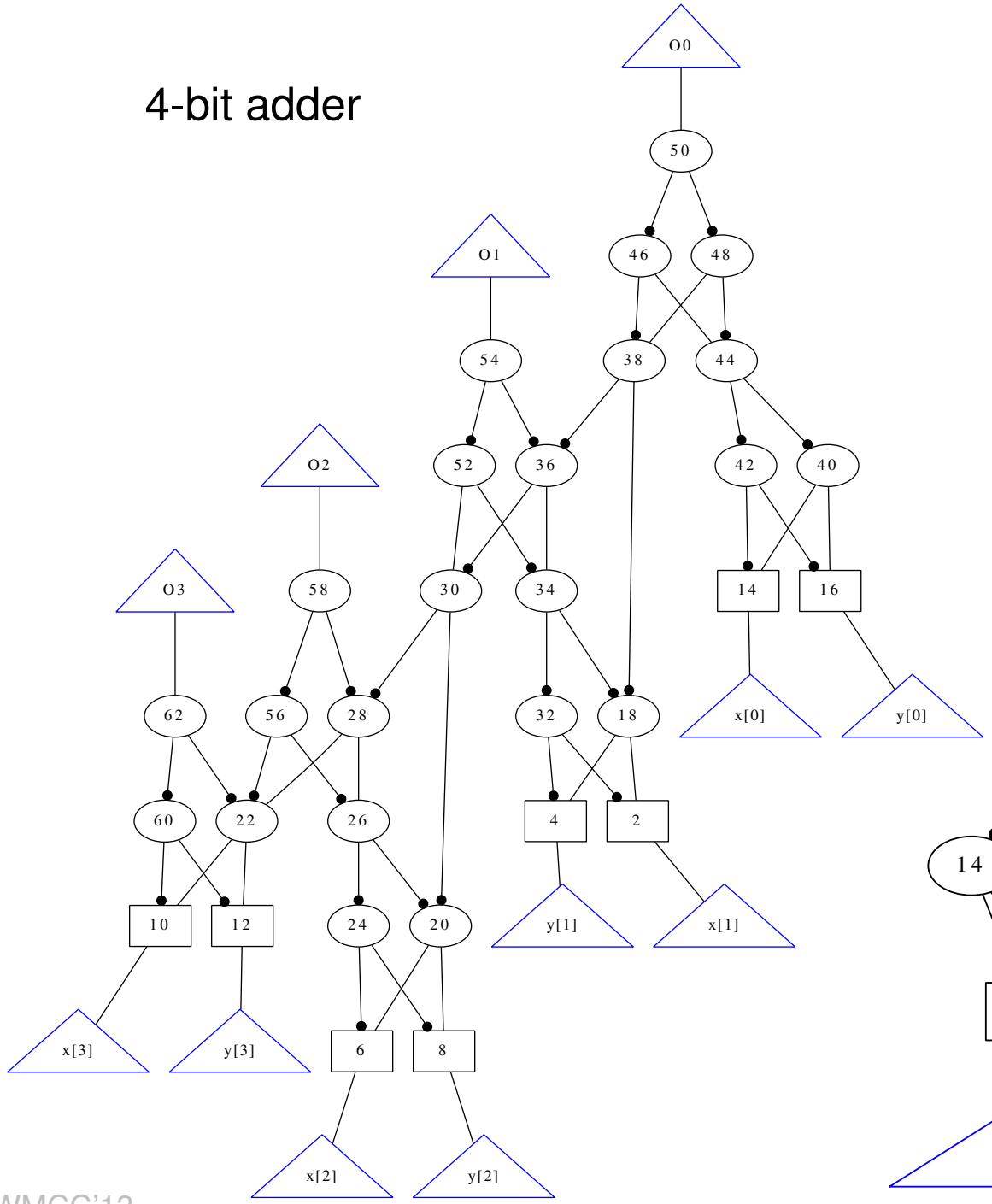
- intensify interest in improving symbolic model checking technology
  - symbolic model checking does not scale *enough* in practice
  - recent new research results, like IC3
  - provide more diverse benchmarks
- repeat success story of SAT/SMT competitions
  - simple standardized input format $\Rightarrow$ AIGER
  - motivation for young researchers to enter this field
  - provide “standard set” of benchmarks
- needs active support by submitters of benchmarks and model checkers

History

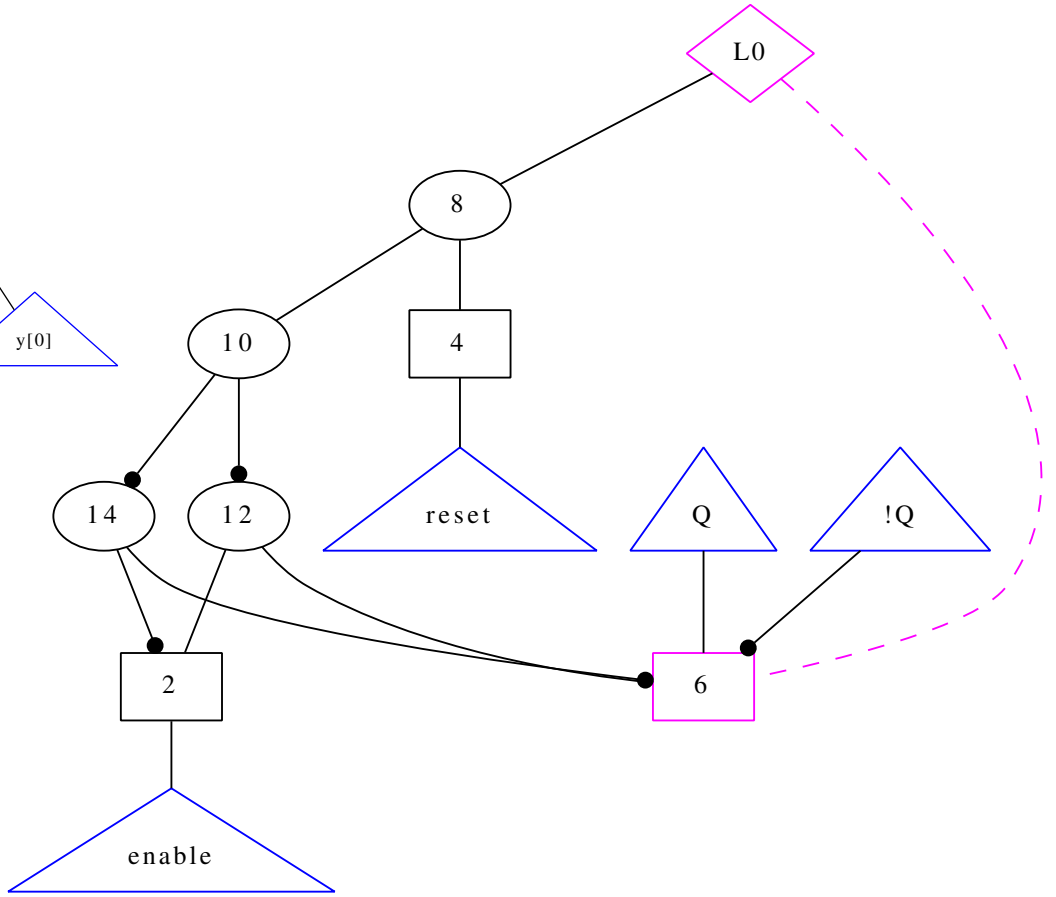
2006	2007	2008	2010	2011	2012	**2013**
AIGER format AVM'06 Ascona	1st HWMCC CAV'07 Berlin	2nd HWMCC CAV'08 Princeton	3rd HWMCC	4th HWMCC	5th HWMCC	**6th** **HWMCC**
Founding Lunch CAV'06 FLOC'06 Seattle		HWMCC Lunch FMCAD'08 Portland	CAV'10 FLOC'10 Edinburgh	FMCAD'11 Austin	FMCAD'12 Cambridge UK	**FMCAD'13** **Portland**

- founding lunch at CAV'06, first competition at CAV'07
- HWMCC lunch at FMCAD'08 $\Rightarrow$ should have benchmarks with multiple properties !!!
- HWMCC'10 at CAV'10 (FLOC'10), since HWMCC'11 with FMCAD
- HWMCC'11: old *single* property track, new *live*ness and new *multi* property track
- HWMCC'12: as HWMCC'11 except for deep bounds track sponsored by Oski
- HWMCC'13: identical to HWMCC'11-12 to **stabilize format and rules**

4-bit adder



toggle flip-flop with reset & enable



- **A**nd-**I**nverter-**G**raph (AIG) file format `http://fmv.jku.at/aiger`
  - structural / circuit SAT and model checking problems
  - compact and (rather) easy to parse
- version 1.9 introduced in 2011 with **new** sections / properties **M I L O A B C J F**
  - **M**aximum variable / index, **I**nputs, **L**atches, **O**utputs, **A**nds
  - **B**ad state properties, (environment) invariance **C**onstraints / assumptions
  - **J**ustice / liveness properties, **F**airness constraints
- new output since HWMCC'12: reached bound **u**⟨**k**⟩ for deep bound track
  only supported by some model checkers
- new binary format 2.0 is still work in progress

Single **Live**'ness (Justice) Property Benchmarks

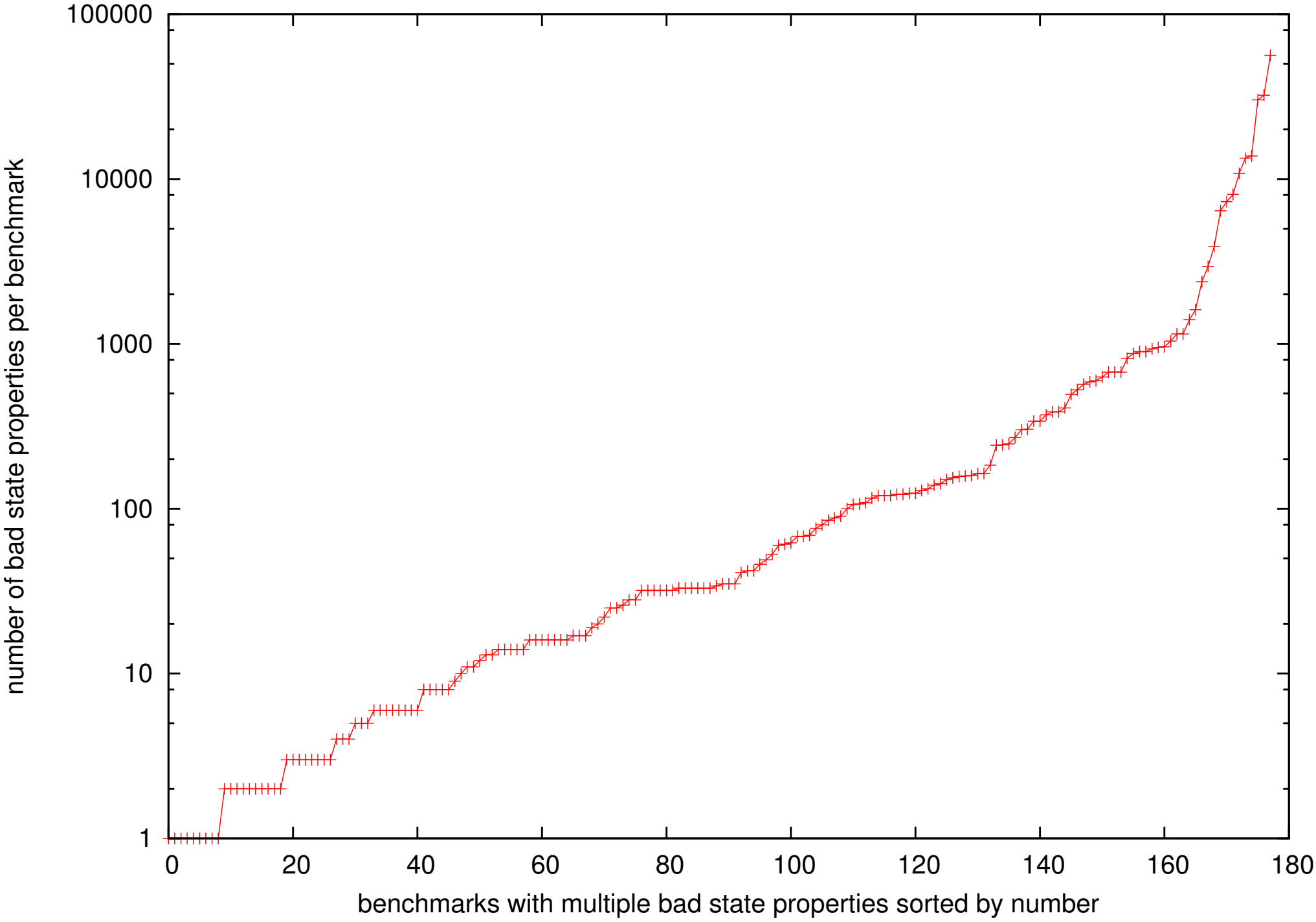
- one of the main goals of the competition is to collect benchmarks
  - last year no new liveness benchmarks
  - used all **118** benchmarks from last two competitions
- **18** single liveness benchmarks submitted by Fabio Somenzi from 5 models
- selected **49** single liveness problems from new 6s suite from all 23 suitable models
  - extracted from 266 new "6s" benchmarks from Jason Baumgartner
  - picked *at most* three justice properties per model randomly
  - required to reset latches to zero
- in summary **185** single liveness benchmarks

Multiple Bad State Properties

- all **76** benchmarks from last year — actually 17 "moved" with `aigmove`
- **one** new multi property benchmark from Vigyan Singhal (Oski)
- **101** new "6s" benchmarks
  - extracted from 266 new "6s" benchmarks from Jason Baumgartner
  - since rules forced AIGER 1.9 format, there was no need to "reset"
- in summary **178** multiple bad state property benchmarks
- in AIGER 1.9 format, thus ...
  - with environment constraints
  - and potentially 1-initialized or no unitialized latches

Largest 15 **Multi**ple Bad State Benchmarks

	M	I	L	O	A	B	C
aig	567561	1048	23957	0	542556	8064	0
aig	739268	3641	102390	0	633237	107	0
aig	740920	1586	80927	0	658407	896	0
aig	774146	1237	42846	0	730063	244	10
aig	768941	5465	23056	0	740420	69	3
aig	1001099	1381	101639	0	898079	32210	0
aig	1405629	83041	83717	0	1238871	157	0
aig	1567435	1760	84925	0	1480750	124	21
aig	1567435	1760	84925	0	1480750	124	21
aig	1804823	2281	243399	0	1559143	60	0
aig	1747497	14458	69115	0	1663924	1	4
aig	2447918	91099	177235	0	2179584	90	0
aig	2661180	3468	186401	0	2471311	56211	0
aig	3544465	4708	260713	0	3279044	1041	0
aig	5623524	252041	467369	0	4904114	960	0


100000
10000
1000
100
10
1
number of bad state properties per benchmark
0
20
40
60
80
100
120
140
160
180
benchmarks with multiple bad state properties sorted by number

Single Safety (Bad State) Property Benchmarks

- all **11** single bad state properties from one Oski model
- selected **123** instances from HWMCC'12
  - based on performance of 18 HWMCC'12 solvers in last competition
  - selected 0 from 52 *trivial* at least 8 UNSAT, 16 SAT
  - selected 8 from 40 *easy* at least 6 UNSAT, 12 SAT
  - selected 23 from 49 *medium* at least 3 UNSAT, 6 SAT
  - selected 47 from 75 *hard* at most 2 UNSAT, 5 SAT
  - selected 48 from 94 *unsolved*
  - which gives 126, but then removed 3 combinatorial benchmarks
- plus **114** from 266 new "6s" benchmarks from Jason Baumgartner
- results in **248** single bad state property benchmarks used this year

  all in *pre* 1.9 AIGER format

Model Checkers

- **aigbmc**, **aigbmclgl**, **blimc**, **blimcn** by Biere (Linz) — new versions
- **pdtrav** by Cabodi, Nocco, Quer (Torino) — new version
- **suprove**, **simpsat**, **simbip**, **simple**, **simplive**, **mulprove** (Berkeley)
  Brayton, Eén, Mishchenko, Sterin — new tools and new versions
- **tip**, **tipbmc**, **tiprbmc** by Sörensson, Claessen (Göteborg) — essentially same as last year
- **v3** by Cheng-Yin Wu, Chi-An Wu, and Chung-Yang (Ric) Huang (Taiwan) — new version
- **fussiasto** by Chien-Yu (Leo) Lai (Taiwan) — **new**
- **minireachic3**, **minireachic3tp** by Martin Suda (Saarbrücken) — **new**
- **smspdr** by Sam Bayless (Vancouver) — **new**
- **shiftbmc** by Norbert Manthey (Dresden) — **new**
- **iimc** by Hassan, Somenzi, Dooley, Bradley (Colorado) — new version
- **nuxmv**, **nuxmvexp** by Alberto Griggio (Trento) — **new**

aigbmc + blimc

aigbmc

- as in HWMCC'11-12 but with new PicoSAT version
- bounded model checker based on FMCAD'04 / CAV'05 papers by Heljanko et.al.
- for multiple properties now goes on until all are reached
- **aigbmclgl** with our SAT solver Lingeling (and cloning `lglclone`)

blimc

- bounded model checker for safety (bad state) properties only
- show- and testcase for the incremental features of our SAT solver Lingeling
- simplifies transition relation with SAT based preprocessing
- uses latest Lingeling and cloning for hard bounds `lglclone`
- **blimcn** does not provide witnesses (eliminates inputs)

After sending an ABC discrepancy to the Berkeley team


```
from:     Robert Brayton <brayton@berkeley.edu> via jku.at
to:       Armin Biere <biere@jku.at>
date:     Tue, Oct 1, 2013 at 5:27 AM
subject:  Re: hwmcc13 entries from berkeley

I got the example and anxiously waiting to debug it, but
there was an explosion at Berkeley yesterday and as far as
I can tell the power is still out (I am in Vermont). So our
servers are not working and without them I can't run anything
from here. I am waiting for people to wake up in Berkeley to
tell me the situation. Fortunately e-mail is still working.

Bob
```

berkeley explosion - Google Search - Mozilla Firefox

berkeley explosion - Google Se...

https://www.google.com/search?client=ubuntu&chanr

Google

+Armin Search Images Maps Play YouTube News Gmail Drive Calendar More

Google

berkeley explosion

Web Images Maps Shopping News More Search tools

About 27,800 results (0.26 seconds)

Wall Street J...

Power Restored at UC **Berkeley** After **Explosion**

ABC News - 2 hours ago

Power was restored to most of the University of California, **Berkeley**, Tuesday morning, after an **explosion** left one person hospitalized for minor **...**

UC **Berkeley Explosion** Knocks Out Power, Causes Multiple Injuries **...**

Huffington Post - 2 hours ago

UC **Berkeley explosion**: Students recall people fleeing, screaming

Los Angeles Times - 9 hours ago

YouTube

YouTube

abc7news.com

NBC Bay Area

Daily Mail

San Francisc...

all 172 news sources »

Submissions from Berkeley

- Single-output - *suprove*
  - Improved simplification
  - Improved reparametrization
  - Improved gate-level abstraction
  - Rarity simulation
- Multiple output - *mulprove*
  - Extensions to rarity simulation, bmc, pdr/ic3 to continue solving with per/output timeouts
    - These run in parallel and some are pushed to start at far reachable states.
  - Use of isomorphism

simplive ABC's liveness model checker

Run in parallel:

- Liveness-to-safety, then abstraction/PDR
- Liveness-to-safety, then BMC
- k-liveness, PDR
- A combinational simplification engine

If any engine returns a result (SAT or UNSAT) the other engines are killed.

If the simplification engine terminates before a result is returned, it kills all engines except the first and run the following in parallel:

- Liveness-to-Safety conversion, then PDR
- Liveness-to-Safety then BMC
- k-liveness and PDR

(Unfortunately, there was a bug in launching the simplification engined, so the actual code just runs the first three engines in parallel)

- Features
  - Multiple engine (threaded) tool
  - Heuristically driven manager (expert system)
  - Initial transformations/reductions (combinational+sequential)
  - Includes: Cudd, Minisat, ABC (combinational synthesis):
    MC Engines: BMC, BDDs, k-induction, IC3, ITP, IGR.

- Just single property track:
  - portfolio-based, static + light weight dynamic classification & engine selection running 5 (threaded) engines concurrently.

- Improvements over 2012
  - new engine:
    IGR, interpolation with guided refinement improvements in BMC, ITP, IC3

- todos
  - improve expert system
  - integrate simple sequential transformations using ABC
  - Salability for multiple property track

tip in single track

The default mode of **tip** is indeed a reasonable choice for single property runs. Since **tip** doesn't have a good approach to time-slice between different engines this just runs my IC3 implementation after first doing temporal decomposition.

---

tip in live track

Yes the default mode will run our k-liveness as published in last years FMCAD. The parameter sets the medium level of "fairness constraints extraction" also described in our paper.

---

tipbmc BMC checking with temporal decomposition and equivalent latch extraction.

---

tiprbmc

interleaves BMC with IC3 in a very stupid (not time-based) way. Since k-liveness can't find counter-examples in our current implementation this is necessary for completeness.

V3 for HWMCC13

- Authors
  - Cheng-Yin Wu, Chi-An Wu, and Chung-Yang (Ric) Huang
- Affiliations
  - Design Verification Lab from National Taiwan University
- Tracks we entered
  - single, liveness, multi
- Safety engines
  - UMC (BMC + induction)
  - Interpolation (NewITP)
  - PDR (two versions)
  - SEC (many versions)
- Liveness engines
  - Liveness-to-safety (L2S) for all safety engines
  - K-liveness (enables different safety engines)

V3 for HWMCC13

- Global configuration
  - Engines share deep bounds, invariants and SEC nets
  - Four threads run different engines in parallel

- Single track configuration
  - Also exploit *dprove* in ABC for proving properties
- Liveness track configuration
  - Two cores for L2S, two cores for K-liveness
- Multi track configuration
  - Interleaving properties among threads after removing trivial ones
  - Verify properties orderly under increasingly relaxed resources
  - Update the order of properties according to resource used
  - NO verification result confirmation for multi track

Technology

- a simple implementation of IC3/PDR with Minisat 2.2 as the backend SAT-solver
- it runs ”forward” and ”backward” versions of the algorithm in parallel
- while **minireachic3** uses standard clause propagation, **minireachic3tp** features a new technique: *Triggered Clause Pushing*
  - the idea is to collect models from unsuccessful clause pushes
  - there is no need to retry a push attempt unless such a witness model falsifies a newly learned clause – with this mechanism we are able to incorporate pushing into the blocking phase of the algorithm and keep all the clauses pushed as far as possible at all times
- for more details see `http://www.mpi-inf.mpg.de/~suda/triggered.html`

Expected outcome

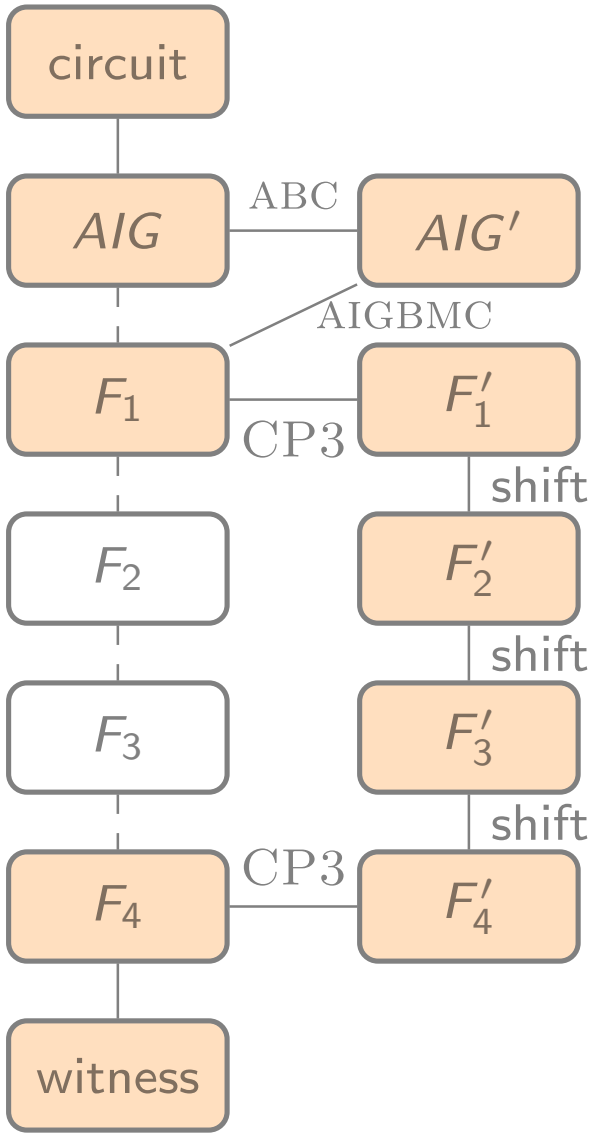
- **minireachic3tp** is expected to beat **minireachic3**
  and thus demonstrate the usefulness of Triggered Clause Pushing

- new implementation of IC3/PDR
- introduces the new concept of *SAT modulo SAT*
  - efficiently answer the kinds of queries IC3 makes
- improves IC3's *clause propagation*
  - so it can avoid quadratic overhead in the number of time frames
- competition submission is single threaded
- running without any pre-processing
  so, its probably not going to win!
- see FMCAD'13 paper

SHIFT-BMC

- Reading and Encoding AIGER to SAT: AIGBMC
- Simplifying Circuit: ABC Library
  - Run commands *dc2* and *scorr* in this order
  - Extracts the number of latches before simplification
  - Currently supports only single bad-property
- Simplifying CNF: Coprocessor 3 (CP3)
  - Uses: TernRes, Unhiding, ELS
    BVE, BVA, Variable densing
- SAT solver: PicoSAT 957

- Unrolling Circuit:
  - Copy the simplified CNF and shift the variables
  - Add equivalences for latch inputs/outputs
- Versions:
  - SHIFT: basic CNF simplification
  - SHIFT–NF: more complex CNF simplification
  - SHIFT–NF–ABC: run circuit simplification
- Author: Norbert Manthey, TU Dresden



Thanks to Armin Biere and Alan Mishchenko for AIGBMC and ABC support

- **IImc** is a model checker for AIGER 1.9 models
- It checks invariants, language emptiness, and CTL properties
- It is written in C++11
- **IImc**'s philosophy is Incremental, Inductive Verification
  - IIV engines: IC3, Fair, IICTL
  - It has a few other engines, both BMC- and BDD-based
- It uses CUDD, zchaff, and minisat
- Version 1.2 is available from `http://iimc.colorado.edu`

- Improvements to IC3: lifting, CTGs
- IC3 engine with localization reduction
- Minisat used as SAT solver for everything but IC3
- Improved preprocessing
  - Ternary simulation
  - Phase abstraction
  - Extraction of unit literal invariants
  - Transition relation reversal for ”backward beems”
  - Removed inefficiencies
- Multithreading based on C++11
  - Four threads in default setup: IC3, reverse IC3, BMC, and either BDDfw or IC3 LR

- FCBMC engine (Fair Cycle BMC)
- GSH engine (BDD-based language-emptiness check)
- Multithreaded (Fair plus two previous engines)

new bit-level engine for nuXmv (eXtended, neXt generation NuSMV)

- only single safety properties for now

for HWMCC'13, portfolio approach with 4 independent engines working in parallel:

- 3 variants of IC3
  1. the "default" one (combines ideas from several implementations)
  2. using CTGs for clause generalization (Hassan et al. @this FMCAD)
  3. using abstraction refinement (Vizel et al. FMCAD'12)
     but with BMC-based refinement instead of IC3
- 1 BMC engine
- k-induction for small bounds (until 45), then plain BMC

SAT solver-independent architecture:

- MiniSAT (version $\geq$ 2.2) and PicoSAT currently supported
- MiniSAT used in the competition

Setup

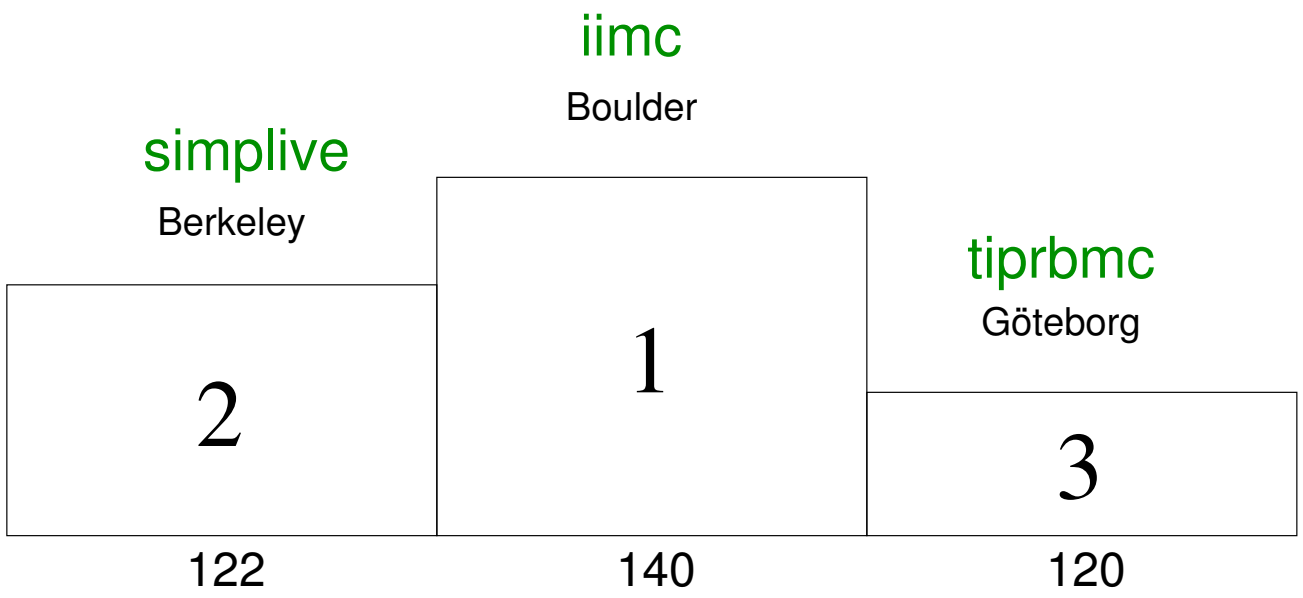
- single property benchmarks (single + live tracks) as in HWMCC'07 - HWMCC'12
  - bad state resp. fair SCC *reachable* $\Rightarrow$ instance *satisfiable* SAT
  - bad state resp. fair SCC *unreachable* $\Rightarrow$ instance *unsatisfiable* UNSAT
- multiple properties per benchmarks (multi track)
  - count the number of solved individual properties
- all solvers read AIGER natively but not all produce full witnesses
- 900 seconds *wall clock* time limit, 7 GB memory limit
  - 32 node cluster, Intel Quad Core 2.6 GHz processors, 8 GB, Ubuntu
  - each solver has full access to one node (4 cores, no hyperthreading)

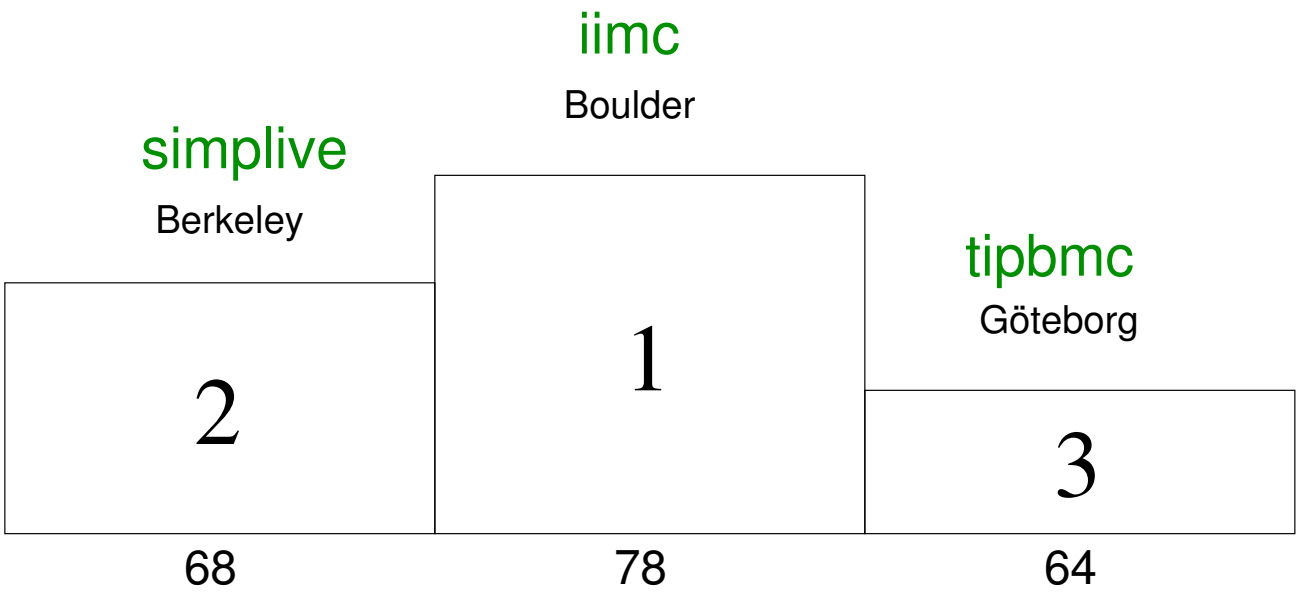
Rankings

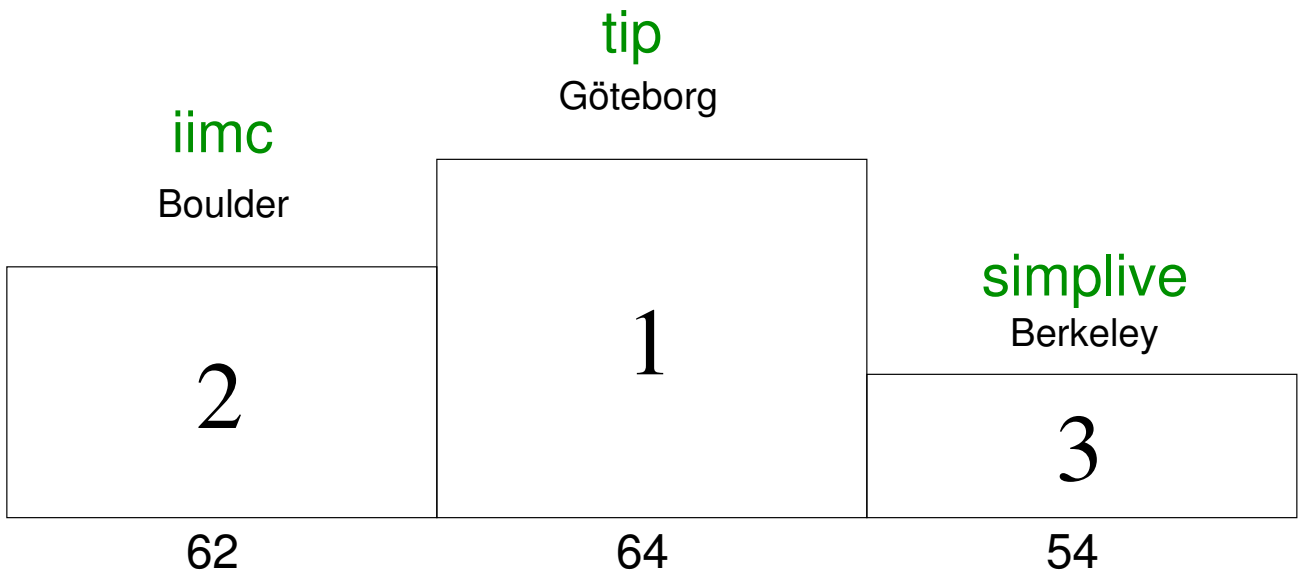
- Main tracks live, multi, single
  - three categories: SAT+UNSAT, SAT, UNSAT
  - no additional single threaded versus multi-threaded ranking

multi threaded ranking	=	wall clock time *limit*	used for ranking
single threaded ranking	=	process time *limit*	not used

- each *group* is only awarded one *virtual medal* per ranking
  - detailed results will be provided for all solvers `http://fmv.jku.at/hwmcc13`
  - you will also get spread sheets and all the log files there


iimc
Boulder
simplive
Berkeley
tiprbmc
Göteborg
2
1
3
122
140
120

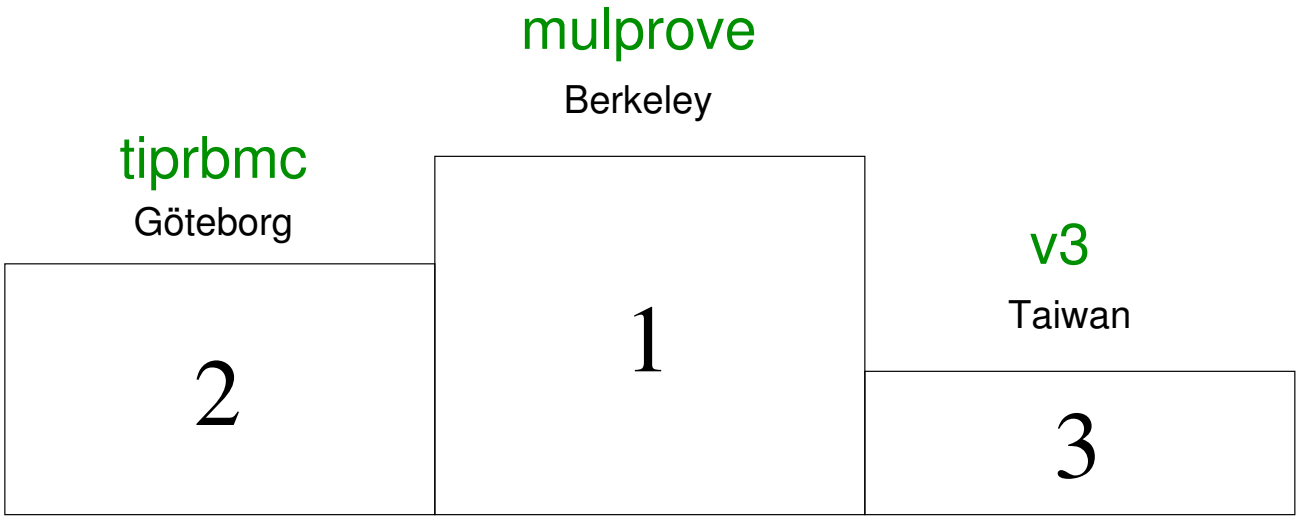

iimc
Boulder
simplive
Berkeley
tipbmc
Göteborg
2
1
3
68
78
64


tip
Göteborg
iimc
Boulder
simplive
Berkeley
2
1
3
62
64
54

solver	sat+uns		sat		uns		to	mo	uk	real sec	time sec	sum mb	max mb
iimc	G	140	G	78	S	62	45	0	0	4954	11634	58232	5136
simplive	S	122	S	68	B	54	43	20	0	7933	23483	30941	2990
tiprbmc12		122		58		64	63	0	0	8818	8751	3295	186
tiprbmc	B	120		58		62	65	0	0	7153	7083	3170	187
v3		108		63		45	75	2	0	10335	28843	26110	6716
tip		79		15	G	64	106	0	0	4307	4261	910	110
tipbmc		64	B	64		0	95	21	5	1681	1633	1119	220
aigbmclgl*		62		62		0	97	26	0	3458	3425	4925	3223
aigbmc*		56		56		0	96	33	0	2744	2715	1105	300

G = gold (winner), S = silver (2nd), B = bronze (3rd)

to = time out, mo = memory out, real = wall clock time
time = process time, sum = sum of max-mem, max = max mem
(restricted to solved benchmarks, e.g. for real, time, sum, max)
uk = unknown (tipmbmc reached 100k bound limit)
* hors concours

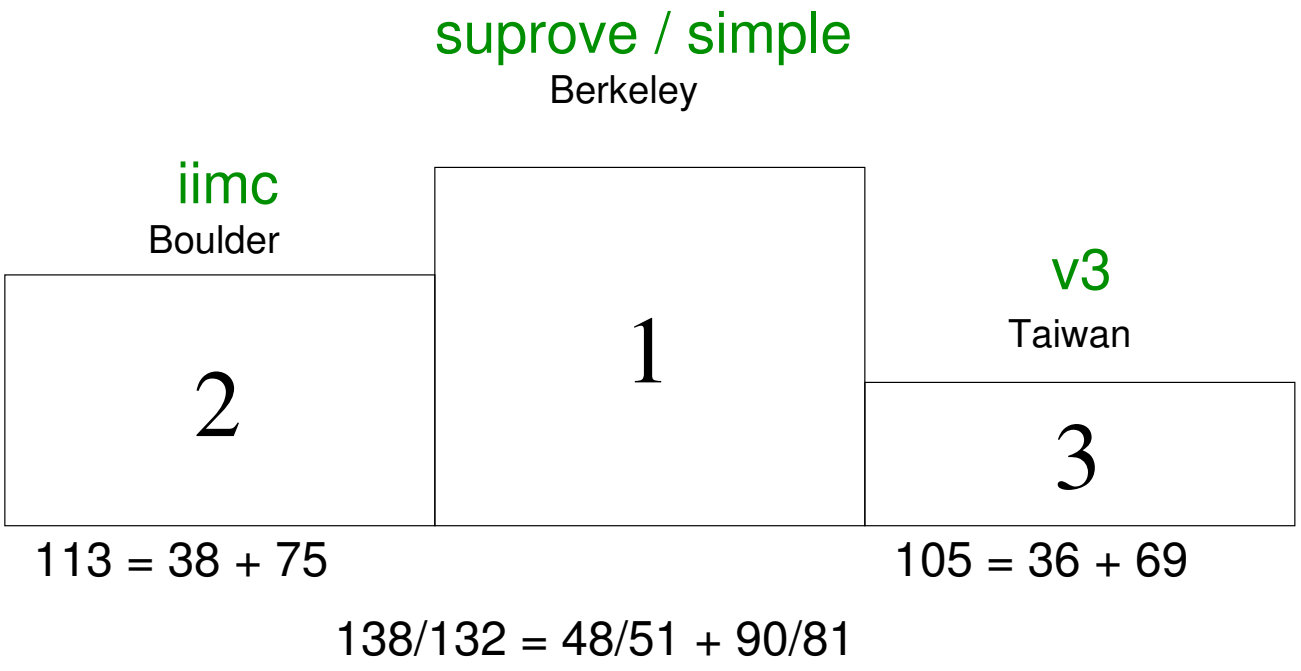

mulprove
Berkeley
tiprbmc
Göteborg
v3
Taiwan
2
1
3

	properties solved			score based				
solver	sat+uns	sat	uns	sat+uns		sat		uns
mulprove	97397	30379	67018	G	60.67%	S	21.42%	G 39.26%
mpmc12+	90520	31039	59481		55.54%		21.17%	34.37%
tiprbmc	76661	31913	44748	S	50.68%		20.71%	29.96%
tip	75166	30302	44864		48.67%		18.17%	S 30.50%
v3	75849	31553	44296	B	44.66%	B	20.67%	B 23.99%
tipbmc	32273	32273	0		21.90%	G	21.90%	0.00%
aigbmclgl*	29693	29693	0		18.07%		18.07%	0.00%
aigbmc*	4063	4063	0		17.60%		17.60%	0.00%

+ mpmc12 winner of last year but has now discrepancies:
  6s249-26 6s386-11 6s386-1 6s386-12 bob12m15m-0

* hors concours

$$\text{score} = \frac{1}{178} \cdot \sum_{i=1}^{178} \frac{\#\text{solved}_i}{\#\text{properties}_i}$$


suprove / simple
Berkeley
iimc
Boulder
v3
Taiwan
2
1
3
113 = 38 + 75
105 = 36 + 69
138/132 = 48/51 + 90/81

solver	sat+uns	sat	uns	to	mo	s11	uk	real	time	space	max
suprove	G 138	48	G 90	100	10	0	0	14466	32981	145583	6636
suprove12	133	48	85	101	14	0	0	13981	29279	113454	4740
simple	132	G 51	81	107	9	0	0	11009	34802	138233	6636
iimc	S 113	S 38	S 75	134	1	0	0	15376	52115	55186	5564
v3	B 105	B 36	B 69	137	6	0	0	10633	41378	100085	5450
nuxmvexp	94	26	68	144	10	0	0	14136	55388	103733	6477
nuxmv	94	27	67	142	12	0	0	15212	59526	113280	6666
simbip	83	26	57	106	14	0	45	1104	2914	74992	4986
tip	78	23	55	170	0	0	0	8540	8499	3749	674
pdtrav	74	20	54	143	1	0	30	8962	27637	82892	3490
minireachic3tp	67	30	37	181	0	0	0	8031	15970	26565	5859
minireachic3	63	29	34	185	0	0	0	7945	15804	22887	5945
simpsat	43	22	21	105	11	0	89	12492	40406	39769	6613
smspdr	42	12	30	164	31	0	11	9903	9864	26731	4936
blimcn*	39	36	3	201	8	0	0	4965	4944	10767	2520
blimc*	37	34	3	202	9	0	0	3354	3333	9424	1327
tipbmc	35	35	0	171	20	0	22	3420	3400	8581	2374
aigbmclgl*	31	31	0	192	25	0	0	3408	3371	14296	2406
shiftbmcn	30	29	1	197	21	0	0	3073	3055	12378	2721
shiftbmc	29	28	1	196	23	0	0	2868	2851	12001	2530
fussiasto	29	12	17	200	1	11	7	2953	2936	7303	1396
aigbmc*	26	26	0	182	40	0	0	3170	3153	12015	2323
shiftbmcna	25	25	0	183	12	0	28	4518	4496	3976	752

* hors concours

Deep Bounds Track

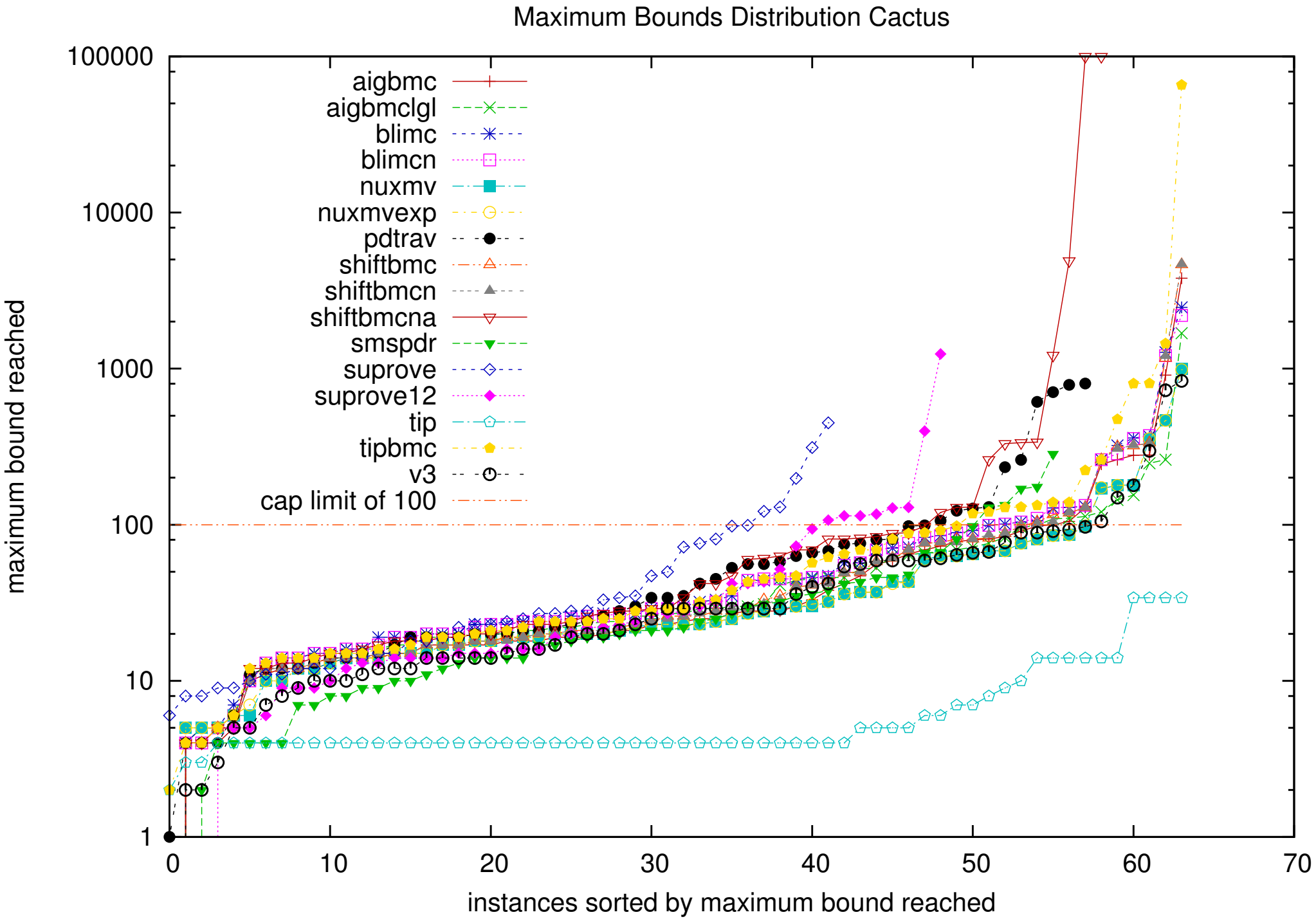
- for some industrial applications deep bounds capacity is important:
  - often actual model checking problems can not be proven fast
  - good metric to measure progress is how deep model checker proved unsatisfiability
- model checkers asked to print bounds
  - e.g. `u10` means bad state not reachable within 10 steps
  - turns model checking into a kind of optimization problem
  - so the deep bounds track is similar to MAXSAT competitions
- **only run on the unsolved instances of the single track**
- $500 award sponsored by Oski Technology

Deep Bounds Track Instances

65 unsolved instances in single track:

6s105 6s119 6s13 6s148 6s160 6s161 6s171 6s188 6s195 6s22 6s23 6s24 6s266rb2 6s267rb3 6s268r 6s274r 6s279r 6s280r 6s29 6s329rb19 6s33 6s340rb27 6s341r 6s343b31 6s351rb02 6s35 6s365r 6s366r 6s367r 6s376r 6s377r 6s382r 6s387rb181 6s392r 6s398b09 6s399b03 6s39 6s402rb0342 6s416r 6s44 6s45 6s7 beemextnc1b1 beemkrebs4b1 beemloyd3b1 beemskbn2b1 bob12s06 bobpcihm bobsmcodic bobsmminiuart intel012 intel013 intel016 intel027 intel032 intel048 intel065 intel066 intel067 oski1rub00 oski1rub01 oski1rub02 oski1rub08 oski1rub09 oski1rub10

no solver proved u0 for 6s398b09


Maximum Bounds Distribution Cactus
maximum bound reached
instances sorted by maximum bound reached
100000
10000
1000
100
10
1
0
10
20
30
40
50
60
70
aigbmc
aigbmclgl
blimc
blimcn
nuxmv
nuxmvexp
pdtrav
shiftbmc
shiftbmcn
shiftbmcna
smspdr
suprove
suprove12
tip
tipbmc
v3
cap limit of 100

Principles

emphasize robustness in reaching deep bounds for many benchmarks

independent of the solvers running on the selected (65) benchmarks

$$\texttt{score} = \frac{1}{65} \cdot \sum_{i=1}^{65} \left(1 - \frac{1}{2 + \texttt{maxbound}_i}\right)$$

`u0`	gives	1/2	=	50.00%
`u1`	gives	2/3	=	66.66%
`u2`	gives	3/4	=	75.00%
`u10`	gives	11/12	=	91.66%
`u100`	gives	101/102	=	99.01%
`u1000`	gives	1001/1002	=	99.90%
`u10000`	gives	10001/10002	=	99.99%

sponsored by

Oski Technology

presented by

Vigyan Singhal

Deep Bound Track Results

solver	bounds	insts	capped	score
tipbmc	3101	64	14	94.53%
blimc*	3022	64	12	94.22%
blimcn*	3010	64	12	94.14%
aigbmclgl*	2757	64	9	93.94%
shiftbmc	2750	64	10	93.71%
shiftbmcn	2745	64	10	93.70%
aigbmc*	2636	64	8	93.51%
nuxmvexp	2412	64	6	93.48%
nuxmv	2411	64	6	93.45%
v3	2534	64	6	92.76%
shiftbmcna	2807	59	11	86.60%
pdtrav	2662	58	10	85.23%
tip	525	64	0	84.43%
smspdr	1748	56	5	79.61%
suprove12	1714	49	8	69.56%
suprove	1615	42	5	61.71%

bounds = sum of bounds reached (capped at bound 100)

* hors concours

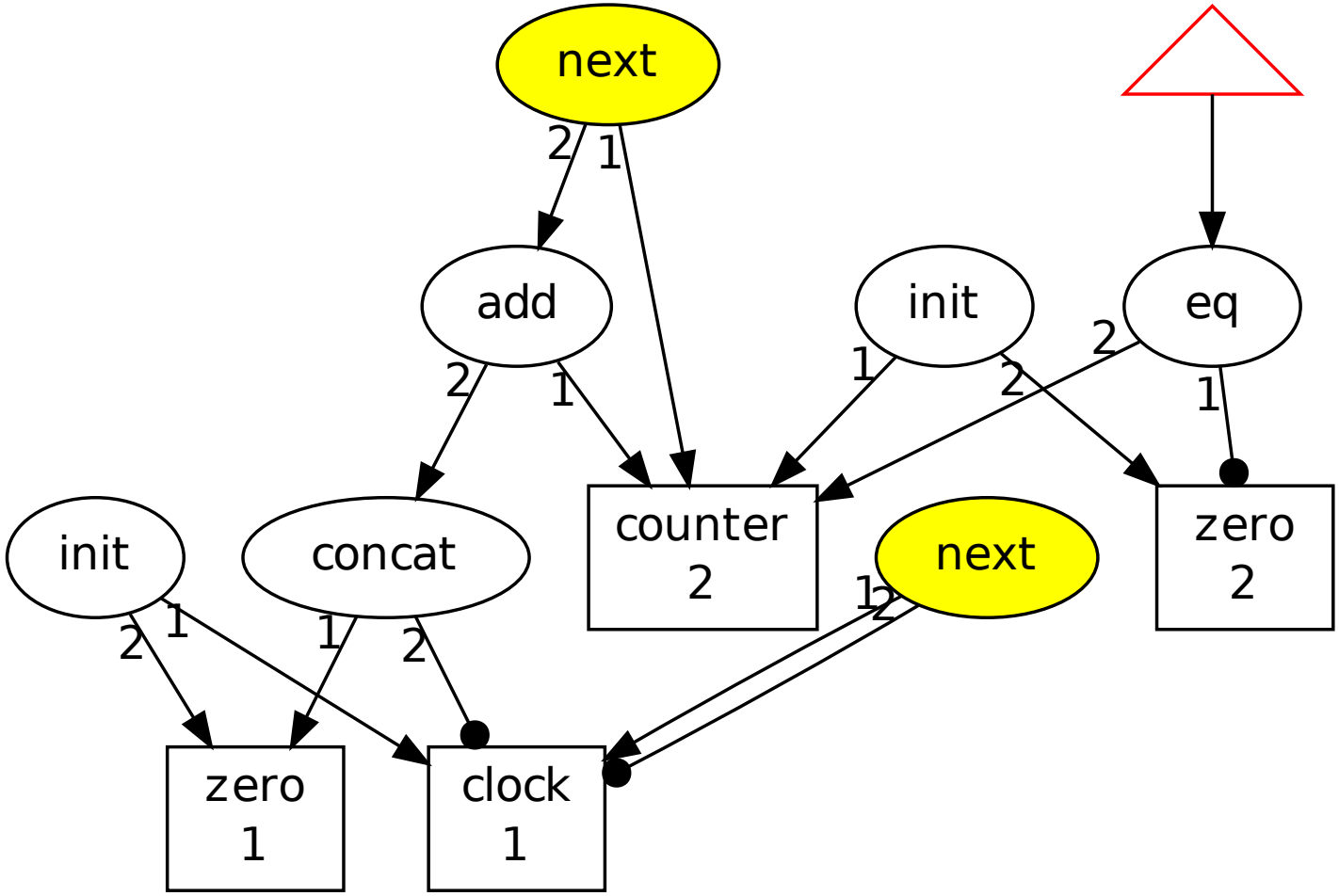
BTOR Word-Level Format

- originally proposed at BPR'08 [BrummayerLonsingBiere'08]
  - easy to parse word level format in contrast to SMT-LIB2
  - for bit-vectors and arrays equivalent to QF_ABV
  - **sequential extension** `next`, `anext`
  - compilers *from* SMT-LIB, compilers *to* AIGER

- recent new features
  - (non-recursive) lambda's or if you want *macros*
  - explicit next state functions
  - explicit initialization, constraints, properties as added to AIGER 1.9

- We need your benchmarks!

```
1 latch 2 counter
2 latch 1 clock
3 zero 1
4 concat 2 3 -2
5 add 2 1 4
6 next 2 1 5
7 zero 2
8 init 2 1 7
9 next 1 2 -2
10 init 1 2 3
11 eq 1 -7 1
12 bad 1 11
```



Conclusion

What has been achieved?

- new benchmarks, new versions, new model checkers
- state-of-the-art improved in all previous categories
- deep bounds track shows improvement too

How to continue?

- 'open' track with multiple safety and/or liveness properties
- issues with parallel model checking: non-determinism
- really really should enforce witnesses next time
- word-level model checking

HWMCC'13 Live Track Table SAT+UNSAT, SAT, UNSAT 32/45

solver	sat+uns	sat	uns	to	mo	uk	real sec	time sec	sum mb	max mb
iimc	G 140	G 78	S 62	45	0	0	4954	11634	58232	5136
simplive	S 122	S 68	B 54	43	20	0	7933	23483	30941	2990
tiprbmc12	122	58	64	63	0	0	8818	8751	3295	186
tiprbmc	B 120	58	62	65	0	0	7153	7083	3170	187
v3	108	63	45	75	2	0	10335	28843	26110	6716
tip	79	15	G 64	106	0	0	4307	4261	910	110
tipbmc	64	B 64	0	95	21	5	1681	1633	1119	220
aigbmclg1*	62	62	0	97	26	0	3458	3425	4925	3223
aigbmc*	56	56	0	96	33	0	2744	2715	1105	300

G = gold (winner), S = silver (2nd), B = bronze (3rd)

to = time out, mo = memory out, real = wall clock time
time = process time, sum = sum of max-mem, max = max mem
(restricted to solved benchmarks, e.g. for real, time, sum, max)
uk = unknown (tipmbmc reached 100k bound limit)
* hors concours

HWMCC'13

HWMCC'13 Multi Track Table SAT+UNSAT, SAT, UNSAT 34/45

solver	properties solved sat+uns	sat	uns	score based sat+uns	sat	uns
mulprove	97397	30379	67018	G 60.67%	S 21.42%	G 39.26%
mpmc12+	90520	31039	59481	55.54%	21.17%	34.37%
tiprbmc	76661	31913	44748	S 50.68%	20.71%	29.96%
tip	75166	30302	44864	48.67%	18.17%	S 30.50%
v3	75849	31553	44296	B 44.66%	B 20.67%	B 23.99%
tipbmc	32273	32273	0	21.90%	G 21.90%	0.00%
aigbmclg1*+	29693	29693	0	18.07%	18.07%	0.00%
aigbmc*	4063	4063	0	17.60%	17.60%	0.00%

+ mpmc12 winner of last year but has now discrepancies:
6s249-26 6s386-11 6s386-1 6s386-12 bob12m15m-0

* hors concours

$$\text{score} = \frac{1}{178} \sum_{i=1}^{178} \frac{\#\text{solved}_i}{\#\text{properties}_i}$$

HWMCC'13

HWMCC'13 Single Track SAT+UNSAT, SAT, UNSAT 36/45

solver	sat+uns	sat	uns	to	mo	s11	uk	real	time	space	max
suprove	G 138	48	G 90	100	10	0	0	14466	32981	145583	6636
suprove12	133	48	85	101	14	0	0	13981	29279	113454	4740
simple	132	G 51	81	107	9	0	0	11009	34802	138233	6636
iimc	S 113	S 38	S 75	134	1	0	0	15376	52115	55186	5564
v3	B 105	B 36	B 69	137	6	0	0	10633	41378	100085	5450
nuxmvexp	94	26	68	144	10	0	0	14136	55388	103733	6477
nuxmv	94	27	67	142	12	0	0	15212	59526	113280	6666
simbip	83	26	57	106	14	0	45	1104	2914	74992	4986
tip	78	23	55	170	0	0	0	8540	8499	3749	674
pdtrav	74	20	54	143	1	0	30	8962	27637	82892	3490
minireachic3tp	67	30	37	181	0	0	0	8031	15970	26565	5859
minireachic3	63	29	34	185	0	0	0	7945	15804	22887	5945
simpsat	43	22	21	105	11	0	89	12492	40406	39769	6613
smspdr	42	12	30	164	31	0	11	9903	9864	26731	4936
blimcn*	39	36	3	201	8	0	0	4965	4944	10767	2520
blimc*	37	34	3	202	9	0	0	3354	3333	9424	1327
tipbmc	35	35	0	171	20	0	22	3420	3400	8581	2374
aigbmclg1*	31	31	0	192	25	0	0	3408	3371	14296	2406
shiftbmcn	30	29	1	197	21	0	0	3073	3055	12378	2721
shiftbmc	29	28	1	196	23	0	0	2868	2851	12001	2530
fussiasto	29	12	17	200	1	11	7	2953	2936	7303	1396
aigbmc*	26	26	0	182	40	0	0	3170	3153	12015	2323
shiftbmcna	25	25	0	183	12	0	28	4518	4496	3976	752

* hors concours

HWMCC'13

Deep Bound Track Results 42/45

solver	bounds	insts	capped	score
tipbmc	3101	64	14	94.53%
blimc*	3022	64	12	94.22%
blimcn*	3010	64	12	94.14%
aigbmclg1*	2757	64	9	93.94%
shiftbmc	2750	64	10	93.71%
shiftbmcn	2745	64	10	93.70%
aigbmc*	2636	64	8	93.51%
nuxmvexp	2412	64	6	93.48%
nuxmv	2411	64	6	93.45%
v3	2534	64	6	92.76%
shiftbmcna	2807	59	11	86.60%
pdtrav	2662	58	10	85.23%
tip	525	64	0	84.43%
smspdr	1748	56	5	79.61%
suprove12	1714	49	8	69.56%
suprove	1615	42	5	61.71%

bounds = sum of bounds reached (capped at bound 100)
* hors concours

HWMCC'13