Hardware Model Checking Competition 2012

HWMCC'12

Armin Biere, Keijo Heljanko, Martina Seidl, Siert Wieringa

presented at

Formal Methods in Computer Aided Design 2012

FMCAD'12

Cambridge, UK

October 24, 2012

updated slides on November 8, 2012

Motivation

- revive interest in improving symbolic model checking technology
  - symbolic model checking does not scale *enough* in practice
  - only recently new academic research results
  - benchmarks have been lacking

- try to 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

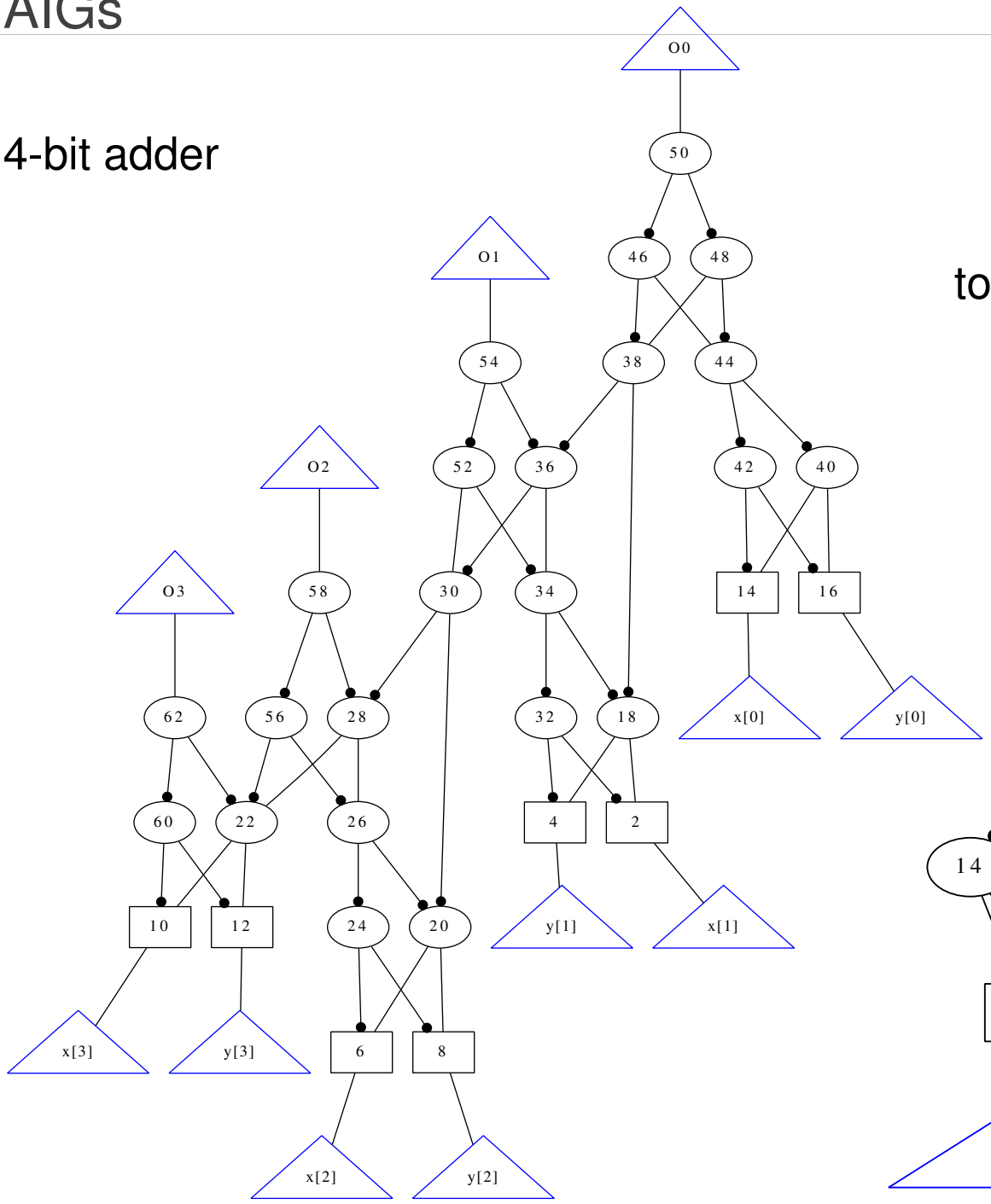
- relies on active support by submitters of benchmarks and model checkers

History

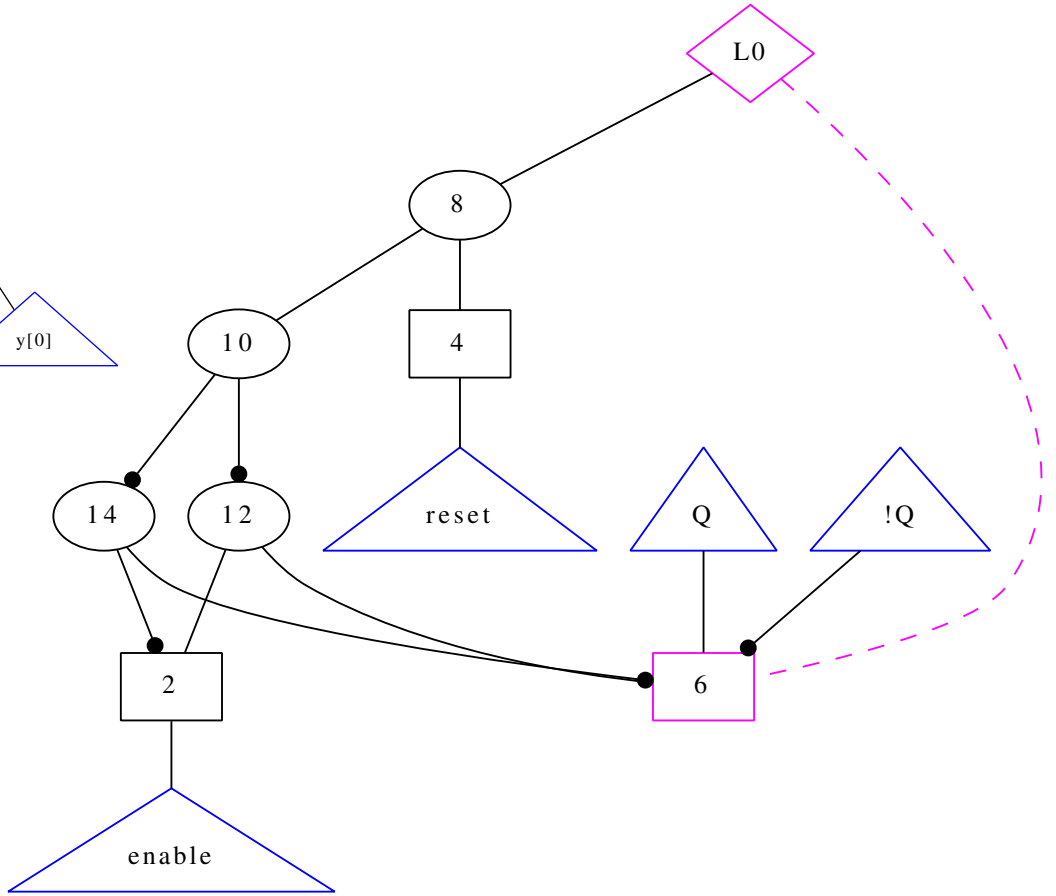
AIGER format AVM'06 Ascona	1st HWMCC CAV'07 Berlin	2nd HWMCC CAV'08 Princeton	3rd HWMCC CAV'10 FLOC'10 Edinburgh	4th HWMCC FMCAD'11 Austin	**5th** **HWMCC** **FMCAD'12** **Cambridge**
Founding Lunch CAV'06 FLOC'06 Seattle		HWMCC Lunch FMCAD'08 Portland			
2006	2007	2008	2010	2011	**2012**

- 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, HWMCC'11 and this year's HWMCC'12 at FMCAD
- HWMCC'11: old *single* property track, new *live*ness and new *multi* property track
- HWMCC'12: identical to HWMCC'11 except for new Deep Bounds Track

4-bit adder



toggle flip-flop with reset & enable



- And-Inverter-Graph (AIG) file format `http://fmv.jku.at/aiger`
  - for structural SAT and model checking problems
  - compact and (rather) easy to parse
- version 1.9 last year introduced new sections / properties `M I L O A B C J F`
  - without really changing syntax
  - after this FMCAD maybe replace ”justice” by ”naughty” so `J` by `N`
- new this year in HWMCC'12: reached bounds output format `u<k>`
- new binary format 2.0 is still work in progress

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

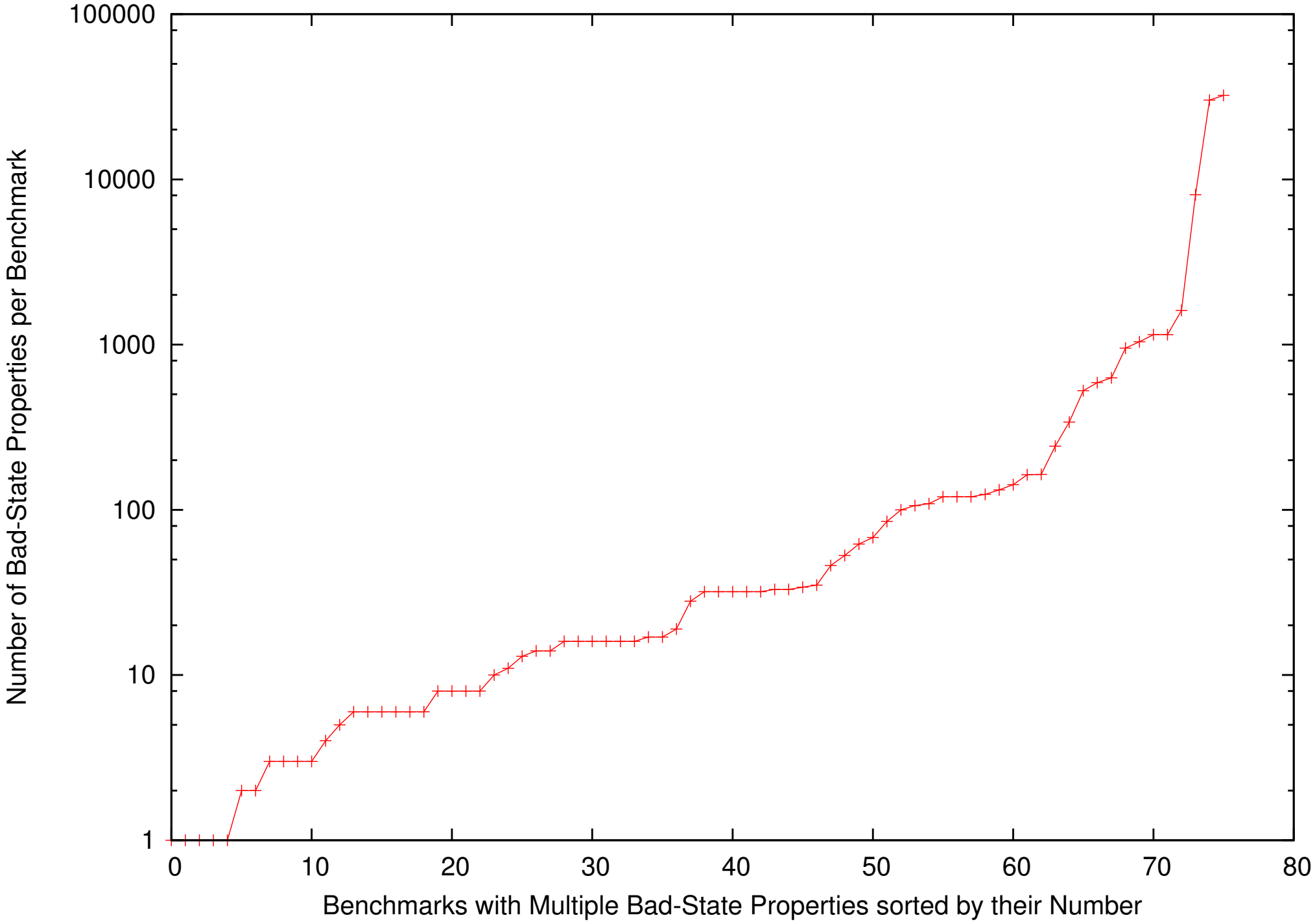
- one of the main goal of the competition is to collect benchmarks
- no new liveness benchmarks this year :-(
- so simply used all the available **118** benchmarks from last year

Multiple Bad-State Properties

- all 24 benchmarks from last year
- 15 new benchmarks submitted by Robert Brayton `bob12mXX.aig`
- 37 new multi property benchmarks submitted by Jason Baumgartner `6s1xx.aig`
- thus running on **76** benchmarks with 81082 properties and 1161 constraints
- substantial 3x increase in number of benchmarks!

Largest **Multi**ple Bad-State Benchmarks


```
aig 90587 25 1308 106 89254
aig 96536 199 6748 0 89589 630 6
aig 123461 1966 0 0 121495 1 47
aig 239708 23651 43950 30228 172107
aig 229205 79 26148 28 202978
aig 255161 237 15560 0 239364 68
aig 287499 4033 16230 0 267236 120 4
aig 333120 647 32922 0 299551 589
aig 349781 1700 14641 0 333440 109 39
aig 347337 5508 0 0 341829 2 2
aig 402625 207 34305 0 368113 8 8
aig 424962 452 13706 0 410804 100 2
aig 567561 1048 23957 0 542556 8064
aig 1001099 1381 101639 0 898079 32210
aig 1567435 1760 84925 0 1480750 124 21
aig 3544465 4708 260713 0 3279044 1041
```


100000
10000
1000
100
10
1
0
10
20
30
40
50
60
70
80
Number of Bad-State Properties per Benchmark
Benchmarks with Multiple Bad-State Properties sorted by their Number

Single Safety (Bad State) Property Benchmarks

- new Beem benchmarks submitted by Jori Dubrovin
  - converted from explicit state model checking benchmark set Beem
  - selected 103 from originally 688 benchmarks

- 62 new single property benchmarks submitted by Jason Baumgartner `6s1XX.aig`
  6s126.aig, 6s127.aig, 6s128.aig, 6s129.aig actually combinational

- 10 new single property benchmarks submitted by Robert Brayton `bob12sXX.aig`

- selected 135 HWMCC'11 benchmarks out of originally 465 benchmarks
  - either unsolved last year or
  - nontrivial (<10 sec) for complete solvers and not solved by all complete solvers
  - only used some of last year's solvers: **blimc**, **iimc**, **pdthrd**, **suprove**, **tipbmc**

- results in **310** single property benchmarks used this year

alphabetically

- **aigbmc**, **blimc** by Biere (Linz) new versions
- **pdtrav** by Cabodi, Nocco, Quer (Torino): 3 variants new versions
- **superprove**, **simplesat**, **mulprove** from Brayton's group (Berkeley) new versions
- **tarmo** by Wieringa (Helsinki) new version
- **tip** by Sörensson, Claessen (Göteborg): lots of variants new versions
- **v3**, **mpmc** by Chi-An Wu, Cheng-Yin Wu, Chung-Yang (Ric) Huang (Taiwan) **new**

aigbmc + blimc

aigbmc

- as in HWMCC'11 but with new PicoSAT version
- bounded model checker based on FMCAD'04 / CAV'05 papers by Heljanko et.al.
- published before competition as a *proof of concept* for new AIGER format 1.9
  - including liveness (justice) properties
  - but in multiple property mode stops as soon one property has a trace

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
- latest Lingeling, uses cloning for hard bounds `lglclone`

Super_prove2

(Brayton, Een, Mishchenko, Sterin)

Similar to last year but improved basic algorithms for BMC PDR

Aimed at SAT and UNSAT problems

- Use concurrency (3-4 PDRs, 3-4 BMCs, 1-2 Interpolates, 1-2 BDD Reachs, Rarity Simulation) throughout
- Simplify (*same as last year*)
  - Retiming, phase and temporal abstraction, signal correspondence, rewriting
- Abstract (*new method*)
  - New gate-level abstraction method (*&gla*)
  - Concurrent proving of current abstraction
  - New method of abstraction refinement.
- Speculate (*improved*)
  - Initial equalities found using “Rarity simulation”
  - Search for cex concurrently
  - If too many equalities, use filters to focus on those that involve latches
- If all fail, concurrently use PDRs, BMCs, Interpolations, and BDD reachs for the remaining time.

Simple_sat

(Brayton, Een, Mishchenko, Sterin)

Aimed at problems that are SAT.

- Simplify first, then
- Use remaining time concurrently with the following engines:
  - 4 PDRs, 1 INTRP, 4 BMCs, 1 BDD Reach, Rarity Simulation

Multi-prove

(Brayton, Een, Mishchenko, Sterin)

Aimed at Multi-output benchmarks

- Remove const-0 outputs
- Run BMC for 5 sec to find easy to prove SAT outputs. Remove these.
- Find isomorphic POs and keep only representatives.
- Simplify
- Find isomorphic POs and remove
- Run BMC for 20 sec. to find SAT outputs and remove
- Try remaining POs each for 10 sec to find easy SAT or UNSAT POs.
- Spend remaining time trying each PO for up to 100 sec using super_prove2.
- Report intermediate results at 1/3, 2/3, final points.

suprove, **simpsat**, **mulprove**

we want to use letters or digits for solver names only

suprove = Super_prove2

simpsat = Simple_sat

mulprove = Multi-prove

- PdTrav (Politecnico di Torino Reachability Analysis & Verification):

  multiple engine tool, submitted in two versions,
  single and multi-threaded (pthreads)

- features:

  initial transformations/reductions (combinational+sequential).
  Heuristically driven manager (expert system)
  Includes: cudd, minisat, abc (combinational synthesis)
  Engines: BMC, BDDs, k-induction, IC3, Interpolation.

- single property:

  portfolio-based,
  static + light weight dynamic classification & engine selection
  running 4 engines concurrently

- multiple properties:

  sequential version.
  static sorting of properties based on COI size
  For each property: BMC + UMC efforts
  reuse of invariants and/or previous CEXES

A) Name: TIP+Tarmo
B) Author: Siert Wieringa
C) Affiliation: Aalto University, Helsinki

D) some bullets on the technology/algorithms, other source:

- This model checker is identical to TIP's BMC algorithm by Niklas Sörensson, except that it uses the Tarmo parallel incremental SAT solver.

- The crucial idea behind Tarmo is its 'asynchronous incremental interface', through which more than one problem can be given to the solver at once.

- TIP+Tarmo solves multiple different 'bounds' in parallel.

- Conflict clause sharing is employed between solver threads.

E) difference to previous versions:

- Last year's Tarmo solver demonstrated the same principle but it used an integrated copy of 'aigbmc' for BMC.

- This year Tarmo has been completely rewritten and it is now easier to integrate in other more complex tools, like TIP.

> (A) name of the model checker (well I have this I think)

Tip

> (B) authors
> (C) their affiliation

Niklas Sörensson (now at Mentor Graphics, work done at Chalmers)
Koen Claessen (Chalmers)

> (D) some bullets on the technology/algorithms, as well
> as to which extend the code is based on other tools

* Safety checking based on Bradley's safety checking algorithm, augmented with (among other things) multiple property checking

* Liveness checking is k-liveness with preprocessing, based on the above safety checker

* SAT-solver used internally is MiniSat

> (E) difference to previous versions if applicable

- improved safety checker

- added preprocessing to the liveness checker

v3, mpmc

- **v3**, **v3final**, **v3i** entered single track only
  - authors: Cheng-Yin Wu, Chi-An Wu and Chung-Yang (Ric) Huang
  - based on ABC, MiniSAT, Boolector, QuteRTL
  - ABC preprocessing, 2xBMC, 2xPDR, new ITP
- **mpmc** entered multi track
  - authors: Chi-An Wu, Cheng-Yin Wu and Chung-Yang (Ric) Huang
  - simulation, probing, solving phases
  - based on ABC, MiniSAT
- affiliation National Taiwan University, Taipei, Taiwan
- *more details will be made available online*

Setup

- single property benchmarks (single + live tracks) as in HWMCC'07 - HWMCC'11
  - 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)

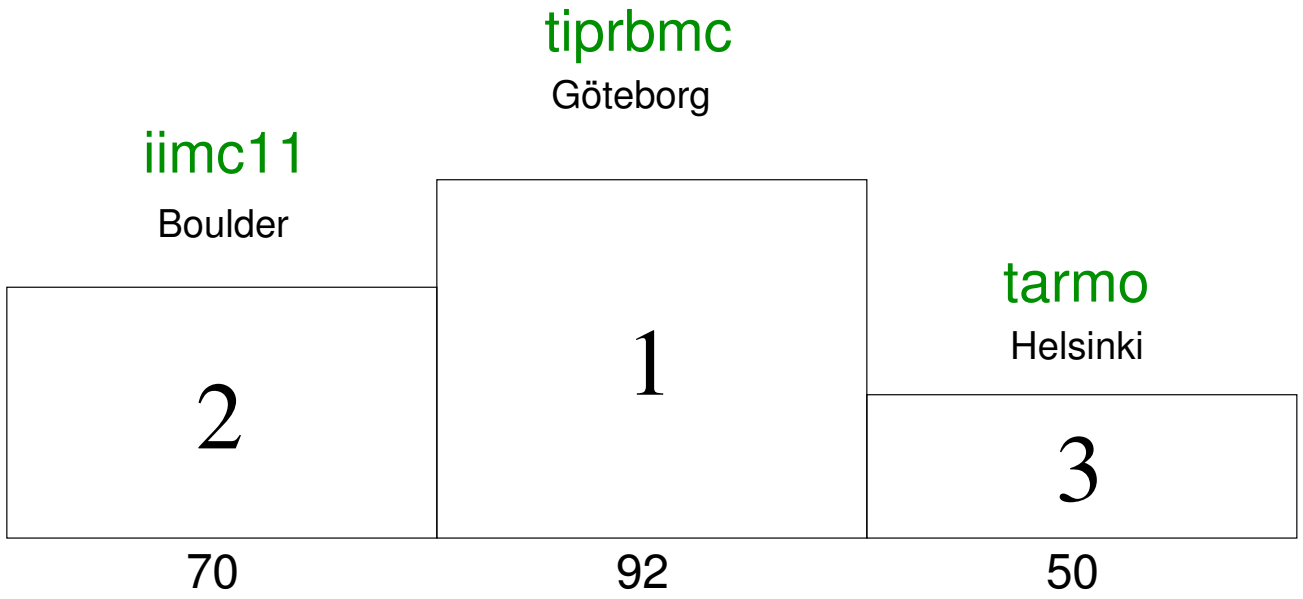
- Main tracks live, multi, single
  - three categories: SAT+UNSAT, SAT, UNSAT
  - no additional single threaded versus multi-threaded ranking

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/hwmcc12`
  - you will also get spread sheets and all the log files there

- Deep Bounds Awards sponsored by Oski Technology

- 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
- decided to only run it on the unsolved instances of the single track
- $500 award


tiprbmc
Göteborg
iimc11
Boulder
tarmo
Helsinki
2
1
3
70
92
50

SAT+UNSAT ranking

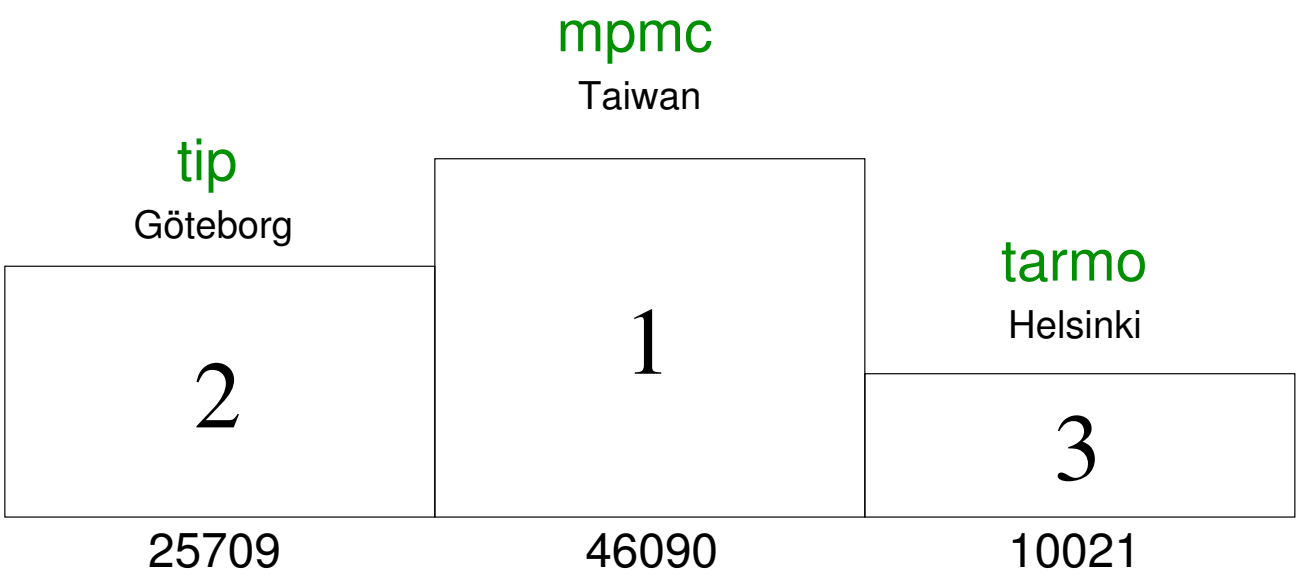
	solver	fnd	ok	sat	uns	fld	to	mo	uk	real time	space max
			SAT+UNSAT								
A	tiprbmc	118	92	46	46	26	26	0	0	4003 3975	1675 158
	tipl2s	118	87	44	43	31	31	0	0	6003 5976	725 33
	tip11	118	77	46	31	41	41	0	0	3513 3490	1608 242
B	iimc11	118	70	27	43	48	48	0	0	4686 4664	2109 240
	tip	118	62	14	48	56	56	0	0	2945 2928	367 25
	tipbmc	118	59	47	12	59	55	0	4	675 661	614 86
C	tarmo	118	50	50	0	68	66	2	0	1155 4516	3411 791
	aigbmc	118	45	45	0	73	68	0	5	1269 1242	541 51

SAT ranking

	solver	fnd	ok	sat	uns	fld	to	mo	uk	real time	space max
A	tarmo	118	50	50	0	68	66	2	0	1155 4516	3411 791
B	tipbmc	118	59	47	12	59	55	0	4	675 661	614 86
	tiprbmc	118	92	46	46	26	26	0	0	4003 3975	1675 158
	tip11	118	77	46	31	41	41	0	0	3513 3490	1608 242
C	aigbmc	118	45	45	0	73	68	0	5	1269 1242	541 51
	tipl2s	118	87	44	43	31	31	0	0	6003 5976	725 33
	iimc11	118	70	27	43	48	48	0	0	4686 4664	2109 240
	tip	118	62	14	48	56	56	0	0	2945 2928	367 25

UNSAT ranking

	solver	fnd	ok	sat	uns	fld	to	mo	uk	real	time	space	max
A	tip	118	62	14	48	56	56	0	0	2945	2928	367	25
	tiprbmc	118	92	46	46	26	26	0	0	4003	3975	1675	158
	tipl2s	118	87	44	43	31	31	0	0	6003	5976	725	33
B	iimc11	118	70	27	43	48	48	0	0	4686	4664	2109	240
	tip11	118	77	46	31	41	41	0	0	3513	3490	1608	242
	tipbmc	118	59	47	12	59	55	0	4	675	661	614	86
	tarmo	118	50	50	0	68	66	2	0	1155	4516	3411	791
	aigbmc	118	45	45	0	73	68	0	5	1269	1242	541	51


mpmc
Taiwan
tip
Göteborg
tarmo
Helsinki
2
1
3
25709
46090
10021

	solver	all	sat	uns
A	mpmc	46167	28515	17652
B	tip	37599	28718	8881
	tipbmc	28748	28748	0
C	tarmo11	4798	1678	3120
	aigbmc	1480	1480	0
	pdtmulti	1445	1338	107
	tarmo	1344	1344	0
	mulprove	39808	28885	10923
	pdtmulti	1445	1338	107

(last two solvers had discrepancies)

mulprove

- lots of discrepancies for 9 benchmarks:
  6s103 6s104 6s110 6s124 6s141 nusmvdme1d3multi
  nusmvdme2d3multi nusmvsyncarb10multi nusmvsyncarb5mult
- partially checked, so **mulprove** clearly is wrong
- detailed analysis, together with the ABC group, showed that **mulprove** did not handle latches correctly, which where initialzed to a value different from zero

pdtmulti

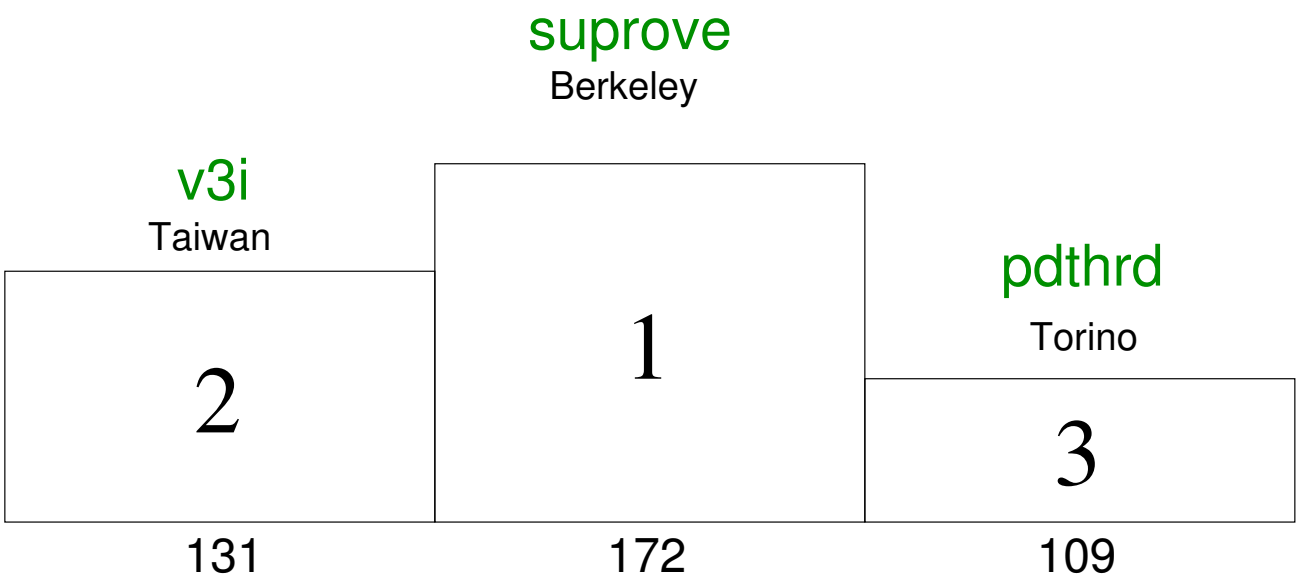
- 5 discrepancies for 6s103
  against 'unsat' of **mpmc** - but not witness by **pdtmul**, so not checked
- 4 discrepancies for bob9234specmulti:
  against more solvers, witness by **tipbmc** checked, so **pdtmulti** clearly wrong

	solver	all	sat	uns	dis
A	tip	46217	40745	5472	0
B	mpmc	46090	28517	17573	0
	tipbmc	20388	20388	0	0
C	tarmo	10021	6781	3240	0
	tarmo11	8538	5298	3240	0
	aigbmc	1480	1480	0	0
	mulprove	39807	28885	10922	328
	pdtmulti	1445	1338	107	9

no difference to SAT+UNSAT ranking

	solver	all	sat	uns	dis
A	mpmc	46090	28517	17573	0
B	tip	46217	40745	5472	0
C	tarmo11	8538	5298	3240	0
	tarmo	10021	6781	3240	0
	tipbmc	20388	20388	0	0
	aigbmc	1480	1480	0	0
	mulprove	39807	28885	10922	328
	pdtmulti	1445	1338	107	9

mpmc and **tip** switched positions


suprove
Berkeley
v3i
Taiwan
pdthrd
Torino
2
1
3
131
172
109

	solver	fnd	ok	sat	uns	fld	dis	to	mo	s11	s6	uk	real	time	space	max
A	suprove	310	172	64	108	138	0	134	3	0	0	1	26654	62035	108431	5660
	simpsat	310	152	63	89	158	0	152	6	0	0	0	19929	68905	92512	5359
B	v3i	310	131	44	87	179	0	173	3	3	0	0	17014	63777	40669	1682
	v3final	310	120	44	76	190	0	184	3	3	0	0	15557	56396	37637	2048
C	pdthrd	310	109	32	77	201	0	190	1	0	0	10	22818	82312	135718	4795
	v3	310	104	40	64	206	0	200	3	3	0	0	13700	54006	40448	3305
	tip	310	93	25	68	217	0	217	0	0	0	0	20493	20437	2598	281
	iimc11	310	58	20	38	252	0	250	0	1	1	0	11097	11063	27425	3257
	tarmo	310	48	48	0	262	0	251	11	0	0	0	6760	26132	27960	2735
	blimc11	310	43	43	0	267	0	265	2	0	0	0	7929	7895	9390	1231
	blimc	310	41	41	0	269	0	266	3	0	0	0	7808	7776	9875	1223
	tipbmc	310	37	37	0	273	0	211	0	0	0	62	5268	5249	8196	1065
	tipbmcf	310	37	37	0	273	0	260	6	0	0	7	5500	5457	10735	2376
	aigbmc	310	26	26	0	284	0	273	9	0	0	2	7584	7557	6218	1231
	suprove11	310	164	51	113	146	1	140	3	0	0	3	26861	51345	87364	3493
	pdthrd11	310	96	20	76	214	10	182	0	2	20	0	17517	49060	48238	2250
	tipbmc11	310	41	41	0	269	1	254	7	0	0	7	5067	5047	10388	2231
	pdtproc	310	81	0	81	229	1	218	1	0	0	9	16768	59648	164496	3963

Last 4 solvers suprove11, pdthrd11, tipbmc11, pdtproc showed discrepancies.

Discrepancies Single Track SAT+UNSAT

solver	fnd	ok	sat	uns	fld	dis	to	mo	s11	s6	uk	real	time	space	max
suprove11	310	164	51	113	146	1	140	3	0	0	3	26861	51345	87364	3493
pdthrd11	310	96	20	76	214	10	182	0	2	20	0	17517	49060	48238	2250
tipbmc11	310	41	41	0	269	1	254	7	0	0	7	5067	5047	10388	2231
pdtproc	310	81	0	81	229	1	218	1	0	0	9	16768	59648	164496	3963

These 4 solvers suprove11, pdthrd11, tipbmc11, pdtproc showed discrepancies:

tipbmc11 one discrepancy on the new benchmark 6s121
(witness invalid, 6 other solvers say 'unsat')

pdthrd11 problems with particular relational encoding in beem instances

pdtproc one discrepancy on the beem instance beemfrogs1f1
(says 'unsat', but 'sat' by 10 other, checked)

suprove11 one discrepancy on the new benchmark bob12s107

(suprove11 said 'sat' without witness but v3(final) 'unsat'
and tip(f) proved 'unsat' too in 9.51 hours, benchmark
is confirmed to be 'unsat' by Robert Brayton)

	solver	fnd	ok	sat	uns	fld	dis	to	mo	s11	s6	uk	real	time	space	max
A	suprove	310	172	64	108	138	0	134	3	0	0	1	26654	62035	108431	5660
	simpsat	310	152	63	89	158	0	152	6	0	0	0	19929	68905	92512	5359
B	tarmo	310	48	48	0	262	0	251	11	0	0	0	6760	26132	27960	2735
C	v3i	310	131	44	87	179	0	173	3	3	0	0	17014	63777	40669	1682
	v3final	310	120	44	76	190	0	184	3	3	0	0	15557	56396	37637	2048
	blimc11	310	43	43	0	267	0	265	2	0	0	0	7929	7895	9390	1231
	blimc	310	41	41	0	269	0	266	3	0	0	0	7808	7776	9875	1223
	v3	310	104	40	64	206	0	200	3	3	0	0	13700	54006	40448	3305
	tipbmcf	310	37	37	0	273	0	260	6	0	0	7	5500	5457	10735	2376
	tipbmc	310	37	37	0	273	0	211	0	0	0	62	5268	5249	8196	1065
	pdthrd	310	109	32	77	201	0	190	1	0	0	10	22818	82312	135718	4795
	aigbmc	310	26	26	0	284	0	273	9	0	0	2	7584	7557	6218	1231
	tip	310	93	25	68	217	0	217	0	0	0	0	20493	20437	2598	281
	iimc11	310	58	20	38	252	0	250	0	1	1	0	11097	11063	27425	3257

	solver	fnd	ok	sat	uns	fld	dis	to	mo	s11	s6	uk	real	time	space	max
A	suprove	310	172	64	108	138	0	134	3	0	0	1	26654	62035	108431	5660
	simpsat	310	152	63	89	158	0	152	6	0	0	0	19929	68905	92512	5359
B	v3i	310	131	44	87	179	0	173	3	3	0	0	17014	63777	40669	1682
C	pdthrd	310	109	32	77	201	0	190	1	0	0	10	22818	82312	135718	4795
	v3final	310	120	44	76	190	0	184	3	3	0	0	15557	56396	37637	2048
	tip	310	93	25	68	217	0	217	0	0	0	0	20493	20437	2598	281
	v3	310	104	40	64	206	0	200	3	3	0	0	13700	54006	40448	3305
	iimc11	310	58	20	38	252	0	250	0	1	1	0	11097	11063	27425	3257
	tipbmcf	310	37	37	0	273	0	260	6	0	0	7	5500	5457	10735	2376
	tipbmc	310	37	37	0	273	0	211	0	0	0	62	5268	5249	8196	1065
	tarmo	310	48	48	0	262	0	251	11	0	0	0	6760	26132	27960	2735
	blimc	310	41	41	0	269	0	266	3	0	0	0	7808	7776	9875	1223
	blimc11	310	43	43	0	267	0	265	2	0	0	0	7929	7895	9390	1231
	aigbmc	310	26	26	0	284	0	273	9	0	0	2	7584	7557	6218	1231

sponsored by

Oski Technology

presented by

Vigyan Singhal

Deep Bound Track Results

solver	bounds	instances	
tipbmcf	21138	94	(incl. 374 unrolled)
tarmona*	15079	93	
blimc*	15078	94	
tipbmcnp	14128	94	(incl. 374 unrolled)
v3final	11932	94	
aigbmc	9165	94	
tipbmc	6820	23	(incl. 106 unrolled)
v3i	5833	94	
tarmo*	5813	5	
tarmof*	4605	4	
suprove	3478	76	
tipf	277	94	
tip	3	1	

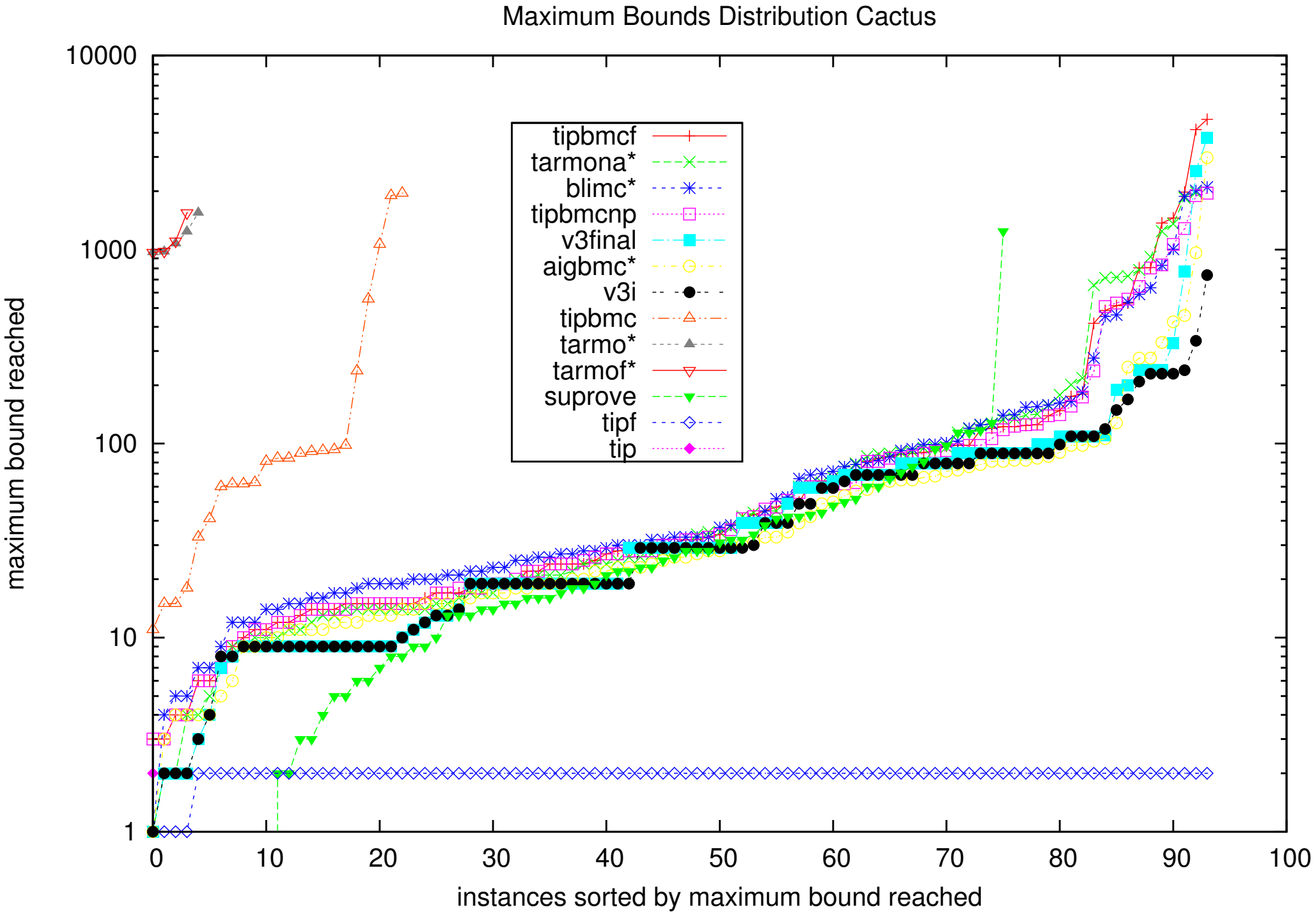
* = organizer solvers run hors concours

'np' in 'tipnp' & and 'tipbmcnp': no piped output filter

'f' in 'tipf' & 'tipbmcf': added 'fflush' and no piped output

'f' in 'tarmof: added 'fflush' to shell script

'na' in 'tarmona: no scripting


Maximum Bounds Distribution Cactus
maximum bound reached
instances sorted by maximum bound reached
10000
1000
100
10
1
0
10
20
30
40
50
60
70
80
90
100
tipbmcf
tarmona*
blimc*
tipbmcnp
v3final
aigbmc*
v3i
tipbmc
tarmo*
tarmof*
suprove
tipf
tip

Deep Bounds Track Instances

calculated for 97 instances unsolved in the single track:

6s100 6s105 6s119 6s122 6s127 6s128 6s129 6s13 6s145 6s148 6s152
6s158 6s160 6s161 6s163 6s171 6s174 6s177 6s178 6s179 6s180 6s185
6s186 6s187 6s188 6s190 6s191 6s192 6s195 6s22 6s23 6s24 6s27
6s28 6s29 6s33 6s35 6s36 6s37 6s38 6s39 6s42 6s44 6s45 6s46 6s7
beemandrsn6b1 beembkry5b1 beemblks2b1 beembrdg3b1 beembrptwo1b2
beemcmbrdg1b1 beemextnc1b1 beemfish6b1 beemfwt1f1 beemhanoi1f1
beemkrebs4b1 beemlann4b1 beemldelec1b1 beemldfilt5b1 beemlifts3b1
beemlmprt5b1 beemlmprt7b1 beemloyd3b1 beemmcs3b1 beemmsmie3b1
beempgmprot1b2 beempgsol2b1 beemplc1b2 beemptrsn4b1 beemptrsn7b1
beemptrsn7f1 beemrether3b1 beemskbn2b1 beemszmsk1b1 beemtlphn4b1
bob12s06 bob12s08 bob12s09 bobpcihm bobsmcodic bobsmminiuart
intel012 intel013 intel014 intel015 intel016 intel025 intel027
intel028 intel032 intel034 intel048 intel066 nusmvdme116
nusmvdme216 pdtswvsam6x8p4

Conclusion

- new benchmarks, new versions, new model checkers
- state-of-the-art improved in all previous categories
- new deep bounds track
- some suprising discrepancies ...

  ... so really should enforce witnesses next time

Discussion

- word-level model checking
- ranking the multiple property track
- how to get more liveness benchmarks
- issues with parallel model checking
- ideas for future awards
- how to continue?